Die Themen des 17. Deutschen IT-Sicherheitskongresses

In Kürze verfügbar

Das Buch zum Kongress

Deutschland.
Digital.
Sicher.
30 Jahre BSI

Herausgegeben vom Bundesamt für Sicherheit in der Informationstechnik

Themen des Kongresses als wertvolles gebundenes Buch:

- **Neuartige Social Engineering Angriffe**
- **Active Directory in Gefahr**
- **Notfallangebot für KMU**
- **Hochsichere Cloud**
- **Cyber Resilience**
- **Post-Quanten-Kryptografie**
- **Umgang mit Datenschutzkatastrophen**
- **Rechtspflicht für Updates?**
- **Awareness Trainings**
- **Identitätsdiebe in die Schranken weisen!**
- **usw.**

Erscheint im Februar 2021. DIN A 5, Hardcover. Subskriptionspreis für Online-Kongresszuhörer bis 15.2.: EUR 49,60, danach EUR 62,00, ISBN 978-3-922746-83-6 www.secumedia.shop

Inhaltsverzeichnis

Sichere Identitäten

Tobias Assmann, Dr. Detlef Hühnlein, Tina Hühnlein, Michael Rauh, Tobias Wich, ecsec GmbH; Ann-Kristin Derst, Hanno Koop, Thorben Pohl, Bundesamt für Sicherheit in der Informationstechnik
Anwendungsintegration sicherer Digitaler Identitäten unter Berücksichtigung von Security by Design Prinzipien

Maximilian Westers, Prof. Dr.-Ing. Andreas Mayer, Hochschule Heilbronn:
Sicherheit von Single Sign-On: Ein Überblick

Jörg Breuer, Christian Stengel, Dr. Friedrich Tönsing, Robert Zastrau, Deutsche Telekom Security GmbH:
OPTIMOS – Trusted-Service-Management-System – eine Infrastruktur für sichere mobile Dienste

Cyber-Sicherheit für die Wirtschaft

Julian Suleder, M.Sc., ERNW Research GmbH, Heidelberg; Dr. Dina C. Truxius, Bundesamt für Sicherheit in der Informationstechnik:
Medical Device Security: Results from Project ManiMed

Dipl. Math. Angelika Jaschob, Bundesamt für Sicherheit in der Informationstechnik:
Ganzheitliches Notfallangebot für KMU

Jan Tiedemann, Peter Rost, Annegrit Seyerlein-Klug, Jens Kulikowski, secunet Security Networks AG:
Multifunktionale und sichere Edge-Architektur für digital transformierte Industrie und Kritische Infrastrukturen

IT-Sicherheit und Recht

Dr. Dennis-Kenji Kipker, Universität Bremen:
IT-Sicherheitsupdates: Pflichten für Hersteller und Verkäufer

Joerg Heidrich, Heidrich Rechtsanwälte Partnerschaftsgesellschaft mbB; Dr. Christoph Wegener, wecon.it-consulting:
Technischer und juristischer Umgang mit Datenschutz-Katastrophen: Vorbereitung, Krisenmanagement und „Lessons learned“

Verena Wingerter, DCSO Deutsche Cyber-Sicherheitsorganisation | Hertie School:
Staatliche Regulierung des Internets in Russland

Sichere Digitalisierung

Prof. Dr.-Ing. Andreas Mayer, Hochschule Heilbronn:
Virtuelle Hauptversammlungen: Ein sicherer Ersatz für Präsenzveranstaltungen?

Jennifer Breuer, Michael Hoppe, Stephan Kohzer, Marina Voigtländer, Bundesamt für Sicherheit in der Informationstechnik
Die sichere Digitalisierung von Wahlen am Beispiel des Modellprojektes Online-Sozialwahlen 2023

Jan-Niclas Hilgert, Martin Lambertz, Fraunhofer FKIE:
Auf den Lime gegangen: Vor- und Nachteile der neuen Mikromobilität

Sichere mobile Kommunikation

Timo Pohl, Arnold Sykosch, Rheinische Friedrich-Wilhelms-Universität Bonn:
Benutzerfreundliche Schutzmechanismen gegen USB-basierte Angriffe unter Linux

Christian Banse, Florian Wendland, Konrad Weiss, Fraunhofer AISEC:
Automatisierte Compliance-Prüfung in Software-Artefakten

Philipp Goldberg, Hochschule Darmstadt:
Zur Security Awareness bei Nutzern von Smartphones

Dr. Helge Kreutzmann, Bundesamt für Sicherheit in der Informationstechnik:
Zukünftige harmonisierte Zertifizierung in Europa am Beispiel 5G, CC und IACS

Dipl.-Ing. Thomas Salvador, QM Experts GmbH; Martina Petersen, CertVision GmbH
Entwicklung der Informationssicherheit im Bereich der automobilen Lieferkette auf dem Weg zum TISAX-Standard in der Version 5.0

Paul Bastian, Micha Kraus, Jörg Fischer, Christoph Bösch Bundesdruckerei GmbH:
Self-Sovereign Identity - Vertrauensbasis für selbstbestimmte Identitätsnetzwerke

Post-Quanten-Kryptografie

Dr. Tobias Hemmert, Dr. Manfred Lochter, Stephanie Reinhardt, Bundesamt für Sicherheit in der Informationstechnik; Prof. Dr. Daniel Loebenberger, Prof. Dr. Marian Margraf, Prof. Dr. Georg Sigl, Fraunhofer AISEC:
Quantencomputerresistente Kryptografie: Aktuelle Aktivitäten und Fragestellungen

Dr. Heike Hagemeier, Dr. Stavros Kousidis, Dr. Thomas Wunderer, Bundesamt für Sicherheit in der Informationstechnik:
Standardisierung von Post-Quanten-Kryptografie und Empfehlungen des BSI

Alexander Zeier, Prof. Dr. Alexander Wiesmaier, Prof. Dr. Andreas Heinemann, Hochschule Darmstadt:
Zur Integration von Post-Quantum Verfahren in bestehende Softwareprodukte

Zurück zum Inhaltsverzeichnis

Das Spiel mit der Angst - Erfolgsfaktoren neuartiger Social-Engineering-Angriffe im Kontext der COVID19-Pandemie

Dr. Niklas J. Hellemann[1]

Kurzfassung:
Social Engineering, also die Manipulation von Menschen zur Erreichung eines Ziels, hat in den letzten Jahren immer mehr an Bedeutung gewonnen. Der Großteil der erfolgreichen Cyber-Angriffe auf Unternehmen enthält mindestens einen Schritt, der auf den menschlichen Faktor zielt. Die Angreifenden bedienen sich dabei bewährter Prinzipien u.a. der Sozialpsychologie, um z.B. Informationen zu extrahieren oder NutzerInnen zur Installation von Schadsoftware zu bewegen.

Im Kontext der COVID19-Pandemie ist hier eine stark erhöhte Dynamik zu beobachten: Angreifende machen sich die allgemeine Verunsicherung zu Nutze und adressieren veränderte Arbeitsbedingungen („Home-Office“) und die Einführung neuartiger Tools und Lösungen, wie z.B. Kollaborationssoftware. Hinzu kommt, dass genau jene Tools auch potenzielle Kanäle für neuartige Angriffe darstellen können. Auch ein Jahr nach Ausbruch der Pandemie bleibt das Aufkommen an Angriffen zudem stark erhöht, sodass eine hohe Relevanz besteht, die veränderte Situation im Bereich des Social Engineerings genauer zu betrachten. Es stellt sich daher ganz konkret die Frage, ob von den beschriebenen veränderten Angriffstaktiken auch tatsächlich eine erhöhte Gefahr, im Sinne einer gesteigerten Erfolgswahrscheinlichkeit, ausgeht.

Dieser Artikel soll diese Fragestellung näher beleuchten. Er bedient sich dabei der Analyse eines großzahligen Datensatzes von Verhaltensdaten, der 1,4 Mio. simulierte Phishing-Angriffe aus der Phase der COVI19-Pandemie im Jahre 2020 enthält. Die entsprechenden Angriffe wurden über E-Mail sowie Messenger-Tools (Microsoft Teams) durchgeführt. Die Ergebnisse dieser Analyse legen nahe, dass in der Tat eine besonders hohe Erfolgswahrscheinlichkeit bei Angriffen besteht, die die aktuelle Thematik ausnutzen und das „Home-Office“-Setting adressieren (z.B. die Einführung neuer Tools). Darüber hinaus lassen die Resultate vermuten, dass von Angriffen, die über Messenger-Tools ausgeführt werden, eine besondere Gefahr ausgeht.

Abschließend werden entsprechende Möglichkeiten diskutiert, wie Unternehmen und öffentliche Organisationen auf die veränderte Angriffslage reagieren sollten, z.B. durch die spezifische Sensibilisierung von Mitarbeitenden und spezielle Absicherung von Cloud-Tools.

Stichworte: Awareness, COVID19, Home-Office, Phishing, Psychologie, Social Engineering

1. Social Engineering, Psychologie und „Faktor Mensch“

Wie der Begriff bereits beschreibt, handelt es sich bei Social Engineering um die Manipulation („Engineering“) anderer Menschen („Social“), mit dem Ziel, sich unmittelbar zu bereichern oder sich mittelbar Zugang zu Informationen oder technischen Systemen zu verschaffen. „Social Engineers“ nutzen dabei u.a. Prinzipien der Sozialpsychologie aus, um z.B. kritisches Denken auszuschalten und das menschliche Ziel zu einer potenziell schädlichen Handlung zu verleiten. Beispielsweise werden sogenannte Heuristiken ausgenutzt, die Menschen über die Entwicklungsgeschichte ausgebildet haben, wie z.B.

[1] SoSafe Cyber Security Awareness (SoSafe GmbH), Köln

das Hören auf Autoritäten (der sog. „Authority Bias“[2]) oder auch Neugier (die sog. „Curiosity Tendency“[3]). Diese Heuristiken werden von Menschen unmittelbar und unterbewusst verwendet, weil sie Schlussfolgerungen erleichtern und dabei helfen, kognitive Ressourcen einzusparen. Für Angreifende und Cyberkriminelle bedeutet das Aktivieren dieser Heuristiken bei ihren Opfern eine Möglichkeit, die kritische Auseinandersetzung zu minimieren und die Erfolgswahrscheinlichkeit eines Angriffs zu erhöhen.

Inhaltlich verwenden die Angreifenden gezielt Themen, die eine möglichst hohe Erfolgswahrscheinlichkeit (wie z.B. für den Klick auf einen Link in einer Phishing-E-Mail) bei einer möglichst großen Gruppe an Opfern versprechen. Dies kann insbesondere dadurch erreicht werden, dass eine besonders hohe Relevanz (aufgrund persönlichen „Involvements“) bei dem Opfer erzeugt wird. Beim „Spear-Phishing“ erzielt der Angreifende dies durch eine vorherige Informationssammlung („Information Gathering“) und eine sehr plausible Gestaltung des Social-Engineering-Versuchs. Im Falle von Flächenangriffen versprechen aber insbesondere Themen, die bei einer großen Gruppe von Menschen eine hohe Relevanz besitzen, die höchste Erfolgswahrscheinlichkeit. Daher lassen sich zahlreiche Phishing-Kampagnen beobachten, die z.B. Sportereignisse, saisonale Feiertage (z.B. Thanksgiving oder Weihnachten) übliche Prozesse in Unternehmen (z.B. IT-Migrationen) thematisieren. Üblicherweise weisen Angriffsversuche mit solchen Themen nämlich höhere Erfolgs- bzw. Klickraten auf.

2. Der Einfluss von Corona auf die Psyche und Arten der Zusammenarbeit

Der Ausbruch der Corona-Pandemie hatte einen grundlegenden Einfluss auf weite Teile unseres Lebens. Zum einen wurden durch das akute gesundheitliche Risiko viele Menschen stark verunsichert und zeigten emotionale Reaktionen von Angst bis hin zur Depression[4]. Zum anderen wurden im Rahmen des Lockdowns im Frühjahr des Jahres 2020 nahezu sämtliche Unternehmen weltweit dazu gezwungen, ihre Belegschaft wenn möglich in einen „Remote-Work“-Modus zu versetzen. Prozesse mussten komplett neu geplant und für viele Mitarbeitende bisher noch neuartige Tools (wie z.B. Videoconferencing- oder Kollaborationssoftware) eingeführt werden. Für einen Großteil der Belegschaft war dies eine völlig neue Arbeitssituation[5].

Auf der Angreiferseite wurden diese veränderten Rahmenparameter in kürzester Zeit erkannt und ausgenutzt. Bereits in den ersten Wochen des Ausbruchs der Pandemie in China und Italien waren die ersten Phishing-Kampagnen mit Corona-Bezug zu beobachten sowie ein stark erhöhtes Niveau von Phishing-Mails, während in Deutschland diese Dynamik zunächst noch nicht zu beobachten war[6]. Die Themen, die bei diesen ersten Corona-Phishing-Mails im Vordergrund standen, zielten einerseits auf psychologische

---

2 Milgram, Stanley (1963). "Behavioral Study of obedience". The Journal of Abnormal and Social Psychology. 67 (4): 371–378.

3 Kidd, C., & Hayden, B. Y. (2015). The Psychology and Neuroscience of Curiosity. Neuron, 88(3), 449–460.

4 Sherman A. Lee, Mary C. Jobe, Amanda A. Mathis, Jeffrey A. Gibbons, Incremental validity of coronaphobia: Coronavirus anxiety explains depression, generalized anxiety, and death anxiety, Journal of Anxiety Disorders, Volume 74, 2020.

5 https://www.bitkom.org/Presse/Presseinformation/Jedes-dritte-Untershynehmen-bietet-Arbeit-im-Homeshyoffice-an.html

6 https://threatpost.com/cynet-the-coronavirus-is-already-taking-effect-on-cyber-security-this-is-how-cisos-should prepare/153758/

Faktoren, wie das Bedürfnis nach Schutz, der Motivation zur Hilfsbereitschaft (z.B. in Form von vermeintlichen Mails der WHO[7]) oder der Suche nach Information (z.B. in Form von angeblichen Nachrichten zu Infektionszahlen[8]).

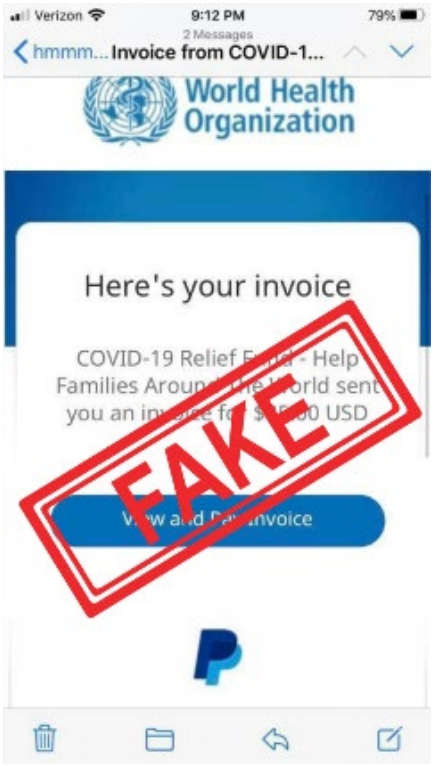


Abbildung 1: Beispiel einer Phishing-Kampagne, die die Motivation zur Hilfsbereitschaft im Zuge der COVID19-Pandemie ausnutzt
(Quelle: https://www.who.int/about/communications/cyber-security)

Andererseits konnte ein Fokus der Angreifer auf neue Cloud-Anwendungen beobachtet werden. So stieg der Anteil von „Credential Theft"-Versuchen für Kollaborationstools wie z.B. „Microsoft Teams" oder „Slack"[9]. Zahlreiche Phishing-Kampagnen versuchten beispielsweise, Zugangsdaten zu den entsprechenden Tools über Einladungsmails und nachgeschaltete Login-Masken zu erbeuten. Gerade wenn derartige Angriffsversuche auf eine verunsicherte und unerfahrene Nutzergruppe treffen, ist hier eine erhöhte Erfolgswahrscheinlichkeit zu erwarten.

Gleichzeitig birgt auch die erhöhte Verbreitung von Kollaborationstools selbst ein verschärftes Angriffspotenzial. Gerade unerfahrene NutzerInnen könnten dazu neigen, die neuartige Umgebung z.B. eines firmeninternen Cloud-Chat-Tools als geschützten Raum wahrzunehmen und unbedarfter auf manipulative Anfragen zu reagieren. Kombiniert mit der Tatsache, dass Angreifende vermehrt Zugangsdaten erlangen, wäre dadurch ein

7 https://www.who.int/about/communications/cyber-security
8 https://www.handelsblatt.com/technik/it-internet/it-sicherheit-der-perfekte-koeder-cyberkriminelle-nutzen-die-corona-panik/v_detail_tab_comments/25641346.html
9 https://www.statista.com/statistics/1175657/increase-identity-theft-coronavirus-outbreak/

stark erhöhtes Risiko abzuleiten. Dies wäre insofern von hoher Relevanz, da die Nutzung von Kollaborationstools weiterhin zunehmen wird – auch unabhängig vom weiteren Verlauf der Pandemie.

3. Daten zu Erfolgsparametern neuartiger Social-Engineering-Angriffe

Im Folgenden wird daher der Versuch unternommen, das beschriebene potenzielle Risiko zu bewerten und die folgenden Fragen zu beantworten:

1. Besitzen COVID19-bezogene Phishing-Angriffe tatsächlich eine erhöhte Erfolgswahrscheinlichkeit?
2. Wie ist die Erfolgswahrscheinlichkeit bei neuartigen Angriffsformen und -Kanälen einzuschätzen?

Die Analyse der geschilderten Fragen stützt sich dabei auf verschiedene Datensätze aus dem Betrieb der „SoSafe-Awareness-Plattform“, die u.a. ein Modul zur Simulation von Phishing-Angriffen beinhaltet. Zum einen wurde hier ein repräsentativer Datensatz verwendet, der ca. 1,4 Mio. während der COVID19-Phase versendete simulierte Phishing-E-Mails und entsprechende (anonyme) Reaktionsdatenpunkte enthält. Daneben wurde eine kleinere Stichprobe analysiert, in der neuartige Angriffskanäle zum Einsatz kamen, namentlich über automatisierte Social-Engineering-Angriffe auf Basis der „Microsoft Teams“-Plattform.

3.1. E-Mail-Phishing: Erfolgswahrscheinlichkeit von COVID19- und „Home-Office“-bezogenen Angriffen

Um zu bestimmen, ob E-Mail-Phishing-Angriffe, die das pandemische Geschehen sowie veränderte Bedingungen der Zusammenarbeit adressieren, eine erhöhte Erfolgswahrscheinlichkeit und damit auch ein gesteigertes Risiko besitzen, kann ein differenzierter Blick auf Klickraten simulierter Phishing-Angriffe zur Erkenntnis beitragen. Dieser Blick erfolgt im vorliegenden Artikel auf Basis von Simulationsdaten der SoSafe-Awareness-Plattform. Diese Plattform versendet im Rahmen des Phishing-Simulations-Moduls laufend und randomisiert nachempfundene Phishing-Mails an MitarbeiterInnen. Bei einer Interaktion mit der entsprechenden E-Mail (z.B. dem Klick auf einen Link) werden die EmpfängerInnen auf eine sog. Lernseite geleitet, die der Aufklärung dient. Gleichzeitig wird die Interaktion entsprechend (aber anonym) registriert. Dadurch ergibt sich die Möglichkeit auf Basis von Klickraten Rückschlüsse auf den Effekt verschiedener Parameter, wie z.B. den Inhalt oder auch technische Vektoren, der E-Mails zu ziehen.

Methodisch sei darauf verwiesen, dass es sich hierbei nicht um ein echtes Experiment handelt, da auf Seite der EmpfängerInnen und Organisationen natürlich eine Vielzahl an Variablen nicht kontrollierbar ist. Am ehesten können diese Daten somit als das Resultat eines „Feldexperiments“ angesehen werden, mit entsprechend verringerter interner Validität aber höherer Generalisierbarkeit und Praxisrelevanz. Daneben ist zu beachten, dass es sich bei den vorliegenden Analysen zudem lediglich um eine deskriptive Auswertung handelt.

Im Folgenden werden drei Gruppen von E-Mails und entsprechende Klickraten beschrieben:

(1) Tabelle 1: E-Mails, die einen direkten Bezug zur pandemischen Situation oder deren Auswirkungen haben (Abbildung 2 zeigt den kompletten Inhalt der Vorlage zu einer E-Mail)
(2) Tabelle 2: E-Mails mit und ohne direkten COVID19-Bezug
(3) Tabelle 3: E-Mails, die entweder die Einführung eines modernen Kollaborationstools thematisieren oder dieses Tool als bereits eingeführt annehmen

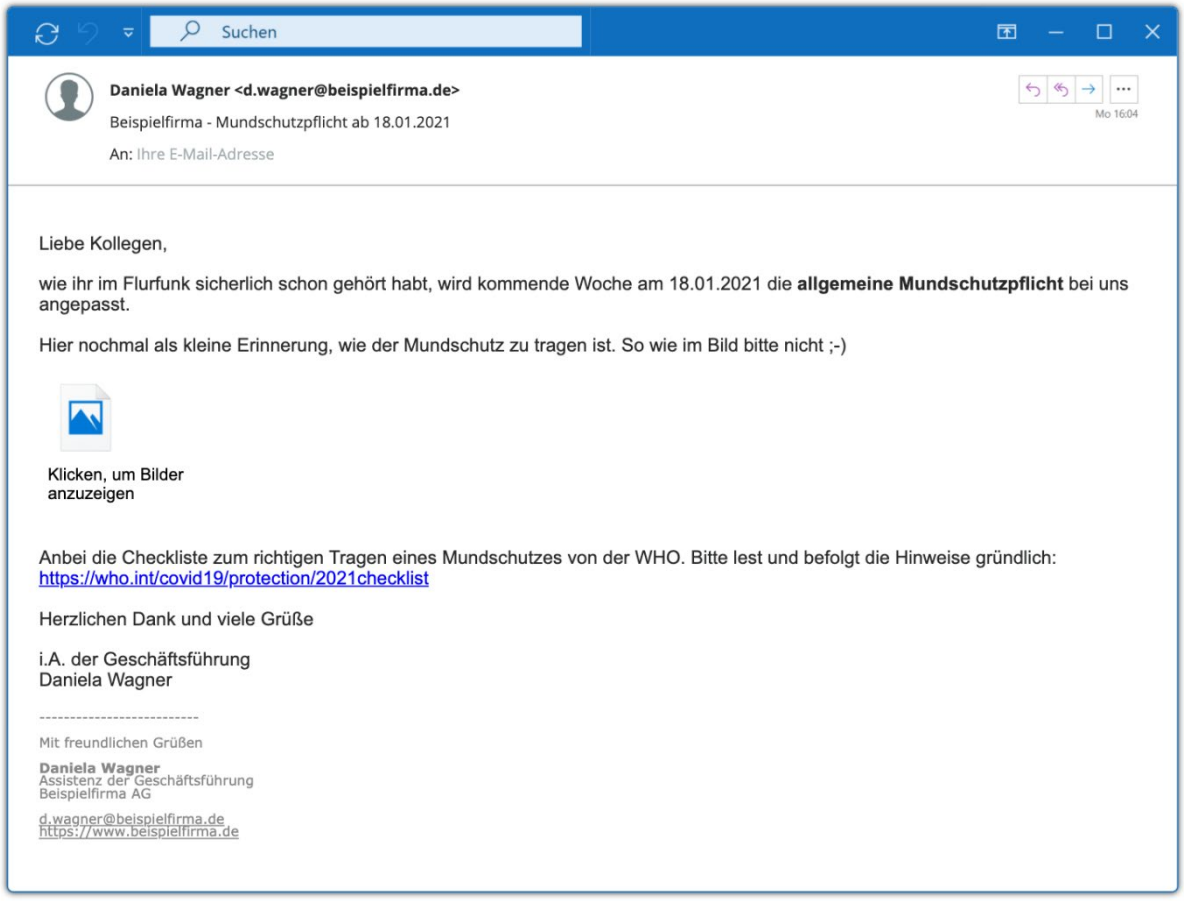


Abbildung 1: Inhalt einer E-Mail mit direktem COVID19-Bezug (entspr. Tabelle 1, E-Mail A1) mit einer Klickrate von 75,1%.

Während der Durchschnitt der Klickraten aller im o.g. Datensatz versendeten E-Mails 43,98% beträgt, weisen die beschriebenen E-Mails teilweise eine stark erhöhte Klickrate auf. Bei den in Tabelle 1 beschriebenen Phishing-Emails sind es 75,1% bzw. 50,7%. Wie auch bereits in Kapitel 1 beschrieben, ist eine mögliche Erklärung die hohe Plausibilität der Thematik für eine große Gruppe von Menschen, da viele Organisationen zum Ausbruch der Pandemie entsprechende Richtlinien eingeführt hatten bzw. Logistikunternehmen ihre Abläufe anpassen mussten. Die gesteigerte Relevanz insbesondere von A1 kann man zudem mit dem Bedürfnis nach Schutz oder Information erklären.

Mail	Betreffzeile	Klickrate
A1	[Firmenname] – Mundschutzpflicht ab [Datum]	75,1%
A2	Lieferabläufe zur Corona-Krise - Sendung 5380499815 nicht zugestellt	50,7%

Tabelle 1: Klickraten von E-Mails mit direktem COVID19-Bezug.

Während Angreifende im späteren Verlauf der Pandemie sehr gezielte Angriffsversuche entwickelten, konnten in der allerersten Phase des Ausbruchs schnell auch einfache Anpassungen an bereits bestehenden Phishing-Kampagnen beobachtet werden. Einen Hinweis zur Frage, ob eine bloße Thematisierung von COVID19 eine Steigerung des Erfolgs eines Angriffs ermöglicht, liefert daher Tabelle 2. Hier lässt sich erkennen, dass eine Betreffzeile, die den Begriff „Corona“ enthält, eine wesentlich höhere Klickrate besitzt.

Mail	Betreffzeile	Klickrate
B1	Agenda für das Meeting nächste Woche	58,8%
B2	Agenda für das Corona-Meeting nächste Woche	78,8%[10]

Tabelle 2: Klickraten von E-Mails mit und ohne COVID19-Bezug.

Im Kontext von „Home-Office“ und neuartigen Softwarelösungen ist zudem zu beobachten, dass Angriffe, die die Einführung neuer Tools thematisieren, ebenfalls eine erhöhte Klickrate aufweisen. Im Gegensatz dazu zeigen E-Mails, die beispielsweise die Verlängerung bereits genutzter Lösungen adressieren, eine eher niedrige Rate auf. Eine mögliche Schlussfolgerung ist, dass insbesondere Remote-Work-unerfahrene NutzerInnen (welche die Mehrheit darstellen[11]) hier einen erhöhten Unterstützungsbedarf aufweisen.

Dies ist insbesondere deshalb brisant, da zum derzeitigen Stand zwar einige Unternehmen bereits auf derartige Tools gewechselt haben, dennoch aber eine große Gruppe von Organisationen dies erst noch vornehmen wird. Es bleibt also zu vermuten, dass das breite Risiko auch nach Abklingen des pandemischen Geschehens bestehen bleibt. Gleichzeitig heben die Ergebnisse den Bedarf hervor, eine Anpassung der Arbeitsprozesse („Home-Office“) und die Einführung neuer Tools mit entsprechenden Sensibilisierungsmaßnahmen für die Mitarbeiter zu begleiten bzw. vorzubereiten.

Mail	Betreffzeile	Klickrate
C1	Microsoft: Bitte authentifizieren Sie Ihr Konto.	45,4%
C2	Wichtig: Verlängerung Office 365	25,2%

Tabelle3: Klickraten von E-Mails zur Einführung neuartiger Tools vs. bei bestehender Nutzung.

10 Indikatives Ergebnis, da auf kleinerer Fallzahl basierend.
11 https://www.bitkom.org/Presse/Presseinformation/Jedes-dritte-Untershynehmen-bietet-Arbeit-im-Homeshyoffice-an.html

3.2. Neuartige Kanäle

Wie bereits geschildert, haben im Zuge der COVID19-Pandemie nicht nur neuartige Themen, sondern auch weitere Kanäle zugenommen, über die Social-Engineering-Angriffe ausgeführt werden können. Die Tatsache, dass zahlreiche Angriffsversuche auf die Erlangung von Zugangsdaten für moderne Kollaborationstools abzielen, lässt vermuten, dass die Angreifenden derartige Tools künftig stärker in ihre Angriffe einbinden möchten. So ist denkbar, dass sich Angreifende innerhalb einer „Microsoft 365"-Instanz eines Unternehmens mithilfe solcher Accounts bewegen, auf Daten zugreifen und sich auch innerhalb der „Teams"-Software durch die Übernahme eines Accounts als Mitarbeitende ausgeben könnten. In diesem Kontext könnten weitere Schritte vorgenommen werden, wie der Austausch von Schadsoftware oder auch die Erlangung von Accounts mit weitergehenden Zugriffsrechten durch Social Engineering.

Es kann angenommen werden, dass die Erfolgswahrscheinlichkeit solcher Angriffe hoch ist, da viele Mitarbeitende
a) noch unerfahren mit der Nutzung der geschilderten Tools sind (auch im Kontext der in Abschnitt 3.1 geschilderten Ergebnisse, insbesondere mit Fokus auf Tabelle 2) und
b) die Umgebung eines firmeninternen Cloud-Tools als geschützten Raum wahrnehmen.

Um diese Annahmen zu bewerten, wurde im Rahmen eines Projektes innerhalb einer Organisation ein „Microsoft Teams"-basierter Bot entwickelt, der Kontakt zu den MitarbeiterInnen der entsprechenden Organisation aufnimmt und diese dazu bewegt, einen Link anzuklicken. Der Bot empfand dabei den Account eines echten Mitarbeitenden der Organisation nach. Abbildung 3 stellt eine der verwendeten Nachrichten dar.

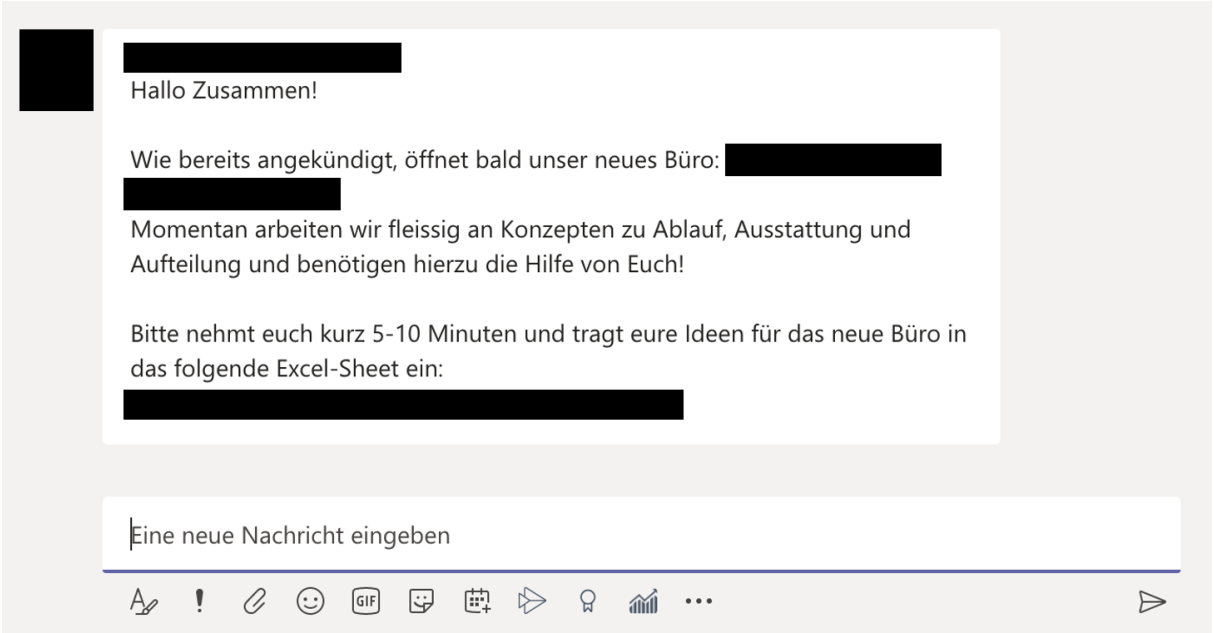


Abbildung 3: Beispielhafte Nachricht eines "Microsoft Teams"-basierten Kommunikations-Bots.

Insgesamt wurden drei verschiedene Nachrichten an 78 MitarbeiterInnen im Verlauf einer Woche versendet. Über alle Nachrichten und EmpfängerInnen hinweg ergab sich dabei eine Klickrate von 48,7%. Man kann somit darauf schließen, dass derartige Angriffe eine sehr hohe Erfolgswahrscheinlichkeit hätten, zumal es sich hier um ein verhältnismäßig einfaches Setup handelte. Durch den Einsatz weiterer Chatbot-Anwendungen könnte ein solcher Angriff auch zahlreiche Dialogebenen abbilden. Zu beachten ist zudem, dass durch bereits verfügbare KI-Modelle (auf Basis von „Natural Language Processing“) die Überzeugungskraft der Interaktion noch zu steigern wäre.

4. Fazit und Möglichkeiten zur Prävention sowie Mitigation

Die COVID19-Krise hat uns einiges abverlangt. Dem Ausbruch der Pandemie zu Beginn des Jahre 2020 folgten zahlreiche Konsequenzen – emotionaler Art aber auch hinsichtlich des Zusammenarbeitens. Angreifende und insbesondere „Social Engineers“ haben diese Situation in hohem Maße ausgenutzt und ihre Taktiken – teils minimal, aber teilweise auch fundamental – auf die neue Situation angepasst. Die Vermutung liegt also nahe, dass dieses Vorgehen darin begründet ist, dass sich die Angreifenden davon auch einen erhöhten Erfolg versprechen.

Die Ergebnisse der vorliegenden Analyse lassen vermuten, dass in der Tat Phishing-Angriffe eine erhöhte Erfolgswahrscheinlichkeit besitzen, wenn sie Themen der COVID19-Pandemie enthalten oder auch Veränderungen im Zusammenarbeiten („Home-Office“) adressieren, wie z.B. die Einführung neuartiger Kollaborationstools. Darüber hinaus lassen die vorliegenden Ergebnisse auch die Annahme zu, dass ebenfalls diese neuartigen Tools und Kanäle in starkem Maße dazu geeignet sind, für Social Engineering ausgenutzt zu werden. Ein Angriff z.B. auf Basis eines internen Messenger-Tools wird wahrscheinlich von einer großen Anzahl von Mitarbeitenden auf Anhieb nicht erkannt.

Das Gesamtrisiko, insbesondere für Angriffe, die den „Faktor Mensch“ ausnutzen, bleibt somit stark erhöht. Zum einen liegt dies an dem weiterhin erhöhten Aufkommen an Angriffen[12], zum anderen an der geschilderten erhöhten Erfolgswahrscheinlichkeit, die auch nach dem Abklingen der COVID19-Problematik in einem veränderten Arbeitskontext weiterbestehen wird.

Für Unternehmen und Organisationen stellt sich also der dringende Bedarf, auf diese veränderten Bedingungen zu reagieren. Während die initiale Phase nach Ausbruch des Infektionsgeschehens im März 2020 davon geprägt war, mobile Infrastruktur für das Arbeiten im „Home-Office“ sicherzustellen, verschiebt sich nun der Fokus noch stärker auf das Thema der Absicherung im dauerhaften „Remote-Kontext“.

Die vorliegenden Ergebnisse legen nahe, dass hierbei dem Thema der Mitarbeitersensibilisierung eine hohe Bedeutung zukommt. Die Tatsache, dass Mitarbeitende besonders stark auf Phishing-Angriffe reagieren, wenn ein unvertrautes Tool eingeführt wird, zeigt

[12] https://www.enisa.europa.eu/news/enisa-news/understanding-and-dealing-with-phishing-during-the-covid-19-pandemic

auf, dass ein Wechsel auf das mobile Arbeiten parallel auch mit entsprechenden Trainingsmaßnahmen begleitet werden sollte, um das Risiko durch Angriffe zu reduzieren. Aber auch allgemein sollte das Arbeiten im „Home-Office“ durch entsprechende Sensibilisierung unterstützt werden, um z.B. die Notwendigkeit der Verschlüsselung oder den sicheren Umgang mit sensiblen Informationen zuhause zu vermitteln. Die besondere Eignung von Messenger- oder Kollaborationstools für die erfolgreiche Durchführung von Social-Engineering-Attacken zeigt zudem auch auf, wie wichtig die entsprechende Absicherung von Cloud-Tools, z.B. durch Zweifaktorauthentifizierung ist.

Literaturhinweise

[1] Milgram, Stanley (1963). "Behavioral Study of obedience". The Journal of Abnormal and Social Psychology. 67 (4): 371–378.

[2] Kidd, C., & Hayden, B. Y. (2015). The Psychology and Neuroscience of Curiosity. Neuron, 88(3), 449–460.

[3] Sherman A. Lee, Mary C. Jobe, Amanda A. Mathis, Jeffrey A. Gibbons, Incremental validity of coronaphobia: Coronavirus anxiety explains depression, generalized anxiety, and death anxiety, Journal of Anxiety Disorders, Volume 74, 2020.

[4] https://www.bitkom.org/Presse/Presseinformation/Jedes-dritte-Untershynehmen-bietet-Arbeit-im-Homeshyoffice-an.html

[5] https://threatpost.com/cynet-the-coronavirus-is-already-taking-effect-on-cyber-security-this-is-how-cisos-should-prepare/153758/

[6] https://www.who.int/about/communications/cyber-security

[7] https://www.handelsblatt.com/technik/it-internet/it-sicherheit-der-perfekte-koeder-cyberkriminelle-nutzen-die-corona-panik/v_detail_tab_comments/25641346.html

[8] https://www.statista.com/statistics/1175657/increase-identity-theft-coronavirus-outbreak/

[9] https://www.bitkom.org/Presse/Presseinformation/Jedes-dritte-Untershynehmen-bietet-Arbeit-im-Homeshyoffice-an.html

[10] https://www.enisa.europa.eu/news/enisa-news/understanding-and-dealing-with-phishing-during-the-covid-19-pandemic

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Kein Deus ex Machina: Warum Mensch und Maschine gemeinsam Cybersecurity machen müssen

Arnold Krille[1], Jeremy Perez[1], und Stephan Schwinger[1]

Kurzfassung:
Angespornt auch durch die Erfolge der bilderkennenden und -generierenden künstlichen Intelligenzen ist auch in der Cybersecurity ein Hype um Künstliche Intelligenz entstanden. Wir wollen einen Überblick geben über die Versprechen und die Realität von KI in Cybersecurity, speziell in Netzwerksicherheit (sowohl IT als auch OT Netzwerke). Wir diskutieren mit Beispielen den „erfolgreichen" Einsatz von KI im Arsenal der Angreifer, gehen auf mögliche Gegenmaßnahmen ein und geben Beispiele für den Einsatz von KI auf Seite der Verteidiger der Netzwerksicherheit. Und wir wollen einen Einblick geben in die aktuellen Entwicklungen und Forschungen bei uns: Wie wir uns vorstellen, dass die „Intelligente Maschine" dem Administrator hilft, sein Netzwerk einfacher zu verstehen und zu überwachen, sodass er den Überblick mit dem richtigen Kontext behält, bei Incidents nicht durch unnötige Informationen abgelenkt wird und auch bei Änderungen im Unternehmen und dem Unternehmensnetzwerk mit intelligenter Assistenz zielgerichtet agieren kann. Diesen Hilfsmitteln muss der Administrator natürlich vertrauen können, weshalb es für uns elementar ist, dass Diese nachvollziehbar und erklärbar sind und etwaige finale Entscheidungen immer in der Zusammenarbeit von Mensch und Maschine getroffen werden.

Stichworte: Antivirus und KI, Botnetzerkennung und KI, KI, Künstliche Intelligenz, Netzwerksicherheit

1. Über KI in Cybersecurity

1.1. Das Versprechen

Die Digitalisierung schreitet in allen Lebens- und Arbeitsbereichen unaufhörlich voran. Neue Technologien und Techniken ermöglichen mehr Automatisierung, mehr Konnektivität, mehr Austausch, mehr Zusammenarbeit, mehr Produktivität und mehr Optimierung als jemals zuvor. Eine wahre Explosion der Anzahl der digitalisierten Prozesse und Vorgänge führt zu einer Vielzahl an neuen Möglichkeiten, komplett neuen und teils disruptiven Geschäftsfeldern und zu einer allgemeinen Hoffnung auf eine bessere Zukunft für alle. Mit dieser Vermehrung der digitalen und „connected" Geräte, digitalen Prozessen und dem Datenaustausch weit über lokale Netzwerke und Organisationen hinaus steigt leider auch die digitale Verwundbarkeit. Die sprunghaft steigende Anzahl digital verbundener Geräte verursacht einen gleichermaßen steigenden Aufwand bei Pflege und Administration des Netzwerkes und bringt einen zugehörigen Anstieg der Sicherheitslücken und der dadurch nötigen Sicherheitsupdates mit sich. Neue Kollaborationen und Arten des Datenaustausch erfordern eine Öffnung der eigenen Netze zu anderen Kommunikationspartnern, vielfach sowohl für eingehende als auch für ausgehende Verbindungen. Auch die Anzahl der Verbindungen und die Menge der übertragenen und damit zu überwachenden Daten

[1] genua GmbH

steigt damit massiv. Dies macht es zunehmend schwerer, den rechtmäßigen Netzwerkverkehr von den bösartigen Attacken und unerwünschten Datenflüssen zu unterscheiden und stellt die IT-Abteilungen vor immer größere Herausforderungen, um den sicheren Betrieb der Netzwerke und damit der Unternehmensprozesse sicherzustellen. Dabei ist sowohl die Sicherheit im Sinne der Ausfallsicherheit, als auch des Schutzes gegen bewusste Attacken, Manipulationen und Datendiebstähle im Fokus. Ähnlich wie die Automatisierung in der industriellen Fertigung eine Steigerung der Produktivität hervorbrachte, ist auch in der IT-Sicherheit mit Automatisierung eine gesteigerte Effektivität erreichbar.

Medienwirksam waren in den letzten Jahren die Potenziale und Optimierungen, die mit Machine Learning und unter dem Stichwort „Künstliche Intelligenz“ in vielen Prozessen und Anwendungsfällen erreichbar sind oder zumindest in Aussicht stehen. Dabei werten Algorithmen große Mengen Daten aus und „erkennen“ manche Korrelationen, die für einen Menschen nicht sichtbar sind, oftmals überraschend für den Menschen. Von diesen Vorteilen soll natürlich auch die Cybersecurity profitieren, versprechen diese Methoden doch, noch mehr Datenverkehr noch besser überwachen zu können und sowohl bekannte Angriffe und Probleme frühzeitiger zu erkennen, aber auch komplett neue Angriffsmuster zu erkennen und automatisch zu unterbinden. Dadurch werden die IT-(Security-)Abteilungen befähigt, bei gleichzeitig stark gestiegenen Nutzungen und Möglichkeiten der digitalen Prozesse mit gleichem oder gar geringerem Aufwand als bisher einen besseren Schutz zu erbringen.

So sieht Gartner die Anwendung von KI auf Probleme der Cybersecurity als Top-10-Trend[2], und über zwei Drittel aller in einer Studie befragten Organisationen erwarten, ohne die Verwendung von Künstlicher Intelligenz Cyberattacken wehrlos gegenüberzustehen[RT19].

Bisher allerdings kann dem gestiegenen Aufwand in der IT-Security nur mit gesteigerten Aufwänden an Personal begegnet werden, wobei ein Großteil der Unternehmen einen Mangel an qualifizierten IT-Security-Experten feststellt. Außerdem erfordern die steigenden Komplexitäten sowohl in den zu schützenden Netzwerken, bei der Erkennung der verschleierten Angriffe als auch bei der Wahrung der gesetzlichen Vorgaben und einzuhaltenden Richtlinien auch eine immer höhere Qualifikation des Personals. Ein Ausbau dieser Fähigkeiten wird durch die bereits hohe Arbeitslast erschwert[3][4]. Häufig wird erwartet, dass in naher Zukunft eine Integration künstlicher Intelligenz in Cyberverteidigungssysteme eine schnellere und effektivere Entscheidung durch Menschen ermöglicht oder sogar autonome Cyberagenten Netzwerkangriffe erkennen und bekämpfen [DR20]. Eine weitverbreitete Ansicht[5] ist, dass KI-Technologien größere Datenmengen verarbeiten und daher schneller mehr Angriffe erkennen können als Menschen. Vor übertriebenen Erwartungen wird an derselbe Stelle wegen der Nichtausgereiftheit dieser Technologien aber ebenfalls gewarnt und empfohlen, kosteneffek-

[2] https://www.computerweekly.com/opinion/Gartner-Top-10-strategic-technology-trends-in-2020
[3] https://www.computerwoche.de/a/personalmangel-bei-der-cyber-abwehr,3547703
[4] https://www.datensicherheit.de/it-sicherheit-budget-personalmangel-studie
[5] https://www.gartner.com/smarterwithgartner/5-questions-to-cut-through-the-ai-security-hype/

tivere, einfachere Lösungen nicht zu vergessen.

1.2. Die Probleme der KI in der echten Welt (mit Beispielen)

Machine Learning-basierte Ansätze in der Cybersecurity scheitern oft an der Realität und Komplexität des Netzwerkverkehrs und moderner Software.

Ein prominentes Beispiel ist die von Zetter [Zet19], [AA19] in dem Antivirusprogramm Cylance gefundene Schwachstelle. Die Entwickler dieser KI-gestützten Erkennung waren zur Senkung ihrer Fehlalarmquote gezwungen, ein explizites Whitelisting in die finale Bewertung einzubauen. Dies konnte ausgenutzt werden, um durch Anhängen entsprechender Strings an die Malware die Erkennung zu umgehen und den Schadcode am Antivirusprogramm vorbeizuschleusen.

Oft sind schon die gewählten Features, die das vorgelegte Sample beschreiben und als Eingabeparameter dem Klassifikator vorliegen, sehr sensibel gegenüber Manipulationen. Auch wenn es sich dabei nicht um ein Produktivsystem handelte, konnte dies beispielsweise von Fleshman [Fle19] ausgenutzt werden, um durch Manipulation nicht ausführbarer Programmbestandteile ohne Einschränkung der Schadfunktionalität mehrere KI-gestützte Erkennungsmethoden gleichzeitig zu täuschen. Um eine geeignete Manipulation jedes Malwaresamples zu finden, die die daraus extrahierten Features gutartig erschienen ließen, benötigte er für diesen Angriff allerdings Zugriff auf die zugrundeliegenden KI-Modelle.

In ähnlicher Weise untersuchen Xu et al.[WXE16] erfolgreich die Möglichkeit, Malwareklassifikatoren für PDF- Dokumente durch Variation der Testdaten zu täuschen. Für einen der untersuchten Klassifikatoren zeigen sie in ihrer Untersuchung sogar, dass dieser unter den gewählten Features von denjenigen dominiert wird, die Artefakte der Malware-Trainingsamples, wie beispielsweise das Fehlen eingebetteter Fonts, hervorheben. Es handelt sich dabei also um Features, die manipulierbar sind, ohne die gewünschte schädliche Wirksamkeit zu mindern. Zur Verhinderung des Lernens auf ungeeigneten Eigenschaften ist es notwendig, beim Erstellen von KI-Modellen Expertenwissen hinzuzuziehen.

Ähnliche Probleme können anhand der über Firewallsysteme gewonnenen Erfahrungen für HTTP- und Email-Verkehr antizipiert werden, da hier bereits klassische, nicht auf Machine Learning basierende Systeme an Abweichungen in der Implementation und in der Interpretation der entsprechenden Protokolle scheitern [Ull17], [Ull15].

Die Erfahrung zeigt [Ull20], dass der Mangel an Expertenwissen oft dazu führt, Features zu verwenden, die sehr spezifisch für die zum Training benutzten Daten sind. Dazu zählen zum Beispiel User-Agents oder XMailer, die bestimmte Spamkampagnen dominierten und dabei einen unberücksichtigten Bias in die Daten einbrachten.

Neben der Realitätsnähe der Trainingsdaten hängt Qualität der trainierten Modelle stark von der in diesen Daten enthaltenen Variabilität ab. Sanders et al. [HS17] argumentieren, dass URL-Klassifikatoren zur Erkennung von DGAs, also URLs, die von bösartiger Software zur Kommunikation mit ihrer Kommandoinfrastruktur aufgerufen werden, aufgrund anderer Ausrichtung des Trainingsdatensatzes versagen können bzw.

unter Overfitting auf spezielle Datensätze leiden.

Generell ist Overfitting ein Problem und bei der Konstruktion von Machine-Learning-Klassifikatoren unbedingt zu vermeiden. Neben dem Training auf zu spezifischen Datensätzen kann auch die Auswahl einer zu großen Menge an Eingabefeatures zu einer Verschlechterung der Vorhersagequalität führen, wie es in [SBE+17] für die Angriffskategorisierung mittels Random Forests demonstriert wird.

Mathas et al. [MSX+18] untersuchten das Open-Source-Tool Apache Spot auf seine Fähigkeiten, einfache Angriffe zu erkennen. Spots KI basiert auf Methoden der Sprachverarbeitung und war nicht in der Lage, die Präsenz von DNS-Tunneln zu erkennen. Auch scheiterte es an der Erkennung einer Flooding-Attacke, da das trainierte Modell sich zu schnell an diese Anomalie anpasste.

Es ist in unseren Augen ein Fehler, beim Nutzer die Erwartung einer inhärenten Zuverlässigkeit der aktuellen KI- gestützten Erkennungsmethoden zu wecken und ihn in einem falschen Gefühl der Sicherheit zu wiegen.

1.3. Die Probleme in der Wissenschaft

Die gesamte menschliche Geschichte wird durch unzählige Beispiele gekennzeichnet, die aufzeigen, wie der Mensch stets den Anspruch hegte, „sichere“ Technologien und Systeme zu entwerfen, die unvermeidlich auf die eine oder andere Weise kompromittiert wurden. Angeblich „unknackbare“ biometrische Verfahren wurden ausgetrickst, Betrugsdetektionssystem umgangen und Unfälle, einschließlich tödlicher, die durch Software oder Industrieroboter verursacht wurden, werden leider immer wieder beobachtet[YS16]. Diese fundamentalen Probleme unterscheiden sich vom Versagen von KI-Systemen insoweit, als diese keine direkte Folge der vermeintlichen Intelligenz des Systems selbst sind. Als Folge dessen treten bei KI-Modellen die bei klassischen Systemen bestehenden Probleme und zusätzlich dazu die KI-spezifische Schwierigkeiten auf. Diese KI-spezifischen Fallstricke, die sich zu katastrophalen Fehlern entwickeln können, haben ihre Ursache oft schon in der Lernphase des KI-Modells. KI-Modelle sind unweigerlich von der Qualität des zum Training verwendeten Datensatzes abhängig, der umgangssprachliche Satz „Garbage In, Garbage Out“ gilt auch hier.

Die Qualität eines KI-Modells wird durch gänzlich unterschiedliche Kriterien bestimmt, die wichtige Fragen hinsichtlich der Integrität des Datensatzes beantworten:

- Inwieweit repräsentiert der genutzte Datensatz die vorherzusagende „Realität“?
- Ist dieser Datensatz eventuell unvollständig?
- Widersprechen sich Daten selbst oder sind die Daten kohärent?

Fehler in der Bereinigung dieser und vieler weiterer Punkte können zu falschen Schlussfolgerungen und folglich zum gravierenden Versagen des Modells führen. Eine weitere Komplikation tritt auf, wenn KI-Modelle problembedingt selbst an Komplexität gewinnen müssen, da diese besonders „datenhungrig“ werden. Sie erfordern eine Datenmenge, die unter bestimmten Umständen nicht zur Verfügung steht.

Ein weiteres, tief liegendes Problem bei der Erforschung und Untersuchung Machine-

Learning-basierter Ansätze in der Cybersecurity liegt im Mangel an brauchbaren Trainings- und Testdaten [TL20], [RWS+19]. So ergibt die Suche nach „kdd 99 ids“ auf www.semanticscholar.org für das Jahr 2019, also 20 Jahre nach Schaffung des KDD-99-Datensatzes, noch immer 191 aktuelle Treffer auf Intrusion-Detection-bezogene Paper, die diesen Datensatz verwenden oder analysieren. Häufig verhindern Datenschutzbedenken das Teilen von in realen Netzwerken aufgezeichneten Daten, auf denen Modelle des Machine Learning trainiert wurden. Dies schränkt die Möglichkeiten zur Reproduktion der Ergebnisse erheblich ein und erschwert den Vergleich verschiedener KI-Ansätze.

Über die mangelnde Aktualität hinaus lässt sich feststellen, dass die öffentlich zugänglichen Traffic-Mitschnitte unter folgenden Problemen leiden:

- Der aufgezeichnete Traffic-Mix weist Disbalancen zwischen Angriffs- und Hintergrundverkehr auf.
- Reale Daten sind nicht ausreichend gelabelt [RWG+17].
- Der Datensatz enthält duplizierte oder redundante Daten.
- Die Anzahl der enthaltenen, verschiedenen Angriffe ist gering.
- Die Daten sind zu spezifisch für die beschriebenen Angriffsarten und damit schlecht auf andere Angriffsarten verallgemeinerbar.
- Die Größe des Datensatzes ist sehr gering und damit für einige Ansätze des Machine Learning unbrauchba.r
- Der Datensatz ist nicht unmittelbar nutzbar, da Einträge aus verschiedenen Log-Quellen nicht korrekt synchronisiert sind[6].

Eine Querschnittsstudie mit dem Fokus auf Fallstricken in akademischen Anwendungen von Machine Learning auf Probleme der Cybersecurity findet sich in [AQP+20]. Die Autoren definieren 10 Klassen von Fehlern entlang des gesamten Machine-Learning-Workflows und erkennen, dass in jedem untersuchten Paper wenigstens drei davon begangen werden. Dabei dominieren Fehler bei der Datenerhebung: Zum einen repräsentiert die Zusammensetzung der zum Training verwendeten Daten oft nicht die Wirklichkeit, zum anderen werden Eigenschaften der Daten zum Training verwendet, die in der Realität nicht ermittelbar sind, bzw. die Grenzen zu den Validierungsdaten verwischen. Weitere häufige Fehler werden in der Verwendung ungeeigneter Baselines, in der fehlenden Validierung der Experimente an realen Systemen und im unterlassenen Härten der Modelle gegen Angriffe erkannt.

Da Angriffe immer ausgefeilter werden, kann Bias bei KI-Modellen ernsthafte Probleme für die Cybersicherheit verursachen. [FP19] argumentieren, dass vor allem zwei weit verbreitete Bias-Formen auftreten: spatial bias, also räumlicher Bias und temporal bias, also zeitlicher Bias. Räumlicher Bias taucht auf, wenn Trainings- und Testdatenverteilungen nicht oder nur unzureichend realistische Gegebenheiten abbilden. Weiterhin operieren KI-Modelle auf sich kontinuierlich ändernden Daten. Im Kontext der Malwareerkennung bestehen zeitliche Änderungen zum Beispiel sowohl aus Weiterentwicklungen bestehender Malware als auch gänzlich neuen Varianten. Daher ist die

[6] Dies wurde beispielsweise für CDX-2009 durch die Autoren von [HBC+13] festgestellt

zeitliche Konsistenz für die Bewertung der Performance eines Modells von großer Bedeutung. Verzerrungen der Ergebnisse können auftreten, wenn der Versuchsaufbau es dem Modell ermöglicht, auf effektiv zukünftiges Wissen zu trainieren, auch zeitlicher bias genannt, oder wenn unvorteilhafte Zeitfenster eingestellt werden. Das Trainieren eines Modells auf Daten, die in der Praxis nicht verfügbar sind – auch bekannt als „data snooping“ – ist nicht ausschließlich auf zeitliche Aspekte beschränkt, sondern kann auf viele Arten auftreten, von denen einige subtil und schwer zu identifizieren sind [AQP+20].

Die Autoren von [FP19] vermuten, dass das Auftreten und die Verbreitung dieser Probleme auf zwei Ursachen zurückzuführen ist: mangelnde Informationen für die Untersuchung des Bias und die anfallende Komplexität bei Berücksichtigung der Zeit bei der Auswertung der Qualität der Modelle.

1.4. Das Problem mit dem Vertrauen

Die Rechtfertigung des Vertrauens in KI wird in [TMF19] diskutiert. Mit „[T]rust in AI [...] to deliver cybersecurity tasks is a double-edged sword“, heben die Autoren hervor, dass KI-Systeme zwar neue Tool sind, aber Angriffe auf diese Tools sowohl möglich als auch schlecht zu erkennen sind, da diese sich nur schwer analysieren lassen und ihre Robustheit sich nicht sicherstellen lässt – insbesondere ließen sich aus Aufzeichnungen über vergangenes Verhalten dieser Systeme keine Robustheitsaussagen für die Zukunft ableiten. Ihrer Meinung nach reichen wegen dieser fehlenden Zugänglichkeit einer Risikoanalyse staatliche Anstrengungen zur Zertifizierung und Standardisierung von Cybersecurity und KI-Systemen nicht aus. Um sie dennoch einsetzen zu können, fordern die Autoren unter anderem ein kontinuierliches Monitoring dieser Systeme. Warum sollten wir nun nicht einfach KI-gestützten (prediktiven) Systemen vertrauen und dabei außer Acht lassen, warum und vor allem wie diese ihre Entscheidungen fällen? Man könnte der Meinung sein, dass diese beiden Fragen keine oder eine mindere Rolle spielten, solange die Performance für einen Testdatensatz gut ausfielen.

Zunächst fällt natürlich auf, dass KIs ähnlich wie Menschen selbst stets ein unvollständiges, vereinfachtes oder manchmal gar gänzlich verzerrtes Bild der „Realität“ abbilden. Dies liegt unter anderem daran, dass Mensch und Maschine im Laufe des Lernprozesses stets nur auf mehr oder weniger umfangreiche Stichproben zurückgreifen können. Diesen grundlegenden Aspekt strebt man zu mindern an, kann ihn jedoch nicht völlig beseitigen. Daher kann und sollte man immer von der Annahme ausgehen, dass KI-Modelle nicht unfehlbar sind. Umgebungen mit geringem Risiko können sich mit dem Auftreten von gelegentlichen Fehlern abfinden. Gänzlich anders sieht es in kritischen Bereichen wie u.a. dem Feld der Cybersecurity aus, wo vereinzelte Fehler schwerwiegende Konsequenzen implizieren können. In diesem Zusammenhang spielt das Verständnis und eine mögliche Interpretierbarkeit der eingesetzten Modelle eine ausschlaggebende Rolle, denn sie erlaubt es mehr über das eigentliche Problem, die Daten und gar im Falle eines Modellausfalls über den Grund des Ausfall zu erfahren. Da KI-Systeme im Feld der Cybersecurity in prekären Umgebungen eingesetzt werden, ist es von größter Bedeutung sicherzustellen, dass der Entscheidungsprozess ver-

trauenswürdig ist, insbesondere in neuen oder gar kontroversen Szenarien. Hierbei erforderlich sind eine klare Definition von Performance-Metriken und die Entwicklung von Algorithmen, die Erklärungen und Zuverlässigkeit und damit einhergehend Rechenschaftspflicht von KI-Systemen garantieren. Da Probleme, die heutzutage von KIs gelöst werden sollen, insgesamt an Komplexität zulegen, werden die hinzugezogenen Modelle selbst vielschichtiger und komplexer. Diese intrinsische Komplexität führt zu Undurchsichtigkeit und schließlich zu sogenannten Black-Box-Modellen, deren innerer Aufbau zwar bekannt ist, die jedoch zu keinen oder wenigen Erkenntnissen hinsichtlich des Gesamtverhaltens führen. Dadurch sind solche Modelle zwei der bereits erwähnten Hauptprobleme unterworfen: einer Anfälligkeit für widersprüchliche Daten bzw. einer mangelnden Robustheit und einer Tendenz, sich der menschlichen Interpretation zu entziehen. Es gibt zwar Ansätze, die Entscheidungsgarantien unter Verwendung einer Vielzahl von Techniken bieten, jedoch weisen diese Einschränkungen auf und/oder führen zu signifikanten Performanceproblemen des Gesamtsystems [MNO19]. In diesem Zusammenhang sind bereits empirische und theoretische Beweise formuliert worden, die aufzeigen konnten, dass beide Faktoren tatsächlich miteinander verbunden sind [NADL19].

Des Weiteren ist nicht nur die Koordination zwischen KI und bestehenden Cybersecurity-Systemen wichtig, sondern auch die Koordination und insbesondere das Vertrauen zwischen Mensch und KI von entscheidender Bedeutung. Als Reaktion auf die in diesem Abschnitt bereits erläuterten Herausforderungen erfordert Cybersecurity hybride Ansätze, welche die exzeptionellen Fähigkeiten von Menschen und KI nutzen und orchestrieren. Um eine erfolgreiche Mensch-KI-Symbiose zu erreichen, müssen Schnittstellen zur KI für eine optimale Entscheidungsfindung so beschaffen sein, dass Menschen die gelieferten Ergebnisse nachvollziehen und ihnen vertrauen können. Einerseits ist die Schulung menschlicher Bediener, beispielsweise zur Abwehr von Datenfälschungsangriffen unumgänglich, andererseits benötigt man KI-Modelle, die mögliche Fehlermodi vorhersagen und sich anpassen können, wenn selbst Menschen Fehlentscheidungen treffen [MLM+20].

1.5. Hohe Kosten / Hohes Risiko durch KI

Das Vertrauen in die Zuverlässigkeit und Robustheit der KI mag im ersten Moment abstrakt erscheinen, kann vielfach aber in eine recht konkrete Zahl, meist einen finanziellen Schaden umgerechnet bzw. ausgedrückt werden. Man darf nicht vergessen, dass die Computernetzwerke in Organisationen nicht als Selbstzweck existieren, sondern dem Unternehmensziel dienen und die Unternehmensprozesse und Abläufe unterstützen. Eine Beeinträchtigung des Netzwerks geht also fast immer mit einer Beeinträchtigung des Unternehmenserfolges einher.

Da es das perfekt sichere Netzwerk nicht gibt, wird in allen Organisationen mehr oder weniger bewusst eine Abwägung zwischen dem Aufwand der Absicherung und dem Nutzen bzw. dem verhinderten Schaden durchgeführt. Zertifizierungen und Frameworks wie etwa die ISO 27001, IEC 62443 oder der Einsatz von Informations-Sicherheits-Management-Systemen (ISMS), aber auch Regulierungen und Vorgaben des Gesetzgebers wie das aktuelle IT-Sicherheitsgesetz 2.0 helfen diese Abschätzung

auf einer soliden und nachvollziehbaren Grundlage zu treffen und die wichtigen Prozesse des Unternehmens und damit die wichtigen IT-/OT-Systeme und -Netzwerke zu bestimmen und adäquat abzusichern.

Zur Beurteilung der eingesetzten Technologien, also auch von eingesetzten KI-Systemen, spielt die Zuverlässigkeit im Finden (und evtl. Verhindern) von Bedrohungen eine Rolle, aber auch die Zuverlässigkeit im „Nicht-finden" von Nicht-Bedrohungen ist entscheidend. Das „Nichterkennen" eines Angriffs wird dabei meist durch den Einsatz multipler verschiedener Systeme mitigiert, sodass letztlich die Rate der gefundenen tatsächlichen Angriffe gegen die Kosten für Anschaffung und Betrieb eines solchen Systems abgewogen werden. Die Kosten von Fehlalarmen, also Reaktionen auf vermeintliche Angriffe aber verursachen je nach Reaktion einen unnötig eingetretenen Schaden in der Produktivität der Organisation.

Dass die Literatur bisher keine belegbaren oder gar „spektakulären" Zahlen für Schäden durch Fehlalarme beim Einsatz von KI-Lösungen in der Cybersecurity zu bieten scheint, legt nahe, dass einerseits diese Werkzeuge noch nicht so zentral eingesetzt werden, wie etwa die klassischen Lösungen Antivirus und Firewalls und andererseits KI-Lösungen wohl nur mit passivem Einsatz und einer Alarmierung des Nutzers implementiert sind und eine automatische Reaktion bisher nicht gewünscht ist. Zu tief sitzt in vielen IT-Abteilungen und Entscheidergremien noch die Erinnerung an den Werdegang, den die vielgelobten Intrusion-Prevention-Systeme durchlebt haben. Auch in diesen Systemen sind die Fehlalarme in Kombination mit der harten Reaktion ein Problem und sorgen meist für die „Degradierung" zum passiven Intrusion Detection System. Der schon erwähnte Report der Capgemini [RT19] unterstützt dies, da die Umfrageteilnehmer zu 88% eine mittlere oder hohe Verwendung von KI-Systemen zur Detektion, aber nur 65% eine mittlere und hohe Verwendung zur Response bescheinigen. Leider werden Art und Konsequenz der Response nicht analysiert.

Eine gute Strategie ist es also, die Versprechen der KI zu nutzen, aber durch den Nutzer als Experten prüfen zu lassen. Das Zusammenspiel zwischen dem Nutzer und dem Algorithmus kombiniert dann die Vorteile beider Seiten.

2. Beispiele, wie die Angreifer KI nutzen

Der Einsatz von KI-Methoden ist für den Angreifer möglich und attraktiv, da er so sein Angriffspotenzial vervielfachen und die Anzahl angegriffener Systeme nach oben skalieren kann. Die potenzielle Verletzbarkeit dieser Algorithmen haben wir bereits in Abschnitt 1.2 diskutiert, insbesondere auf die Mutation von Malware, um diese an Firewalls vorbeizuschleusen. In ähnlicher Weise lassen sich die Texte von Phishing Mails maschinell optimieren, um Mailfilter zu täuschen [PM15].

Phishing-URLs lassen sich durch Methoden der künstlichen Intelligenz so wählen, dass automatische URL-Klassifikatoren sie schwerer erkennen [BTCV18].

Neben der delivery gibt es entlang der cyber kill chain noch weitere Phasen, in denen ein Angreifer vom Einsatz von KI profitieren kann:

Ein wesentliches Werkzeug moderner Cyberkrimineller ist das sogenannte social en-

gineering, das heißt die geschickte Beeinflussung potenzieller Opfer, um ein Eindringen in ein System zu ermöglichen oder zu begünstigen. So wurden möglicherweise bereits Deep Fakes zur Nachahmung von Stimmen eingesetzt, um durch sogenannten CEO Fraud[7] hohe Geldzahlungen zu veranlassen.

Ähnlich zu klassischen Verfälschungsangriffen wie ARP-Spoofing, die die Netzwerkarchitektur anvisieren, kann auch ein auf Machine Learning basierendes IDS attackiert werden. Dies geschieht beispielsweise durch graduelle Verschiebung der Entscheidungsgrenze der zugrundeliegenden Klassifikatoren durch explizite Generierung und Injektion von Grenzfällen während ihres Trainings [LLZ+18]. Ebenso ist es möglich, eine angreifende KI bewusst gegen ein verteidigendes KI-System zu trainieren. Dies verlangt aber normalerweise Zugang zu dem verteidigenden System.

Generative neuronale Netzwerke lassen sich verwenden, um die Kommandoinfrastruktur von Botnets zu verbergen, indem sie die zur Kommunikation nötigen Domainnamen erzeugen (DGAs) [AWF16]. Experimente deuten darauf hin, dass die Vielfalt in den so erzeugten Domainnamen ein effektives Training von Detektoren verhindern kann[CLH19].

Neben der bereits diskutierten Veränderung von file patterns ist auch die Veränderung von Kommunikationsmustern (z.B. auf Paket- und Timingebene) eine für den Angreifer interessante Option, um im Netzwerk unerkannt zu bleiben. Es gibt Hinweise auf kursierende Malware, die den beobachteten Netzwerkverkehr nachahmen, um ihre Entdeckung zu erschweren[8]. Weiterhin wird in [RG18] eine Malware derart modifiziert, dass sie ihre C2-Kommunikation mit Hilfe eines neuronalen Netzes mutiert. Das so erhaltene Trafficprofil ist dann auf Flow-Ebene nicht mehr von Facebooktraffic unterscheidbar und wird von einem IPS nicht gestoppt. In solch einem Fall spielt ein DPI- und verhaltensbasiertes System wie der cognitix Threat Defender seine Vorteile aus: die Art des Verkehrs wird durch Einbeziehen der Applikationsebene erheblich genauer ermittelt, und die Auswertung historischer Daten über die (Nicht-)Benutzung einer bestimmten Anwendung durch ein bestimmtes Gerät kann ein Blockieren des Verbindungsaufbaues auslösen.

Für die nahe Zukunft können aber selbst-optimierende „Next Generation Threats“ erwartet werden [TZ19], die autonome Entscheidungen hinsichtlich ihres Vorgehens oder relevanter Ziele fällen, ohne dabei eine auffällige C2- Infrastruktur zu benötigen

Auch für zielgerichtete Attacken ergeben sich dabei neue Möglichkeiten: So konnte in [DK18] eine Malware konstruiert und in ein Videokonferenzsystem eingeschleust werden, die sich nur bei Erkennen eines bestimmten Gesichtes aktiviert und daher sehr spezifische Zielsysteme schädigen kann.

3. Wie können Verteidiger KI nutzen?

Wenn die Möglichkeiten von Machine Learning und „künstlicher Intelligenz“ von Angreifern auf IT-Systeme recht einfach und effektiv genutzt werden können, welche

[7] https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402
[8] https://www.wsj.com/articles/artificial-intelligence-transforms-hacker-arsenal-1510763929

Möglichkeiten bieten sich dann auf der Seite der Verteidiger?

3.1. Automatisierung in Antivirus

Die Erkennung von Schadcode mittels Antivirussoftware erfolgt hauptsächlich über das Finden markanter Binärmuster (auch als Fingerabdruck bezeichnet) in den zu untersuchenden Dateien. Dabei bringt die Antiviruslösung eine regelmäßig aktualisierte Liste von Mustern oder Checksums bekannter bösartiger Dateien mit und prüft die Dateien im lokalen Speicher/Netzwerk unentwegt auf Übereinstimmung mit diesen.

Angreifer können damit durch kleinste Variationen ihrer Binärdateien diesem simplen Matching entgehen, wie etwa Fleshman et al. [FRZ+18] zeigen. Der Hersteller der Antivirussoftware muss diese neuen Variationen aufwendig analysieren und dann als Variation einer bekannten Malware seiner Datenbank hinzufügen.

Methoden des Machine Learning sind gut darin, Ähnlichkeiten zu erkennen. Diese Methoden werden von Herstellern von Antiviruslösungen verwendet, um neue Varianten bekannter Malware zu erkennen und ihrem Ursprung zuzuordnen. Diese automatisierte Klassifizierung und Zuordnung wird dann von den Analysten geprüft und bestätigt (oder eben verworfen), reduziert aber für Variationen bekannter Malware den Aufwand der Analyse und ermöglicht damit eine schnellere Bereitstellung von passenden Mustern an die Kunden. Da aber diese Unschärfe in der Erkennung von Malwarevariationen auch zu Falschmeldungen führen kann, sollten die Ergebnisse auf jeden Fall durch Analysten geprüft werden. Damit ergibt sich also immer noch eine gewisse Zeitspanne bis Variationen von bekannter Malware auch im Praxiseinsatz erkannt und bekämpft werden.

Verschiedene Hersteller von Antiviruslösungen werben auch mit KI-Methoden zur Erkennung völlig neuer Schadsoftware. Dabei werden Heuristiken und Methoden des Machine Learning in Ergänzung zu den traditionellen Mustern in der lokal betriebenen Instanz des Anti-Virus eingesetzt. Wie gut diese Mechanismen funktionieren, scheint allerdings fraglich angesichts der immer wieder über die IT-Landschaft ziehenden Wellen von neuen Bedrohungen, welche erfolgreich Zero-Day Attacken nutzen.

Es bleibt letztlich ein Wettlauf zwischen den Angreifern und den Antivirusherstellern mit dem Verteilen der Antivirus- Patterns.

3.2. Angriffe erkennen aus logfiles über viele Netzwerke (verschiedener Kunden) hinweg

Da es müßig ist, die Logdateien jedes einzelnen Gerätes nur auf dem jeweiligen Gerät untersuchen zu können, haben sich zentrale Security Incident and Event Management Systeme (kurz SIEM) als Erweiterung einer zentralen Logserverlösung schon länger durchgesetzt. Auch auf diesen Datenbeständen kann mit Methoden des Machine Learning nach typischen Gruppen von gemeinsam auftretenden Ereignissen gesucht werden. Da die Logauswertung damit über die Grenzen eines einzelnen Gerätes oder Systems hinweg erfolgt, können also auch Ereignisse verschiedener Systeme korreliert werden. Auch bei dieser Datenauswertung ergeben sich bessere und verlässlichere Ergebnisse je größer die Menge der verfügbaren Daten ist.

Speziell in größeren Organisationen sind sowohl die dafür nötigen technischen Ressourcen vorhanden als auch die nötige Anzahl und Expertise der Mitarbeiter. Der Aufwand, ein solches System gewinnbringend mit Machine-Learning-Methoden einzusetzen, sollte nicht unterschätzt werden [Bou20], macht er doch den effektiven und kosteneffizienten Einsatz dieser Methoden für kleine und mittlere Organisationsgrößen schnell unerreichbar.

Eine Reihe Anbieter hat dies erkannt und bietet zentrale SIEM-Lösungen als Cloud-Lösung an. Damit können die verwendeten Algorithmen auch über Organisationen hinweg Daten auswerten, es erfolgt also implizit ein gegenseitiger Austausch von Informationen über die Bedrohungslage. Allerdings ist bei diesen Lösungen die Einsicht wie der Algorithmus zu seiner Bewertung kommt noch weniger möglich als bei lokalen on-premise Lösungen. Oftmals ist dieses Wissen auch das Geheimnis des jeweiligen Anbieters**.**

Auch ein ungewollter Datenaustausch zu Dritten kann bei diesen Lösungen nicht ausgeschlossen werden. Das können offensichtliche Datenlecks mit Vollzugriff auf die Datenbank des Anbieters sein, die Risikobewertung basiert in diesen Fällen „lediglich“ auf dem Vertrauen gegenüber dem Anbieter. Es können aber auch Datenlecks auftreten, wo aus der Ausgabe eines Algorithmus auf die Trainingseingabe geschlossen werden kann [FJR15]. Da speziell in logfiles auch regelmäßig persönliche Informationen erscheinen, ist aber in beiden Fällen der Weg zur Verletzung des Datenschutzes nicht weit.

3.3. Erkennung von Botnetzen

Botnets sind Netzwerke infizierter Maschinen, die genutzt werden, um konzertiert Schadaktionen im Internet, wie beispielsweise DDoS-Attacken, durchzuführen. Sowohl die Erkennung infizierter Maschinen als auch die Identifikation der beteiligten Kommandoinfrastruktur (sogenannter C2-Server) sind Ziele, die mittels Machine Learning unterstützt werden können. Für letzteres ist DISCLOSURE [BBR+12] ein prominentes Beispiel. Hier wird zwischen normaler Kommunikation und Verkehr zu C2-Servern mittels der Abweichungen im Kommunikationsvolumen, in der zeitlichen Verteilung und in den Zugriffsmustern der Clients unterschieden. BotMiner [GPZL08] findet mögliche infizierte Clients durch das Clustern und Korrelieren von aus Verbindungsstatistiken gewonnenen Charakteristika und von IDS gemeldeten Alarmen. Eine systematische Diskussion weiterer Ansätze zur Botnet-Erkennung findet sich auch in [GZC14].

3.4. Erkennung von automatisch generierten Domainnamen

Die Erkennung von durch Algorithmen erzeugten Domainnamen dient unter anderem der Bekämpfung von Botnets. Zwei konkurrierende Ansätze sind zu unterscheiden: Featurebasierte Methoden extrahieren bestimmte strukturelle, statistische oder linguistische Features, und wenden auf die so gewonnenen Repräsentationen klassische Ansätze des Machine Learning wie beispielsweise Random Forests an. solche Features sind beispielsweise die Verteilung von n-Grammen oder der Anteil der Vokale im Domainnamen. Als Beispiel für einen solchen Klassifikator sei FANCI [STHM18]

genannt. Featurefreie Ansätze verzichten auf dieses Expertenwissen und ermitteln die relevanten Features eigenständig. Sie basieren deswegen meist auf neuronalen Netzen. Ein populäres Beispiel für einen derartigen Klassifikator ist der auf LSTMs basierende ENDGAME [WAAG16], aber auch andere Klassen neuronaler Netze können verwendet werden.

Die Klassifikatoren scheinen nicht sehr robust gegenüber Variationen in den zu klassifizierenden Domainnamen zu sein [PNS+19]. Möglicherweise lässt sich dieses Problem ohne Einbeziehung des Domainkontextes nicht lösen.

Die in der Realität verwendeten domainerzeugenden Algorithmen bilden zwei Klassen – die Domains können zeichen- weise oder aus Wörterbüchern erzeugt werden. Da die Künstlichkeit der durch letztere Methoden generierten Namen oft auch für das menschliche Auge nicht offensichtlich ist, ist diese Klasse erheblich schwerer zu erkennen und ein aktives Forschungsgebiet [HPLJ20].

3.5. Trafficgenerierung

Die Erhebung von Daten stellt sowohl für die Forschung als auch für den Endnutzer ein erhebliches Problem dar. Wie wir bereits in Abschnitt 1.3 festgestellt haben, altern Datensätze schnell und sind deswegen für praxisnahe Forschung von eingeschränktem Wert. Methoden, die neueste Angriffsmethoden erkennen sollen, können deswegen nur schlecht entwickelt und getestet werden. Auch die Verwendung von in Testbeds erzeugten Datensätzen stellt aufgrund der eingeschränkten Realitätsnähe dieser Testbeds nur bedingt eine Alternative dar. Auch für produktiv eingesetzte Systeme sind die rechtlichen Rahmenbedingungen ungünstig, insbesondere wenn diese Daten off-premise, also z.B. in einer Cloud und durch Dritte verarbeitet werden sollen.

Eine Möglichkeit dem zu begegnen ist es synthetische Daten zu generieren. In [RSLH19] geschieht dies mittels spezieller neuronaler Netze. Diese sogenannten GANs sind in der Lage, aus vorhandenen Daten neue Datensätze mit ähnlichen Eigenschaften zu erzeugen, welche dann zum Lernen und Validieren benutzt werden können. Dies erhöht die Variabilität der verfügbaren Daten und stellt auch einen weiteren Schutz gegen Deanonymisierung dar. Darüber hinaus reduziert es unserer Meinung nach auch die Angreifbarkeit der (möglicherweise in einer Cloud laufenden) KI-Modelle.

Diese Art der Datensynthese ist allerdings zum gegenwärtigen Zeitpunkt Objekt aktiver Forschung und noch nicht praxistauglich – so geben die Autoren von [RSLH19] zu, dass wichtige statistische Korrelationen mit ihrer Methode nicht realitätsgetreu abgebildet werden.

3.6. genuas Beteiligung am Forschungsprojekt WINTERMUTE

Die genua GmbH beteiligt sich stark am Forschungsprojekt WINTERMUTE[9]. Zusammen mit den universitären Partnern JMU Würzburg, Universität Bamberg und Universität Bremen sowie weiteren deutschen Unternehmen ist es hier das Ziel, Netzwerkadministratoren in die Lage zu versetzen, mit Unterstützung KI-basierter Hilfs-

[9] https://www.projekt-wintermute.de/

mittel ihr Netz besser verstehen und schützen zu können. Das Forschungsprojekt richtet neben der Analyse und Steuerung des Netzwerkverkehrs und der Struktur hochgradig vernetzter Systeme den Fokus stark auf Fragen der Interpretierbarkeit der Ergebnisse, des Schutzes der Privatheit der Daten sowie der benutzerfreundlichen Mensch-Maschine-Interaktion, um ein effektives Risikomanagement zu erreichen.

4. KI als Helfer für den Administrator

4.1. Praxistaugliche KI

Für einen tatsächlich praxistauglichen Einsatz ist nötig, einen Schritt zurück zu machen und sich darauf zu besinnen, was die grundlegenden Probleme in der Cybersecurity, speziell der Netzwerksicherheit, heutzutage sind und was Methoden der künstlichen Intelligenz an dieser Stelle beitragen können. Denn es ist zwar spannend wenn eine Maschine in Zukunft automatisch ein Netzwerk gegen Angreifer wirkungsvoll verteidigen soll oder auf Bedrohungen für die Betriebs- und Datensicherheit reagiert bevor die Störfälle eintreten, aber das ist eben auch Zukunftsmusik.

Der Netzwerk- und Sicherheitsadministrator von heute ertrinkt im Tagesgeschäft mit seinen ganzen „normalen“ Aufgaben und kann sich um die gefährlichen Situationen sehr häufig nur kümmern, nachdem sie aufgetreten sind. Eine effektive Hilfe sollte also darin bestehen, dem Administrator Routineaufgaben abzunehmen. Dies kann durch Automatisierung erfolgen, und viele Organisationen setzen schon solche Werkzeuge ein. Diese Automatisierungsvorgänge sind aber bisher statisch und individuell entwickelt, erfordern einen initialen Aufwand und bedürfen dann einer konstanten Pflege, um durchgängig effektiv zu sein. Außerdem sind diese Methoden nur geeignet, um schon bekannte Aktionen in der Masse anzuwenden. Ein erster Schritt für die KI ist daher der Vorschlag von Massenaktionen, in denen neue Erkenntnisse empfohlen werden.

Eine wichtige Arbeitsgrundlage für die sichere Administration von Netzwerken ist das Wissen über die Anzahl und Art der vorhandenen Netzwerkteilnehmer. Nur wenn man weiß, welche Geräte überhaupt in welchen Netzwerksegmenten vorhanden sind und welcher Funktion diese Geräte dienen, kann man die passenden Sicherheitsregelungen implementieren. Auch der Vergleich zwischen den laut Plan im Netzwerksegment agierenden Teilnehmern und den tatsächlich vorhandenen Geräten liefert wichtige (und eventuell überraschende) Informationen.

Die Sicherheitsregelungen in einem Netzwerk unterliegen verschiedenen Anforderungen: Sie müssen natürlich gegen Angriffe und Störungen effektiv schützen, dürfen den Betrieb des Netzwerkes und damit der Unternehmensprozesse nicht einschränken oder unterbinden und müssen den Pflegeaufwand für den Administrator gering halten. Um diese Anforderungen zu erfüllen, ist es unverzichtbar, Verallgemeinerungen über den Bestand und die Aufgaben und Verhaltensweisen der Geräte im Netzwerk zu treffen. Diese Verallgemeinerung entspricht einer Gruppierung von Netzwerkteilnehmern nach verschiedenen Gesichtspunkten. Es ist dabei zu beachten, dass in der Praxis eine Gruppierung mit exklusiven Gruppen nach nur einem Merkmal nicht ausreichend ist. Server bieten häufig mehrere Dienste an, sind also Teil mehrerer Gruppen gleichzeitig.

Auch Mitarbeiter arbeiten häufig in verschiedenen Projekten und mit unterschiedlichen Aufgaben und Rollen, die zugeordneten Netzwerkgeräte (Workstations, etc.) erfüllen damit häufig auch die Eigenschaften mehrerer Gruppen.

Aus unserer Sicht bieten sich verschiedene Möglichkeiten der Hilfestellung an:

Wenn ein einzelnes Asset im Inventarsystem (also ein Netzwerkteilnehmer, der bekannt ist oder dynamisch erkannt wurde) bearbeitet wird, ist der Administrator für eine sinnvolle Bewertung des Gerätes auf Kontextinformationen angewiesen. Dann ist es interessant zu wissen, welche anderen Assets gleiches oder zumindest ähnliches Verhalten in ihrer Netzwerkkommunikation zeigen, oder welche Geräte ähnliche Sicherheitsvorfälle auslösen. Damit kann die Bestimmung eines Assets und seiner Funktion leichter erfolgen, die Zuteilung in Gruppen lässt sich damit verifizieren. Aber auch die Reaktion auf Sicherheitsvorfälle kann besser vorbereitet werden.

Um dabei den Nutzer nicht mit der Anzeige von zu viel und irrelevantem Kontext abzulenken, kann und sollte auch die Auswahl des anzuzeigenden Kontextes dynamisch festgelegt werden. Ein Algorithmus würde also beurteilen, welcher Kontext wichtiger erscheint als andere und dann diesen hervorheben. Natürlich muss der Nutzer hierbei immer die Möglichkeit haben, wieder den gesamten verfügbaren Kontext sichtbar zu machen.

- Regelmäßig existieren in Netzwerken viele gleichartige Geräte, die ein gleiches oder ähnliches Verhalten zeigen, gleiche oder ähnliche Aufgaben erfüllen und damit von den gleichen Sicherheitsbestimmungen erfasst werden. Dem Nutzer hilft es also, wenn auch Massenoperationen vorgeschlagen werden, wo ein Algorithmus ähnliches oder gleiches Verhalten mehrerer Geräte findet und dann alle diese Geräte für eine Gruppierung vorschlägt. Dabei liegt es in der Natur der Methoden, dass zwar Korrelationen, aber keine Kausalitäten gefunden werden. Die vorgeschlagenen Gruppen können also eine für den Administrator offensichtliche Gruppierung darstellen, wo die Zeitersparnis der Gruppierung ein benefit ist. Es können dabei auch Gruppen vorgeschlagen werden, deren Zuordnung dem Administrator nicht offensichtlich ist, aber bei genauerem Bedenken eine neue Erkenntnis und damit eine sinnvolle Gruppierung darstellt. Allerdings können auch falsche Korrelationen zu nicht sinnvollen Gruppierungen führen. Darum ist dies nur als Vorschlagssystem für den Administrator gedacht.
- Ergänzend ist es ebenfalls eine Erleichterung für den Administrator, wenn auch eine Zuordnung von Geräten zu existierenden passenden Gruppen vorgeschlagen wird. Damit würden auch automatisch die festgelegten Sicherheitsbestimmungen auf diese Geräte ausgeweitet. Allerdings gilt auch hier, dass dies nur ein Vorschlag für den Administrator sein darf.
- Da sich das Verhalten von Geräten dynamisch ändern kann, ist auch das Verhalten, das eine Gruppe beschreibt, nicht statisch, sondern sollte regelmäßig „mitgelernt“ werden, um das Gruppenverhalten adäquat zu erfassen. Dabei können Abweichungen im Verhalten eines Gerätes auffallen, wenn also ein Teilnehmer der Gruppe sich nicht mehr so verhält, wie es die

Mehrzahl der Gruppe tut. Aber auch die Änderung des Gruppenverhaltens über die Zeit sollte dann überwacht werden, um bei signifikanten Änderungen eine Benachrichtigung an den Nutzer auszulösen. Dabei kann eine große beziehungsweise signifikante Änderung des Gruppenverhaltens auf eine Bedrohung hinweisen, aber eben auch gewünscht sein. Beispielsweise kann sich das Verhalten einer Gruppe nach einem Systemupdate ändern und die Bedrohung geht nicht von den Teilnehmern mit dem geänderten Verhalten aus, sondern vielmehr von den Gruppenteilnehmern, deren Verhalten sich eben nicht ändert, weil etwa das Systemupdate bei diesen Geräten nicht erfolgreich war und damit eine Sicherheitslücke weiterhin ausnutzbar bleibt.

Dabei ist das Finden von Gruppen mit gleichem oder ähnlichem Verhalten einer der Bereiche, die von uns weiter untersucht werden und auf die wir im Folgenden eingehen wollen.

4.2. Beispiel aus der Produktentwicklung: Unterstützung des Administrators bei der Organisation seines Netzwerkes

Der cognitix Threat Defender (cTD) erlaubt das Vergeben nutzergewählter Tags für die beobachteten Assets des internen Netzwerkes und führt über sie verschiedene Statistiken. So sammelt er Zeitreihen über volumetrische Informationen für jedes Protokoll und jede Applikation, die ein bestimmtes Asset als Empfänger oder als Sender verwendet hat. Diese Informationen werden über vorbestimmte Zeitintervalle aggregiert und erlauben den Aufbau assetspezifischer, hochdimensionaler Featurevektoren, die die jeweilige Protokoll- und Applikationsnutzung widerspiegeln.

Von diesen Featurevektoren nehmen wir an, dass sie das Verhalten der Assets bereits gut charakterisieren und deswegen zu Verhaltensgruppen zusammengefasst werden können.

Es ist aber bekannt, dass Clusteringverfahren in großen Dimensionen aufgrund des Fluches der Dimension [Yiu19] häufig schlecht funktionieren. Um ein Clustern zu ermöglichen und den Einfluss von Hintergrundrauschen zu vermindern, führen wir eine Dimensionsreduktion durch. Anschließend wenden wir eine Variante von DBSCAN[RSJM10] an, die die Clusterqualität optimiert.

Aus der so gewonnenen Zuordnung von Assets zu Verhaltensclustern können wir Tags, die nur einzelnen Assets zugeordnet sind, auf die jeweilige Gruppe ähnlicher Assets verallgemeinern. Da die Policy-Rules, die der cTD enforced generiert, auf Tags basieren können und damit unmittelbaren Einfluss auf den erlaubten Netzwerkverkehr haben, werden diese Tags zunächst dem Netzwerkadministrator vorgeschlagen, anstatt automatisch zur Beschreibung der Assets hinzugefügt zu werden.

Wir zeigen die Funktion unseres algorithmischen Ansatzes am CIC-IDS-2017-Datensatz [SHLG18]. Dieser Datensatz wurde in einem Testbed künstlich erzeugt und dient der Validierung von Intrusion Detection Systemen. Obwohl der cTD auch als IDS fungieren kann, soll dies in diesem Beispiel vernachlässigt werden. Deswegen verwenden wir nur den Teil, der frei von Angriffsverkehr ist. Dieser Teil des Datensat-

zes umfasst lediglich die Aufzeichnung eines fünfzehnstündigen Zeitraumes. Der Datensatz liegt als packet capture vor, jedoch sind weder Assets explizit identifiziert noch mit Tags versehen. Allerdings sind die Mitglieder des Opfernetzwerkes, die installierten Betriebssysteme und im Falle der Server teilweise auch ihre Funktion bekannt. Insgesamt handelt es sich um einen Domain Controller, zwei Webserver sowie elf Clientmaschinen, auf denen verschiedene unixoide Betriebssysteme oder Windows-Varianten laufen. Die verwendete Firewall bleibt unsichtbar, da wir in diesem Experiment nicht nach Assets auf Layer 2, sondern nur nach IP-Adressen auf Layer 3 analysieren können. Die bekannten Betriebssysteme, sowie die Server-Client-Unterscheidungen sollen hier die userdefinierten Tags darstellen.

Die Anwendung unserer Verhaltensgruppencharakterisierung erzeugt das in Abbildung 1 gezeigte Ergebnis. Die Achsen besitzen hier keine Bedeutung und sind Resultat des Dimensionsreduktionsprozesses. Eine visuelle Inspektion zeigt im Wesentlichen eine Dreiteilung der Maschinentypen, nämlich in einen nur von Windows-Clients besetzten Bereich (oben rechts), einen von Ubuntu-Clients dominierten Bereich (unten links) und ein Gebiet, das serverartiges Verhalten zeigt (unten rechts). Der Mac-Rechner erscheint recht isoliert. Die als Internal IP 1 und External IP 1 bezeichneten Assets sind in der Dokumentation des Datensatzes nicht beschrieben und daher von unbekannter Funktion.

Die Nähe des Ubuntu-12-Servers und mehr noch des Windows-8.1-Clients zu dem Ubuntu-Bereich deutet auf eine eventuelle Fehlklassifikation hin.

Eine Interaktion mit einem derartigen Graphen ist ab einer hinreichend großen Menge an Assets nicht mehr effizient durchführbar. Auf diese Daten wurde deswegen, wie oben beschrieben, ein Clusteringverfahren angewandt.

Der Algorithmus stuft weiterhin den Domain Controller und den Web Server sowie die Broadcasting-IP als ähnlich ein (Cluster-ID 3). Der optisch gut getrennte Cluster von Windows-Maschinen zerfällt leider in einen Cluster (Cluster-ID 0) und weitere, nicht diesem Cluster zugeordnete Windows-Clients. Der Ubuntu-Client-dominierte Cluster zerfällt ebenfalls in zwei Subcluster (Cluster-IDs 1 und 2). Die verbleibende Ubuntu-Maschine kann keinem dieser Cluster sicher zugeordnet werden. Gleiches gilt auch für die Windows-8.1- und Mac-Clients – dies ist hier allerdings ein wünschenswertes Ergebnis. Wir möchten betonen, dass keine Parameter aus unserem Algorithmus speziell auf diesen Datensatz abgestimmt wurden. Verwendet wurden vielmehr andere, von uns aus nicht-öffentlichen Netzwerkdaten ermittelte Einstellungen. Als Variation eines dichtebasierten Clusteringverfahrens sind die Clusteraufspaltungen durch die geringe Gerätezahl erklärbar, da das Verhältnis von Intra- zu Interclusterabständen ungünstig ist. Angesichts der Kürze des verfügbaren Mitschnittes bewerten wir die gewonnene Assetcharakterisierung insgesamt aber als zufriedenstellend.

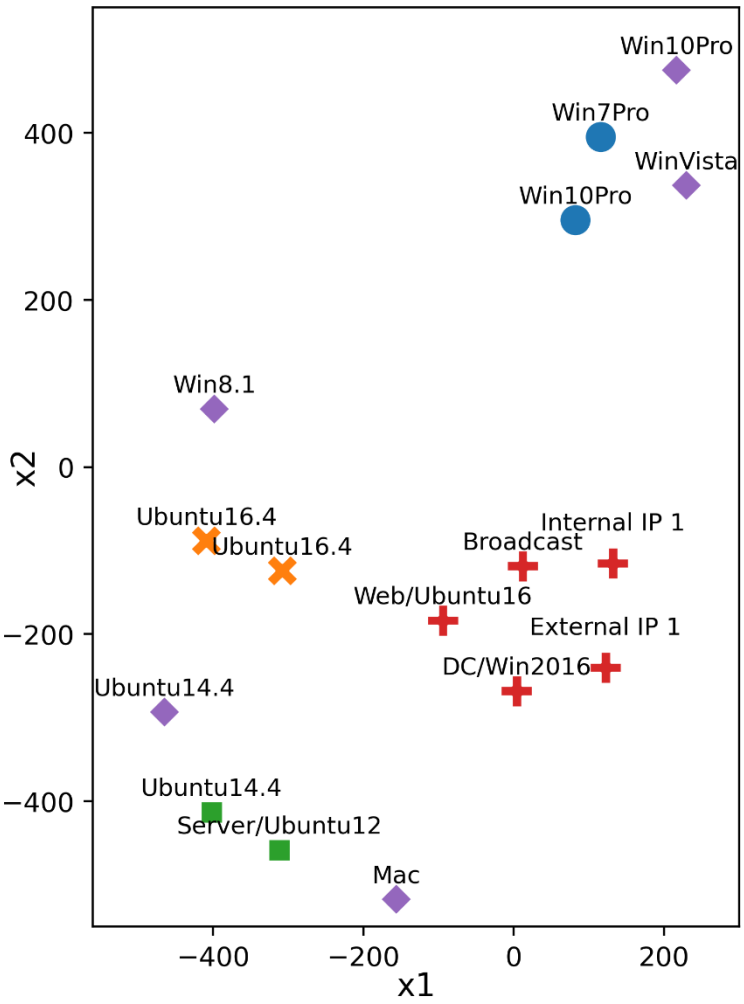


Abbildung 1: Verhaltenscharakterisierung der im Opfernetzwerk des CIC-IDS-2017-Datensatz vertretenen Assets.

5. Fazit

Im vorliegenden Beitrag haben wir das Anwendungspotenzial von KI auf Probleme der IT-Sicherheit beleuchtet. Auf Angreiferseite dominieren dabei neuartige Evasionsstrategien zur Täuschung von Detektoren. Auf Verteidigerseite offenbart sich die Anwendbarkeit sowohl auf Netzwerk- und Logdaten als auch auf die Analyse von Dateien.

Eine weiterführende Diskussion des Offensiv- und Defensivpotentials KI-basierter Methoden findet sich in [TQBZ20]. Wir schilderten weiterhin kurz die aus dem digitalen Wandel und den damit einhergehenden Bedrohungen erwachsenen Hoffnungen, die in verteidigende KIs gesetzt werden.

Als wesentliche Hindernisse zur Nutzbarmachung von KIs in der Netzwerksicherheit haben wir unter anderem die Menge, Qualität und Zugänglichkeit an realistischen und aktuellen Trainingsdaten, Fehler bei der Modellgenerierung, die mangelnde Robustheit

der vorhandenen KI-Systeme und ihre aufgrund hoher Komplexität oft fehlenden Interpretierbarkeit identifiziert.

Der aktuell nötige Aufwand, um diese Probleme zu umgehen, ist höher als ihn die meisten Organisationen und Firmen zu leisten im Stande sind und kann realistisch im Moment nur durch große Cloud-Anbieter gestemmt werden. Dies führt derzeit zu vielen kleinen Pilotprojekten, welche richtig und wichtig sind und auch dementsprechend vermarktet werden, aber häufig auch den Anschein erwecken, dass KI um der KI willen passiert und der tatsächlich greifbare Nutzen noch nicht gegeben ist.

Die großen Erwartungen, die an KI-Systeme in der Cybersecurity gestellt werden, führen unserer Meinung nach zu einem derzeit nicht zu rechtfertigenden Vertrauen in autonom agierende KI-Systeme. Damit dieses Vertrauen nicht ins Gegenteil umschlägt und um Risiken zu vermeiden, sollte von autonomen Systemen für Aufgaben der Cybersecurity abgesehen werden, da die Gefahr der Fehlentscheidungen und der daraus resultierenden Schäden noch zu groß ist. Wenn der Mensch aber die Analysen und Erkenntnisse der Maschine im konkreten Fall jeweils als Empfehlung betrachtet und validiert, lässt sich das Vertrauen in die Fähigkeiten der KI begründet aufbauen und stellt gleichzeitig die Akzeptanz der von etwaigen Reaktionen betroffenen Nutzer sicher.

Dann aber ergeben sich mit dem Einsatz von KI-Systemen durchaus Aufgaben und Möglichkeiten auch auf Seite der Verteidiger der Cybersecurity, die den Administratoren effektiv helfen den Bedrohungen entgegen zu wirken.

Experimentell zeigten wir auf, wie KI menschliche Akteure unterstützen kann, indem DPI-Informationen zur halbautomatischen Assetcharakterisierung und -gruppierung verwendet werden. Dadurch wird die Formulierung und Durchsetzung von feingranulareren und spezifischeren Sicherheitsbestimmungen auch in großen Netzwerken absehbar leichter und stellt in unseren Augen ein lohnendes Ziel der Entwicklung der nächsten Jahre dar. Bei allen Bedenken und Risiken möchten wir auch betonen, dass die Vorteile mittel- und langfristig überwiegen werden und die Entwicklung der KI auch im Hinblick auf die Cybersecurity definitiv weiter zu verfolgen und voranzutreiben sind.

Sowohl in dem aktuellen Entwicklungsstadium als auch in der Zukunft kommt es auf das richtige Zusammenspiel von Mensch und Maschine an.

“By far the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.” – Eliezer Yudkowsky [Yud06]

6. Danksagung

Die Autoren danken Steffen Ullrich für die wertvollen Hinweise und fruchtbare Diskussionen.

Literaturhinweise

[AA19] Shahar Zini Adi Ashkenazy. Cylance, I kill you! https://skylightcyber.com/2019/07/18/cylance-i-kill-you/, 2019. Accessed: 2021-01-04.

[AQP+20] Daniel Arp, Erwin Quiring, Feargus Pendlebury, Alexander Warnecke, Fabio Pierazzi, Christian Wressnegger, Lorenzo Cavallaro, and Konrad Rieck. Dos and don'ts of machine learning in computer security, 2020.

[AWF16] Hyrum S. Anderson, Jonathan Woodbridge, and Bobby Fi- lar. DeepDGA: Adversarially-tuned domain generation and detection. CoRR, abs/1610.01969, 2016.

[BBR+12] Leyla Bilge, Davide Balzarotti, William Robertson, Engin Kir- da, and Christopher Kruegel. Disclosure: Detecting botnet command and control servers through large-scale netflow analysis. pages 129–138, 12 2012.

[Bou20] Daniel Bourke. Introducing the 2020 machine learning roadmap. https://www.mrdbourke.com/2020-machine-learning-roadmap/ , 2020. Accessed: 2021-01-04.

[BTCV18] Alejandro Correa Bahnsen, Ivan Torroledo, L. Camacho, and Sergio Villegas. DeepPhish : Simulating malicious AI, 2018.

[CLH19] Isaac A. Corley, Jonathan Lwowski, and Justin Hoffman. Do- mainGAN: Generating adversarial examples to attack domain generation algorithm classifiers. CoRR, abs/1911.06285, 2019.

[DK18] Marc Ph. Stoecklin Dhilung Kirat, Jiyong Jang. Deeplocker: Concealing targeted attacks with AI locksmithing. Blackhat USA, 2018.

[DR20] J. Eaton D.F. Reding. Science & Technology Trends 2020-2040. Technical report, NATO, 2020.

[FJR15] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Mo- del inversion attacks that exploit confidence information and basic countermeasures. pages 1322–1333, 10 2015.

[Fle19] William Fleshman. Evading machine learning malware classifiers for fun and profit! https://towardsdatascience.com/

[FRZ+18] William Fleshman, Edward Raff, Richard Zak, Mark McLean, and Charles Nicholas. Static malware detection & subterfuge: Quantifying the robustness of machine learning and current anti-virus, 2018.

[GPZL08] Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee. BotMiner: Clustering analysis of network traffic for protocol- and structure-independent botnet detection. pages 139–154, 01 2008.

[GZC14] Sebastián García, Alejandro Zunino, and Marcelo Campo. Survey on network-based botnet detection methods. Security and Communication Networks, 7, 05 2014.

[HBC+13] Ivan Homoliak, Maroš Barabas, Petr Chmelar, Michal Drozd, and Petr Hanacek. Asnm: Advanced security network metrics for attack vector description. 07 2013.

[HPLJ20] Kate Highnam, Domenic Puzio, Song Luo, and Nicholas R. Jennings. Real-time detection of dictionary DGA network traffic using deep learning, 2020.

[HS17] Joshua Saxe Hillary Sanders. Garbage in, garbage out: How purportedly great ML models can be screwed up by bad data. Technical report, Blackhat USA, 2017.

[LLZ+18] Pan Li, Qiang Liu, Wentao Zhao, Dongxu Wang, and Siqi Wang. BEBP: A poisoning method against machine learning- based IDSs. 03 2018.

[MLM+20] Patrick McDaniel, John Launchbury, Brad Martin, Cliff Wang, and Henry Kautz. Artificial intelligence and cyber security: Opportunities and challenges technical workshop summary report. NETWORKING & INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT SUBCOMMITTEE and the MACHINE LEARNING & ARTIFICIAL INTELLIGENCE SUB- COMMITTEE of the NATIONAL SCIENCE & TECHNOLOGY COUNCIL, 2020.

[MNO19] Ravi Mangal, Aditya V Nori, and Alessandro Orso. Robustness of neural networks: a probabilistic and practical approach. In 2019 IEEE/ACM 41st International Conference on Software En- gineering: New Ideas and Emerging Results (ICSE-NIER), pages 93–96. IEEE, 2019.

[MSX+18] Christos M. Mathas, Olga Segou, George Xilouris, Dimitris Christinakis, Michail Kourtis, Costas Vassilakis, and Anasta- sios Kourtis. Evaluation of Apache Spot's machine learning capabilities in an SDN/NFV enabled environment. pages 1–10, 08 2018.

[NADL19] Adam Noack, Isaac Ahern, Dejing Dou, and Boyang Li. Does interpretability of neural networks imply adversarial robust- ness? arXiv preprint arXiv:1912.03430, 2019.

[PM15] S. Palka and D. McCoy. Fuzzing e-mail filters with generative grammars and n-gram analysis. In WOOT, 2015.

[PNS+19] Jonathan Peck, Claire Nie, Raaghavi Sivaguru, Charles Gru- mer, Femi Olumofin, Bin Yu, Anderson Nascimento, and Mar- tine De Cock. CharBot: A simple and effective method for evading DGA classifiers. IEEE Access, PP:1–1, 07 2019.

[RG18]Maria Rigaki and Sebastián García. Bringing a GAN to a knife- fight: Adapting malware communication to avoid detection. 05 2018.

[RSJM10] Anant Ram, Jalal Sunita, Anand Jalal, and Kumar Manoj. A density based algorithm for discovering density varied clus- ters in large spatial databases. *International Journal of Computer Applications*, 3, 06 2010

[RSLH19] Markus Ring, Daniel Schlör, Dieter Landes, and Andreas Hotho. Flow-based network traffic generation using generative adversarial networks. Computers & Security, 82:156 – 172, 2019.

[RT19] Geert van der Linden Allan Frank Luis Delabarre Jerome Bu- vat Jeff Theisler Sumit Cherian Yashwardhan Khemka Ron To- lido, Anne-Laure Thieullent. Reinventing cybersecurity with artificial intelligence: The new frontier in digital securi- ty. https://www.capgemini.com/wp-content/uploads/2019/ 07/AI-in-Cybersecurity_Report_20190711_V06.pdf, 2019. Accessed: 2020-01-08.

[RWG+17] Markus Ring, Sarah Wunderlich, Dominik Grüdl, Dieter Lan- des, and Andreas Hotho. Creation of flow-based data sets for intrusion detection. Journal of Information Warfare, 16:41–, 12 2017.

[RWS+19] Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Lan- des, and Andreas Hotho. A survey of network-based intrusion detection data sets. Computers & Security, 86:147–167, 2019.

[SBE+17] Tara Salman, Deval Bh, Aiman Erbad, Raj Jain, and Moham- med Samaka. Machine learning for anomaly detection and categorization in multi-cloud environments. pages 97–103, 06 2017.

[SHLG18] Iman Sharafaldin, Arash Habibi Lashkari, and Ali Ghorba- ni. Toward generating a

new intrusion detection dataset and intrusion traffic characterization. pages 108–116, 01 2018.

[STHM18] Samuel Schüppen, Dominik Teubert, Patrick Herrmann, and Ulrike Meyer. FANCI : Feature-based automated nxdomain classification and intelligence. In 27th USENIX Security Sympo- sium (USENIX Security 18), pages 1165–1181, Baltimore, MD, August 2018. USENIX Association.

[TL20] Ankit Thakkar and Ritika Lohiya. A review of the advan- cement in intrusion detection datasets. Procedia Computer Science, 167:636–645, 04 2020.

[TMF19] Mariarosaria Taddeo, Tom McCutcheon, and Luciano Floridi. Trusting artificial intelligence in cybersecurity is a double- edged sword. Nature Machine Intelligence, 1, 11 2019.

[TQBZ20] Cong Truong, Diep Quoc Bao, and Ivan Zelinka. Artificial intelligence in the cyber domain: Offense and defense. Sym- metry, 12:410, 03 2020.

[TZ19] Cong Truong and Ivan Zelinka. A survey on artificial in- telligence in malware as next-generation threats. Mendel, 25:27–34, 12 2019.

[Ull15] Steffen Ullrich. HTTP evader - automate firewall evasion tests. Technical report, genua GmbH, 2015.

[Ull17] Steffen Ullrich. Umgehung von Firewalls auf Applikations- ebene. Technical report, genua GmbH, 2017.

[Ull20] Steffen Ullrich. Persönliche Mitteilung. 2020.

[WAAG16] Jonathan Woodbridge, Hyrum Anderson, Anjum Ahuja, and Daniel Grant. Predicting domain generation algorithms with long short-term memory networks. 11 2016.

[WXE16] Yanjun Qi Weilin Xu and David Evans. Automatically evading classifiers: A case study on PDF malware classifiers. Technical report, University of Virginia, 2016.

[Yiu19] Tony Yiu. The curse of dimensionality: Why high dimensional data can be so troublesome. https://towardsdatascience.com/ the-curse-of-dimensionality-50dc6e49aa1e, 2019. Accessed: 2020-12-14.

[YS16] Roman V Yampolskiy and MS Spellchecker. Artificial intelli- gence safety and cybersecurity: A timeline of AI failures. arXiv preprint arXiv:1610.07997, 2016.

[Yud06] Eliezer Yudkowsky. Artificial intelligence as a positive and negative factor in global risk, 2006.

[Zet19] Kim Zetter. Researchers easily trick Cylance's AI-based antivirus into thinking malware is 'goodware'. https://www.vice.com/en_us/article/9kxp83/researchers-easily-trick-cylances-ai-based-antivirus-into-thinking-

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

CyberSecurity Challenges: Serious Games for Awareness Training in Industrial Environments

Tiago Gasiba[1,2], Prof. Dr. Ulrike Lechner[2], Maria Pinto-Albuquerque[3]

Abstract:
Awareness of cybersecurity topics, e.g., related to secure coding guidelines, enables software developers to write secure code. This awareness is vital in industrial environments for the products and services in critical infrastructures. In this work, we introduce and discuss a new serious game designed for software developers in the industry. This game addresses software developers' needs and is shown to be well suited for raising secure coding awareness of software developers in the industry. Our work results from the experience of the authors gained in conducting more than ten CyberSecurity Challenges in the industry. The presented game design, which is shown to be well accepted by software developers, is a novel alternative to traditional classroom training. We hope to make a positive impact in the industry by improving the cybersecurity of products at their early production stages.

Keywords: Awareness; Critical Infrastructures; CyberSecurity Challenges; Development; Industry; Secure Software; Serious Games; Software Developers

1. Introduction

If not addressed during the early stages of software design and implementation, software development errors and security vulnerabilities can end up in a final product or service. Security vulnerabilities can result in serious negative consequences for society, the customer, and the company that produced the software. Think, e.g., of critical infrastructures as the grid, transportation, or production lines: a security vulnerability in the code may cause interruptions in service quality for individual customers when critical machinery or information systems fail or even for society when critical infrastructure fails. Over the last years, the number of industrial security-related incidents has been increasing, which has resulted in severe incidents, leading to a substantial financial impact, reaching up to 1.6% of GDP in some EU countries [7].

To address these issues, products and services provided by the industry must follow IT security standards. These standards mandate the implementation of a secure software development lifecycle and secure coding guidelines that must be followed to write secure code. Prominent examples of these standards for industrial environments are the IEC 62443 [23], ISO 27001 [24], and the Grundschutzkatalog from the Bundesamt für Sicherheit in der Informationstechnik (BSI) [4]. Examples of secure coding guidelines widely used in the industry are the SEI-CERT Java Secure Coding Guidelines and SEI-CERT C/C++ Secure Coding Guidelines, both from Carnegie Mellon [5]. The Open Web Application Security Project (OWASP, [27]) and the BSI (BSI 5.21, [3]) provide

[1] Siemens AG, Munich, Germany, tiago.gasiba@siemens.com
[2] Universität der Bundeswehr München, Munich, Germany, tiago.gasiba@unibw.de, ulrike.lechner@unibw.de
[3] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisboa, Portugal, maria.albuquerque@iscte-iul.pt

secure coding guidelines which are specific for web application development and widely used in the industry.

These standards provide a much-needed basis that establishes ground rules required to produce secure products and services. The effectiveness of these standards is related to the level of awareness and understanding of the standards by the persons directly affected by them: software developers. However, a recent study by Patel et al. [28] has shown that more than 50% of software developers cannot spot software vulnerabilities in source code. This lack of awareness about secure coding is a problem that needs to be addressed.

Among others, a possible way to address this issue is to provide training to software developers on secure coding. We present a new serious game designed to raise awareness and train software developers in secure coding in this work. The serious game, named CyberSecurity Challenges, is an adaption of the capture-the-flag game genre. Capture-the-flag was initially developed in the penetration testing community to practice and train offensive IT-security skills. The idea is that by attacking a system, well-trained penetration testers can discover vulnerabilities in products and services that can be fixed before final shipment to the customer. However, since these activities require a full or partially developed project, they often occur late in the software development stages. We propose using an adapted version of the game, which targets software developers, focuses on the defensive perspective, and has the primary goal of increasing awareness of secure coding guidelines and secure coding best practices. Furthermore, we show how our concept can be used for onsite IT-Security Awareness Workshops and how it can be adapted for online training.

This work is organized as follows: in section 2, the authors briefly discuss previous work related to the cybersecurity challenges. Section 3 introduces the CyberSecurity Challenges and discusses challenges based on open-source components and the Sifu platform. Section 4 discusses the games' evaluation in an industrial context through survey results, participant feedback, and lessons learned. Finally, section 5 summarizes and concludes the paper.

2. Related Work

Although several methods exist to deal with software vulnerabilities, e.g., requirements engineering and code reviews, we focus on awareness training for software developers. Several previous studies indicate that software developers lack secure programming awareness and skills [1,28,32]. In 2020, Bruce Schneier, a well-known security researcher and evangelist stated that less than 50% of software developers can spot security vulnerabilities in software [30]. His comment adds to a discussion on secure coding skills. In 2011, Xie et al. [33] did several interviews with 15 senior professional software developers in the industry with an average of 12 years of experience. Their study has shown a disconnect between software security concepts and the role they have in their jobs. Awareness training on information security is addressed in McIlwraith [25], which provides a systematic methodology and a baseline for implementing awareness training.

There is a stream of literature on compliance with security policies, which deals with general employees, not with software developers specifically. This stream of literature explores many reasons why people do not comply with IT security policies. The unified framework by Moody et al. [26] summarizes the academic discussion on compliance with IT-security policies. Empirical findings conclude that neither deterrence nor punishment, such as, e.g., public blame, works to increase compliance. However, increasing IT-security awareness increases the level of compliance [31]. In their seminal review article, Hänsch et al. [22] define IT-security awareness in the three dimensions: Perception, Protection, and Behavior. The concept of IT-security awareness is typically used in IT security management contexts. We adapt these concepts to software developers as follows [16]: perception - knowledge of existing software vulnerabilities, protection - knowing the existing mechanisms, e.g., secure coding guidelines and software development best practices, that avoid software vulnerabilities, and behavior - knowledge and intention to write secure code.

Graziotin et al. [21] show that happy developers are better coders, i.e., produce higher quality code and software. Their work suggests that by keeping developers happy, we can expect that the code they write has a better quality and, by implication, be more secure. Davis et al. [6] show, in their construct, that cybersecurity games have the potential to increase the overall happiness of software developers. Their conclusions support our approach to use a serious game to train software developers in secure coding. Awareness games are a well-established instrument in information security. They are discussed in de-facto standards as the BSI Grundschutz-Katalog [4] (M 3.47, Planspiele) as one means to raise awareness and increase the level of security. Frey et al. [8] show the potential impact of playing cybersecurity games on the participants and also the importance of playing games as a means of cybersecurity awareness. They conclude that cybersecurity games can be a useful means to build a common understanding of security issues. Rieb et al. [29] provide a review of serious games in cybersecurity and conclude that there are many approaches. The games listed mainly address information security rather than secure coding. Documented and evaluated games are [2] and [29].

Capture-the-flag is one particular genre of serious games in the domain of Cybersecurity [6]. Game participants win flags when they manage to solve a task. Forensics, cryptography, and penetration testings are typical skills necessary for solving tasks and capturing flags. The present work refines capture-the-flag games to achieve the goal of raising secure coding awareness of software developers in the industry. Previous work on selected design aspects on the CSC includes [10,13–15,18–20].

3. CyberSecurity Challenges

In this section, we introduce the CyberSecurity Challenges (CSC), which were developed to raise awareness on secure coding. We also present a detailed discussion on creating these games (1) by using existing open-source components, and (2) using the open-source Cybersecurity Challenge platform developed by the authors - the Sifu platform.

3.1. What are CyberSecurity Challenges

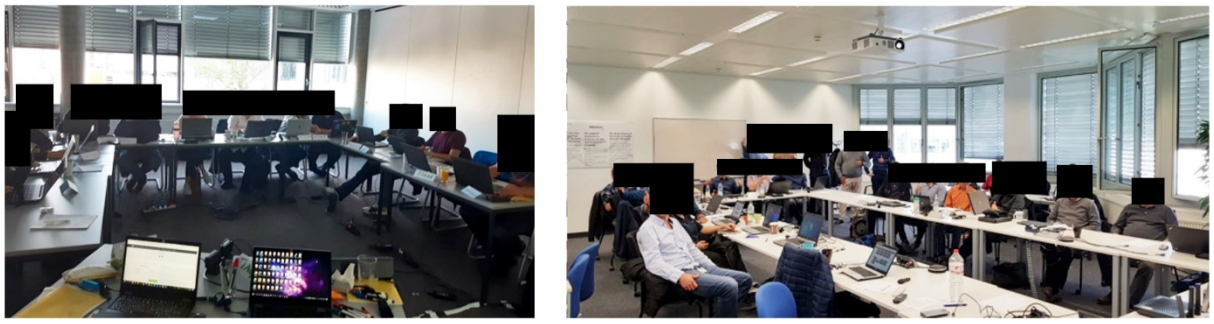


Fig. 1. CyberSecurity Events - On-site Events

CyberSecurity Challenges (CSC) are a genre of serious games developed with the specific purpose of raising awareness of industrial software developers in the topic of secure coding and secure coding guidelines. Figure 1 shows two examples of CSC events that took place in the industry.

The game consists of a platform where several participants (i.e., software developers) form teams that compete against each other in solving secure coding challenges. The challenges consist of exercises that are developed primarily to address software development vulnerabilities. Solving the challenges requires the participants to know and follow secure coding guidelines. Figure 2 depicts the general architecture of CyberSecurity Challenges (CSC), which consists of the following components: Challenges, Dashboard, and Countdown.

The challenges represent the individual exercises that the participants must solve to capture flags and gain points. The dashboard displays the available challenges and is used to control each team's current status regarding the number of gathered points. Figure 3 shows an example of a dashboard based on the open-source CTFd platform. Upon solving a challenge, the participants receive a flag. This flag is represented by a random-like string that can be redeemed for points in the dashboard. The reward on the number of points is related to the difficulty level of the challenge. The countdown component consists of a timer that, when expired, automatically locks the dashboard, preventing further submission of flags. The countdown timer is also used to incentivize the competitiveness of the players on solving the challenges. One or more coaches take part in the game by aiding every team and every participant during the gameplay, such that no one gets stuck or lost while solving the exercises. The coaches also supervise the game to ensure that the desired game objectives, e.g., learning goals, are achieved. In the end, the team with the highest amount of points wins the challenge. Nevertheless, all teams and players are winners since, by participating in the game, awareness of secure coding is stimulated. The game's competitive nature increases the fun, contributes to the overall awareness level of every player, and ensures a memorable event that can potentially have long-lasting impressions.

The different CSC challenges can be implemented in two ways: 1) using open-source components or 2) using self-developed components. In the first case, the challenges are implemented through adaptation, re-use, and re-purposing existing open-source projects and components. This method's main advantage is the reduced cost of implementation of individual challenges while outsourcing their maintenance. In the second case, the challenges can be better adapted to internal company policies while also focusing more on the defensive perspective. The architecture shown in Figure 2 was initially developed for onsite events. A recent installment of the game [15] allows the game not only to be played remotely but also to include an intelligent coach based on artificial intelligence techniques. In the following, we present a more detailed introduction of the CSC game implementation based on open-source components and the Sifu platform.

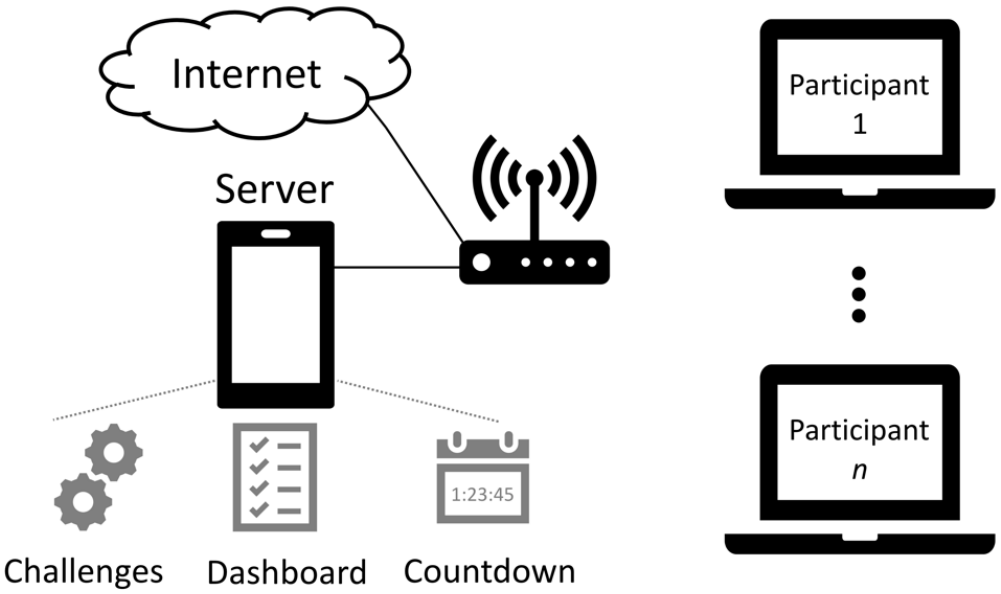


Fig. 2. Architecture of CyberSecurity Challenges infrastructure

3.2. CyberSecurity Game

The CSC game was developed in the industry, focusing on Web and C/C++ developers. In contrast to C/C++, for the web challenges, it was decided not to focus on a single programming language or framework since many of these programming languages and frameworks are in everyday use in the company where the CSC game was developed. In this case, we chose a generic approach based on the Open Web Application Security Project - OWASP [27]. The challenges' design took two approaches: 1) based on open-source components and 2) design of own challenges. A common approach to the design of the challenges is given in [19]. Each challenge is presented to the participants according to the following phases: Phase 1 - introduction, Phase 2 - challenge, and Phase 3 - conclusion. Phase 1 presents an introduction to the challenge and sets up the scenario; the core part of the challenge is phase 2; phase 3 concludes the challenge by presenting additional text related to secure coding guidelines, or further questions related to phase 2. The types of challenges are Single-Choice Questions, Multiple-Choice Questions, Text-Entry Questions, Associate-Left-Right, Code-Snippet Challenge, and Code-Entry Challenge.

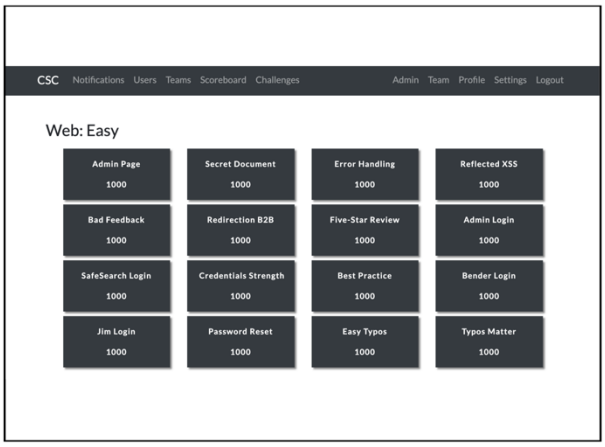


Fig. 3. Dashboard

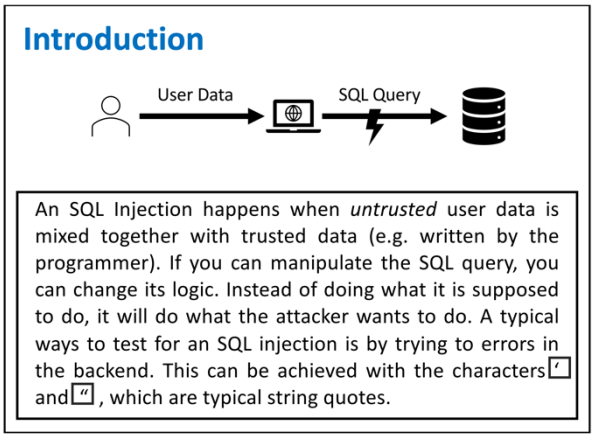


Fig. 4. Web Challenge: Phase 1

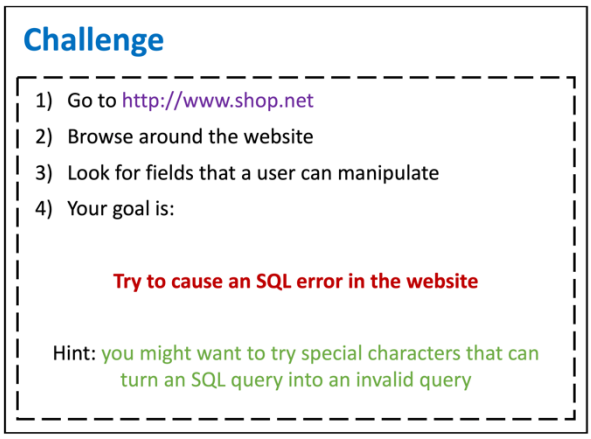


Fig. 5. Web Challenge: Phase 2

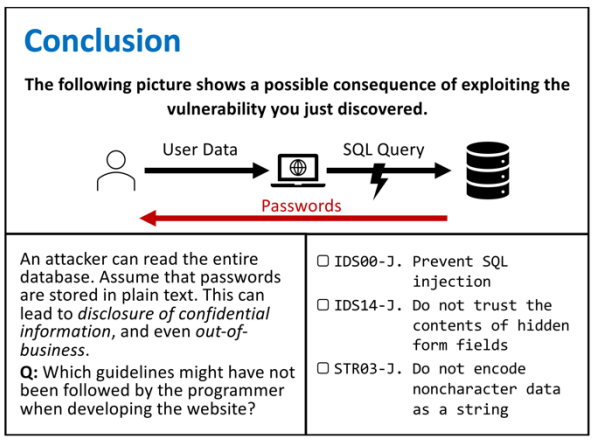


Fig. 6. Web Challenge: Phase 3

Challenges using Open-Source Components

Challenges on secure coding for software developers can be implemented by using and adapting existing open source components. Since most of the available projects focus on the offensive perspective, the following adaptations are suggested: 1) include an incomplete description on how to solve the challenge, and 2) provide follow-up questions related to secure coding guidelines. Fig. 4-6 shows an example of a challenge for Web developers using OWASP JuiceShop. The challenge's learning goal is to understand what SQL injections are and how to identify an SQL injection quickly. Phase 1 sets the stage for the challenge (Fig. 4). In Phase 2, the player is assisted with how to find the vulnerability, through the textual description, as in Fig 5, or also directed by the game coaches. The last phase consists of an additional question related to the exercise, as shown in Fig 6, which enquires and directs the player to corresponding secure coding guidelines.

Table 1 shows the open-source projects and components which have been used to design CSC challenges for Web and C/C++, along with the expected effort required to modify them. Note that the design of these challenges is based on open source components that include an offensive perspective. Therefore, after the components' adaptation, it is more natural and accurate to describe these types of challenges as defensive/offensive (D/O).

Type	Project	Effort	Description
Web/Java	Juice Shop	Minimal	Insecure web application for training purposes from the OWASP project.
Web/Java	Java SEI-CERT	Medium	Secure coding guidelines dedicated to Java from Carnegie Mellon University
Web	Vulnerable API	Medium	REST API containing several vulnerabilities
C/C++	MBE	Small	Vulnerable code from RPISEC course at Rensselaer Polytechnic Institute
C/C++	C/C++ SEI-CERT	Medium	Secure coding guidelines dedicated to C/C++ from Carnegie Mellon University
C/C++	Vulnerable code snippets	High	Vulnerable C/C++ code from NIST (Juliet Set)

Table 1. Open-Source Tools used for CyberSecurity Challenges

Defensive Challenges using Sifu Platform

The Sifu platform hosts code projects containing vulnerabilities in a web application. A web interface is chosen to avoid the players' need to install software on their machines, which might be difficult or impossible in an industrial setting. The players' task is to fix the project's source code to bring it to an acceptable solution (therefore focusing on the defensive perspective). An acceptable solution is when the source code is compliant to secure coding guidelines and does not have known vulnerabilities. The Sifu platform contains two main components: 1) challenge assessment and 2) an automatic coach. The challenge assessment component analyses the proposed solution submitted by a player and determines if it is acceptable. Analysis is based on several tools, e.g., compiler output, static code analysis, and dynamic code analysis. The automatic coach component is implemented through an artificial intelligence technique that provides hints to the participant when the solution is not acceptable, with the intent to guide the participant to an acceptable solution.

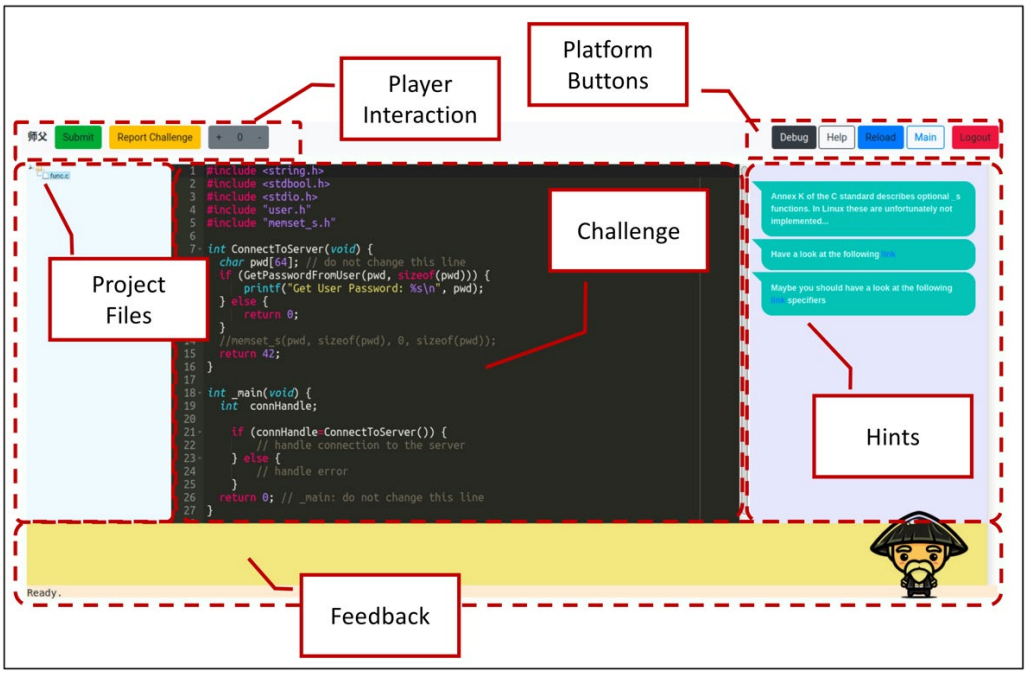


Fig. 7. Sifu Platform - User Interface

Figure 7 shows the web user interface of the Sifu platform. Note that only phase 2 is shown in the figure. The player can browse the different files of the project. All the hints issued by the automatic coach are available on the right-hand side. If the player experiences errors when using the platform, these can be reported for later analysis and improvement. Since untrusted and potentially malicious code will be executed in the platform during the analysis stage, several security mechanisms need to be implemented to guarantee that the players cannot hack it. Further detailed information on the implementation is available in [15, 18]. The open-source Sifu platform can be downloaded from Github [9].

4. Evaluation of CyberSecurity Challenges

The authors have implemented the CSC game and have held a total of thirteen CSC events in the industry: nine onsite events (from November 2017 to October 2019) and four CSC online events (from June 2020 to July 2020). Furthermore, two events in November 2020 were held in the academia.

No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Type	D/O	D/O	D/O	D/O	D/O	D/O	D/O	D/O	D/O	D	D	D	D	A	A
Date	11/17	5/18	7/18	7/18	9/18	7/19	7/19	9/19	10/19	6/20	7/20	7/20	7/20	11/20	11/20
NP	11	12	6	30	16	14	15	7	23	15	21	20	15	12	4
Where	DE	DE	DE	DE	DE	CH	CH	DE	TK	OL	OL	OL	OL	OL	OL

D/O: Defensive/Offensive, **D**: Defensive, **NP**: Number of participants, **DE**: Germany, **CH**: China, **TK**: Turkey, **OL**: Online

Table 2. CyberSecurity Challenge Events

Table 2 summarizes all the events. To evaluate and refine the CSC game, we have performed empirical studies together with the CSC events. The results presented in this work summarize our empirical studies by focusing on the following six dimensions:

- **Know-how** - evaluate if the CSC game contributes to learning new techniques and principles to be used during software development
- **Significance** - evaluate if the CSC game contributes to understanding the importance of secure coding guidelines
- **Skills** - evaluate if the CSC game contributes to improve the participants' secure coding skills
- **Clarity** - evaluate if the challenges in the CSC game are clearly presented
- **Coaching** - evaluate if the help provided by coaches is adequate during gameplay
- **Behavior** - evaluate if the participants, after playing the CSC game, feel prepared to write secure code

The answers to the survey questions were based on a 5-point Likert scale on agreement and are summarized through negative (-) answers (strongly disagree and disagree), neutral (N), and positive (+) answers (agree and strongly agree). Answering the survey was not mandatory, and the participants that took part in the study have given their consent; additionally, their answers were anonymized. Although the total number of participants to the CSC events exceeded 200, the total number of participants that answered the survey were: 56 - for defensive/offensive (D/O) events 1-9, 25 - for

defensive (D) events 10-13, and 14 for defensive challenges in the academia (A) in events 14-15. Additional results were captured through open feedback, questions, and discussions with the participants. The main positive and negative quotes from the participants were also collected. In the following sub-sections, we present a brief overview and discussion of the survey's main results, participant feedback, and an overview of the lessons learned on the design of CSC games and events. For a more in-depth overview of the empirical studies, we refer the reader to the work published by the same authors in [10–20].

4.1. Results

Table 3 shows a summary of the results for the six different questions, both for the industry (81 participants) and the academia (14 participants). The two highest-ranked questions are: Defensive/Offensive Challenges - Q2, Q5; Offensive Challenges - Q2+Q3+Q5, Q1; Offensive Challenges - Q3, Q4+Q5. The results in this table leads to the following conclusions: (1) defensive challenges have a higher level of agreement than defensive/offensive challenges, (2) there is a higher amount of neutral answers in defensive/offensive than in purely defensive challenges, (3) nevertheless both defensive/offensive and defensive challenges show a high level of agreement on the suitability as an method to increase awareness.

Question	Industry						Academia			Description
	D/O			D						
	-	N	+	-	N	+	-	N	+	
Q1	12.5	7.1	80.4	0.0	10.0	90.0	6.2	12.5	81.3	I learned new techniques and principles of secure software development
Q2	0.0	5.3	94.7	0.0	0.0	100.0	18.7	12.5	68.8	I understand the importance of secure coding guidelines
Q3	3.6	14.3	82.1	0.0	0.0	100.0	0.0	6.2	93.8	Focusing on the challenges improves my practical secure coding skills
Q4	8.9	8.9	82.2	8.0	8.0	84.0	0.0	12.5	87.5	The learning goals of the challenges were clearly explained
Q5	1.8	12.5	85.7	0.0	0.0	100.0	12.5	0.0	87.5	The help from the coaches was adequate
Q6	8.9	26.8	64.3	0.0	20.0	80.0				I feel that I am prepared to handle issues related to secure coding at work

-: Negative agreement, **N**: Neutral answers, **+**: Positive agreement
D/O: Defensive/Offensive, **D**: Defensive

Table 3. CyberSecurity Challenge – Empirical Results

These results mean that, while there are good indicators that both challenge types are suitable to raise secure coding awareness on software developers, the indicators for defensive challenges shows a higher level of adequacy. The presented results also show promising results for the three awareness constructs as introduced by Hänsch et al. [22] - perception (Q2), protection (Q1), and behavior (Q3). An extended experiment, using the same artifact but in an academic setting, also shows good indicators of its suitability to train future generations of junior industrial software developers. For a more in-depth discussion on the presented results, we refer the reader to the literature by the same authors [10–20].

4.2. Participant Feedback

Table 4 shows the main positive and negative quotes from participants to the CSC games. Most of the collected feedback was positive and indicated that the CSC game is suitable for raising secure coding awareness. The feedback obtained by the authors, during all the events that took place in the industry, has also shown that the software developers highly appreciate playing the CSC game.

	Quotes from Participants
Positive	I really enjoyed participating in the challenges.
	I am well excited in trying to crack the answers to the challenges
	Enjoyed the challenges, different topics and how competitive we became
	It was lots of fun. Questions inbetween were nice.
	Enjoyed and lots of fun. I've learned many interesting things
	Quite fun and nice to work, especially work in team
	Enjoyed and learned very much
	It was really funny and I leaned a lot
	Funny and interesting; learned a lot - hope to remember and use in practice
	Really liked and enjoyed the exercises
	Enjoyable to try everything and very fun
Negative	Hints not always accurate or precisely leading to the problem in the code
	We do not perform attacks on systems
	Could not understand what to do in the challenge
	Some hints are very generic
	The user interface is very minimalist
	User interface could be improved

Table 4. Quotes from Participants to CyberSecurity Challenges

For one of the groups that participated in the CSC event, the players have joined forces together after the event and searched the internet for further similar games, thus giving a good indicator of possible long-term effects. Another success factor was the positive feedback from management, leading to recurring CSC events and establishing good impression on managers. Nevertheless, we collected some negative feedback related to the user interface and the hints' precision. Additional negative feedback is related to the fact that defensive/offensive challenges still include an offensive part. The offensive part's presence can lead to difficulty in understanding what to do in the challenge due to the participants' background (i.e., software developers). In a separate discussion, we could conclude that coaches' help can positively improve the game experience.

4.3. Evaluation of the design

Figure 8 shows an overview of the lessons learned on the different aspects related with the design, deployment and refinement of CyberSecurity Challenges. These have resulted from all the thirteen deployments that were performed in the industry. The five top-level design aspects are: 1 - learning goals, 2 - time management, 3 - game roles, 4 - game components, and 5 - challenges. Learning goals (L) are related to the game's content and adaptation to the target group of software developers and considers programming language, secure coding guidelines, alignment with management, and the current status quo of know-how. Time management is an essential aspect of deploying and using games in the industry. This aspect includes the agenda of the event and the temporal dimensioning of the challenges.

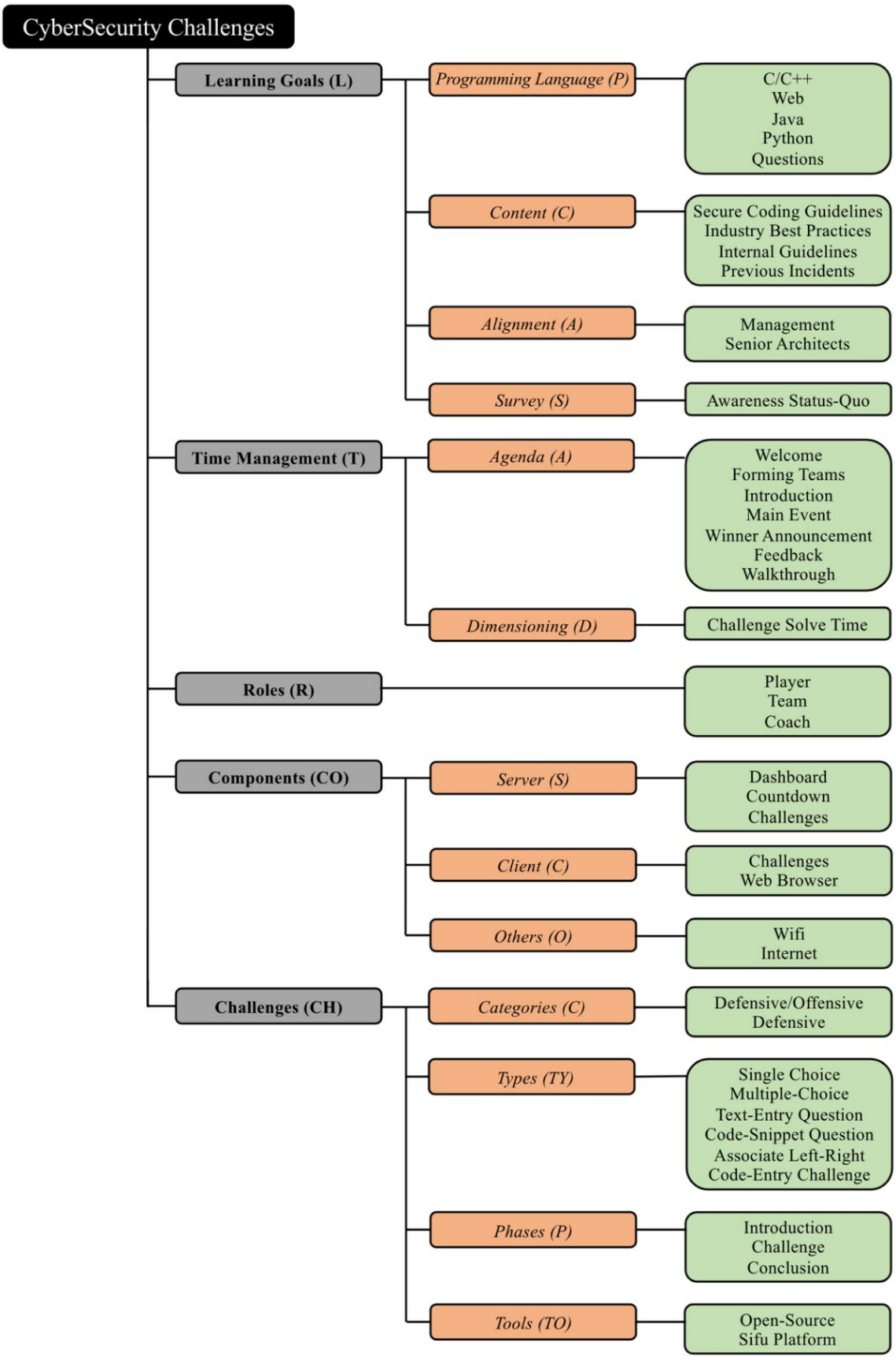


Fig. 8. Dimensioning CyberSecurity Challenges

A clear definition of roles in a serious game is also a critical aspect of such a game's design. The CyberSecurity Challenges game defines three roles: individual player, team, and coach. These games are typically deployed in a computer network infrastructure. Therefore, the different components present in the network and their management are also essential aspects of the game. Finally, the aspect challenges (CH) looks at the different categories of challenges (as introduced before), challenge types suitable for the industry, the different phases of a challenge, and tools to create the challenges. Detailed discussions on each of these aspects can be found in [10–20].

5. Conclusions

If not addressed appropriately, software vulnerabilities can result in serious negative consequences. A good time to address these issues is in the early stages of software development by raising the awareness of software developers on secure coding. This paper presents CyberSecurity Challenges (CSC) as a possible solution. CyberSecurity Challenges is a genre of serious games developed to raise the awareness of industrial software developers on secure coding and secure coding guidelines. CSC games have been developed since 2017 in the industry. They were extensively studied as part of the Ph.D. research by the first author, resulting in more than ten publications. The CSC game can be used both for onsite training and remote training, thus easily adapting to possible travel restrictions imposed by the current COVID-19 situation.

Our results, through empirical studies, show that this game is adequate to raise secure coding awareness, both when using defensive/offensive challenges and purely defensive challenges. Furthermore, preliminary results indicate that the same artifact could be used in academia to prepare the future industry workforce. Feedback obtained from software developers in the industry also indicates this community's acceptance and welcoming of the game. During gameplay, software developers have fun and practice the usage secure coding guidelines for secure software development. Furthermore, CSC games found additional success by being well accepted by management. We conclude that, by raising awareness, this type of game is a viable approach to tackle possible software vulnerabilities due to bad code quality (e.g., vulnerable code).

Acknowledgements

The authors would like to thank the participants of the CyberSecurity Challenges for their time and their valuable answers and comments. The authors would also like to thank Kristian Beckers and Thomas Diefenbach for their helpful, insightful, and constructive comments and discussions.

This research is partially financed by national funds through FCT - Fundação para a Ciência e Tecnologia, I.P., under the projects FCT UIDB/04466/2020 and UIDP/04466/2020. Furthermore, the third author thanks the Instituto Universitário de Lisboa and ISTAR, for their support.

References

[1] Assal, H., Chiasson, S.: 'Think secure from the beginning' A Survey with Software Developers. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–13. CHI '19, Association for Computing Machinery, New York, NY, USA (2019)

[2] Beckers, K., Pape, S.: A Serious Game for Eliciting Social Engineering Security Requirements. In: 2016 IEEE 24th International Requirements Engineering Conference (RE). IEEE (08 2016)

[3] Bundesamt für Sicherheit in der Informationstechnik: Baustein B 5.21 - Webanwendungen (2014), https://tinyurl.com/y25m2kxl

[4] Bundesamt für Sicherheit in der Informationstechnik: BSI IT-GrundschutzKatalog, 2016, 15. ed. (2016), https://tinyurl.com/zkbmfb6

[5] Carnegie Mellon University: Secure Coding Standards (2019), https://tinyurl.com/y29mwsyj, online

[6] Davis, A., Leek, T., Zhivich, M., Gwinnup, K., Leonard, W.: The fun and future of CTF. 2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14) pp. 1–9 (2014), https://tinyurl.com/y97enbtr

[7] ENISA: The cost of incidents affecting CIIs (8 2016), https://tinyurl.com/y3v4rv8x

[8] Frey, S., Rashid, A., Anthonysamy, P., Pinto-Albuquerque, M., Naqvi, S.A.: The Good, the Bad and the Ugly: A Study of Security Decisions in a Cyber-Physical Systems Game. IEEE Transactions on Software Engineering 45(5), 521–536 (2019)

[9] Gasiba, T.: Sifu Platform (12 2020), https://github.com/saucec0de/sifu, Siemens AG, MIT License, online

[10] Gasiba, T., Beckers, K., Suppan, S., Rezabek, F.: On the Requirements for Serious Games Geared Towards Software Developers in the Industry. In: Damian, D.E., Perini, A., Lee, S. (eds.) Conference on Requirements Engineering Conference. pp. 286–296. IEEE, Jeju, South Korea (09 2019). https://doi.org/10.1109/re.2019.00038

[11] Gasiba, T., Hodzic, S., Lechner, U., Pinto-Albuquerque, M.: Raising Security Awareness using CybersecurityChallenges in Embedded Programming Courses. In: forthcoming (2021), in preparation

[12] Gasiba, T., Lechner, U.: Raising secure coding awareness for software developers in the industry. In: 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW). pp. 141–143. IEEE, Jeju, South Korea (09 2019). https://doi.org/10.1109/REW.2019.00030

[13] Gasiba, T., Lechner, U., Cuellar, J., Zouitni, A.: Ranking Secure Coding Guidelines for Software Developer Awareness Training in the Industry. In: Queirós, R., Portela, F., Pinto, M., Simões, A. (eds.) First International Computer Programming Education Conference (ICPEC 2020). OpenAccess Series in Informatics (OASIcs), vol. 81, pp. 11:1–11:11. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany (2020)

[14] Gasiba, T., Lechner, U., Pinto-Albuquerque, M.: Awareness of Secure Coding Guidelines in the Industry - A first data analysis. In: The 19th IEEE International Conference on Trust, Security and Privacy in Computing and Communications. IEEE, Online (12 2020), to appear

[15] Gasiba, T., Lechner, U., Pinto-Albuquerque, M.: Sifu - A CyberSecurity Awareness Platform with Challenge Assessment and Intelligent Coach. In: Cybersecurity Journal, Special Issue on Cyber-Physical System Security. SpringerOpen (12 2020). https://doi.org/10.1186/s42400-020-00064-4

[16] Gasiba, T., Lechner, U., Pinto-Albuquerque, M.: CyberSecurity Challenges for Software Developer Awareness Training in Industrial Environments. In: 16th International Conference on Wirtschaftsinformatik (2021), to appear

[17] Gasiba, T., Lechner, U., Pinto-Albuquerque, M.: Is Secure Coding Education in the Industry Needed? An Investigation Through a Large Scale Survey. In: 43rd International Conference on Software Engineering (2021), to appear

[18] Gasiba, T., Lechner, U., Pinto-Albuquerque, M., Porwal, A.: Cybersecurity Awareness Platform with Virtual Coach and Automated Challenge Assessment. In: 6th Workshop On The Security Of Industrial Control Systems & Of CyberPhysical Systems (CyberICPS). pp. 67–83. Springer, Cham, Online (12 2020). https://doi.org/978-3-030-64330-0 5

[19] Gasiba, T., Lechner, U., Pinto-Albuquerque, M., Zouitni, A.: Design of Secure Coding Challenges for Cybersecurity Education in the Industry. In: 13th International Conference on the Quality of Information and Communications Technology. pp. 223–237. Springer, Online (09 2020). https://doi.org/978-3-030-58793-2 18

[20] Gasiba, T., Lechner, U., Rezabek, F., Pinto-Albuquerque, M.: Cybersecurity Games for Secure Programming Education in the Industry: Gameplay Analysis. In: Queirós, R., Portela, F., Pinto, M., Simões, A. (eds.) First International Computer Programming Education Conference (ICPEC 2020). OpenAccess Series in Informatics (OASIcs), vol. 81, pp. 10:1–10:11. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany (2020)

[21] Graziotin, D., Fagerholm, F., Wang, X., Abrahamsson, P.: What happens when software developers are (un)happy. Journal of Systems and Software 140, 32–47 (2018)

[22] Hänsch, N., Benenson, Z.: Specifying IT security awareness. In: 25th International Workshop on Database and Expert Systems Applications, Munich, Germany. pp. 326–330. IEEE, Munich, Germany (Sep 2014). https://doi.org/10.1109/DEXA.2014.71

[23] IEC 62443-4-1: Security for industrial automation and control systems - part 4-1: Secure product development lifecycle requirements. Standard, International Electrotechnical Commission (01 2018)

[24] ISO 27001: Information technology – Security techniques – Information security management systems – Requirements. Standard, International Standard Organization, Geneva, CH (10 2013)

[25] McIlwraith, A.: Information Security and Employee Behaviour: How to Reduce Risk Through Employee Education, Training and Awareness. Gower Publishing, Ltd. (2006)

[26] Moody, G.D., Siponen, M., Pahnila, S.: Toward a Unified Model of Information Security Policy Compliance. MIS quarterly 42(1), 1–50 (2018)

[27] OWASP Foundation: Open Web Application Security Project, https://owasp.org/

[28] Patel, S.: 2019 Global Developer Report: DevSecOps finds security roadblocks divide teams (July 2020), https://about.gitlab.com/blog/2019/07/15/globaldeveloper-report/, online; posted on July 15, 2019

[29] Rieb, A.: IT-Security Awareness mit Operation Digitales Chamäleon. Ph.D. thesis, Universität der Bundeswehr München, Neubiberg (2018)

[30] Schneier, B.: Software Developers and Security. Online (July 2020), https://www.schneier.com/blog/archives/2019/07/software develo.html

[31] Stewart, G., Lacey, D.: Death by a Thousand Facts: Criticizing the Technocratic Approach to Information Security Awareness. Information Management & Computer Security 20(1), 29–38 (2012)

[32] Tahaei, M., Vaniea, K.: A Survey on Developer-Centred Security. In: 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). pp. 129–138. IEEE (2019)

[33] Xie, J., Lipford, H.R., Chu, B.: Why do Programmers Make Security Errors? 2011 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) pp. 161–164 (09 2011). https://doi.org/10.1109/VLHCC.2011.6070393

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Cyber Resilience im digitalen Wandel – Strategische und operative Lösungsansätze

Wolfram Girg[1]

Kurzfassung:
Zunehmende Cyber-Gefahren begleiten Unternehmen in die digitale Transformation. Punktuelle Cyber-Sicherheit wird nicht mehr ausreichend sein, um Unternehmen vollumfänglich abzusichern. Es sind sowohl die technischen und organisatorischen Grundlagen als auch die Herausforderungen in der eigenen Organisation zu berücksichtigen, damit eine optimale Wirksamkeit der Schutzmaßnahmen erzielt wird. Die Cyber Resilience, eine Kombination aus Sicherheit und Resilienz im digitalisierten Umfeld, kann hierzu als Grundlage für ein beständiges und erfolgreiches Business etabliert werden. Dabei ist es hilfreich zu verstehen, welchen Cyber-Risiken wir ausgesetzt sind und was notwendig ist, um die digitale Widerstandsfähigkeit zu verbessern.
Zunächst ist es empfehlenswert, neben den erforderlichen organisatorischen Rollen und Verantwortlichkeiten detaillierte Informationen aller kritischen Unternehmenswerte zusammenzustellen und vorhandene Maßnahmen aus unterschiedlichen Sicherheitsperspektiven einzubeziehen. Eine solche Übersicht dient nach einer Risikobetrachtung als Basis für weitere Entscheidungen und Maßnahmen. Die verwendeten Werkzeuge sollten strukturierte Ergebnisse liefern und Aufschluss über den aktuellen Cyber-Reifegrad der Organisation geben. Weitere Orientierung bietet die Leitfrage: Welche Schwerpunkte sind im Anschluss individuell für das entsprechende Unternehmen zu priorisieren? Die Optimierung der technischen Sicherheit sollte pragmatisch und praxisnah erfolgen. Die Schwerpunkte auf der organisatorischen Seite umfassen das Ausmerzen prozessualer Schwachstellen in Anlehnung an bestehende Sicherheitsstandards und unter Berücksichtigung des „Faktors Mensch“.

Stichworte: Cyber-Resilience, ISMS, Digitalisierung, NIST Cyber Security Framework (CSF), Prozesse, Reifegrad, Resilience, Risiken, Schwachstellen-Management, SIEM

1. Die Bedeutung der Cyber-Sicherheit

Das Schlagwort Digitalisierung ist in aller Munde. Neben den zunehmenden gesetzlichen Anforderungen heizen die notwendigen Innovationen die Digitalisierung eines Unternehmens ebenfalls an. Dadurch wird eine grundlegende Abhängigkeit in die Sicherheit und Resilienz der digitalen Prozesse, der Technik sowie der ausführenden und verantwortlichen Personen erzeugt. In der Regel gehen in Unternehmen zukunftsorientierte Persönlichkeiten wie der Chief Executive Officer (CEO), der Chief Information Officer (CIO), oder sofern etabliert der Chief Digital Officer (CDO) die digitale Transformation an. Zudem ist der Grad der anzustrebenden Digitalisierung bedingt durch die Bedeutung und Abhängigkeit der dringend benötigten Technik. Ist das Unternehmen eher mit verwaltenden Tätigkeiten betraut oder sogar aktiv an kritischen technologischen Entwicklungen beteiligt, ergeben sich hierdurch Änderungen an dem Unternehmens-Risikoprofil. Die digitale Umgebung wird unabhängig von der Ausrichtung des Unternehmens immer mehr zum sensiblen und kritischen Nervensystem und rückt dadurch weiter in den Geschäftsmittelpunkt. Es entstehen neue Abläufe und Kommunikationskanäle, die

[1] Controlware GmbH

Geschäftsentwicklung soll schneller vonstattengehen – mit den zusätzlichen Herausforderungen paralleler Kostensenkung. Auch wenn nicht alle Anforderungen der Digitalisierung neu sind oder tiefgreifende Änderungen mit sich bringen, bedeutet die provozierte Abhängigkeit von digitalen Prozessen weitere Risiken und darf nicht unterschätzt werden. Eine existenzielle Voraussetzung für ein funktionierendes Geschäft ist nach wie vor das durch die Kunden entgegengebrachte Vertrauen. Auf dem Spiel stehen nicht nur die zwingend einzuhaltende Vertraulichkeit, sondern auch die Sicherstellung der notwendigen Integrität. Die zugesagte Verfügbarkeit von Infrastruktur, Anwendungen, Prozessen und Produkten darf ebenfalls nicht vergessen werden.

1.1. Mythen der Cyber-Sicherheit

An vielen Stellen reagieren Unternehmen mit einer gewissen Distanz auf die notwendigen Veränderungen digitaler Projektvorhaben und scheuen sich, die zwingend erforderlichen Sicherheitsaspekte von Beginn an einzubeziehen.

- „Wir fallen nicht auf – bei uns gibt es nichts zu stehlen."

  An erste Stelle sollte immer die Wertschöpfung des eigenen Unternehmens gestellt werden. Dabei ist die Firmengröße eher sekundär. Die oftmals verlautete Vermutung, dass es bei kleineren Unternehmen „weniger zu stehlen gibt", ist nicht passend, sondern abhängig von den Unternehmenswerten sehr individuell. An dieser Stelle darf nicht vergessen werden, dass größere Unternehmen auch entsprechend mehr in Ressourcen – sei es finanzieller oder personeller Art –, in die zu schützende IT und in die Informations- bzw. Cyber-Sicherheit investieren als dies kleineren Firmen möglich ist. Die Etablierung grundlegender Sicherheitsmaßnahmen, sowohl technischer als auch organisatorischer Art, scheint unabhängig von der Unternehmensgröße angebracht.

- „Wir haben doch schon die neueste Technologie."

  Cyber-Sicherheit bedeutet nicht nur Technologie, sondern erzwingt ein Zusammenspiel von Technik, Mensch und Prozessen. Dies bedingt allerdings auch für jeden Teilbereich die Auseinandersetzung mit den jeweiligen individuellen Schwachstellen. Dabei sind nicht nur die Technik-IT-Einheit oder die allerneuesten Technologien gefragt: Jeder Mitarbeiter des Unternehmens ist aufgefordert, an der Sicherheit mitzuwirken.

- „Bei uns sind die Daten doch sicher."

  Weiterhin ist ein alleiniger Schutz der Verbindungen innerhalb der eigenen Unternehmensgrenzen aufgrund der vielfältigen Schnittstellen in andere Unternehmen, zu Partnern, Kunden und Cloud-Providern keine angemessene Strategie für die Cyber-Sicherheit. Gerade bei der Nutzung von Infrastruktur, Anwendungen oder Rechnerkraft von Cloud-Providern hängt ein sicherer Gebrauch dieser Dienstleistungen unter anderem von einer durchdachten IT-Strategie und einem pragmatischen Cloud-Nutzungs- und Umsetzungskonzept ab. Ein Teil eines Gesamtkonzeptes der Cyber-Sicherheit besteht darin, alle kritischen, zulässigen

Verbindungen der IT unter Kontrolle zu bekommen, was entsprechende Priorisierungen erfordert.

1.2. Die Gefahr von Cyber-Angriffen und die aktuelle Gefährdungslage

Die Frage zur derzeitigen Gefährdungslage bzw. zu der Wahrscheinlichkeit, dass das eigene Unternehmen Opfer eines Cyber-Angriffs wird, ändert sich immer häufiger von „ob“ zu „wann“. Die aktuelle Gefährdungslage rund um die Cyber-Sicherheit ist angespannt. Mittlerweile wird es immer wahrscheinlicher, große Schäden aus einem Cyber-Angriff davonzutragen. Dies belegen viele aktuelle Untersuchungen wie beispielsweise der Cyber Resilient Organization Report von IBM aus dem Jahr 2020[2]. In dieser weltweit angelegten Studie wurden insgesamt circa 3.500 IT- und Security Experten befragt: 51 Prozent der Befragten gaben an, zwischen 2018 und 2020 durch einen Cyber-Angriff erheblich gestört worden zu sein. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) beschreibt in seinem Lagebericht 2020[3] den Schwerpunkt der Angriffe aus dem digitalen Raum im Bereich der Cyber-Kriminalität. Es stehen aber nicht nur Konzerne oder größere Unternehmen im Fokus der Cyber-Kriminellen: Die Angreifer haben zum großen Teil finanzielle Beweggründe, sodass auch kleine und mittelständische Unternehmen (KMU) in der Regel nicht verschont bleiben. 2020 wurde im Lagebericht des BSI über einen täglichen Zuwachs von neuen Schadcode-Varianten berichtet. Explizit erwähnt werden ernste Bedrohungen durch „Ransomware“-Angriffe: Dabei verhindert eine erfolgreiche Attacke mit dieser Schadsoftware den Zugriff auf lokale oder im Netzwerk erreichbare Daten und Systeme und verschlüsselt diese. Im Anschluss wird durch den Angreifer ein Lösegeld gefordert, um die Daten wieder zu entschlüsseln. Zusätzlich werden die Unternehmen unter Zeitdruck gesetzt: Oft wird mit einer Erhöhung des Lösegeldes oder einer dauerhaften Verschlüsselung der Daten gedroht, sofern die Zahlung nicht innerhalb eines kurzen Zeitfensters erfolgt. Ein weiteres eindrucksvolles Detail: Laut der IBM-Studie haben nur rund 45 Prozent der befragten Unternehmen die Gefährdung „Ransomware“ in ihren Notfallplänen berücksichtigt. Dies ist äußerst bedenklich, da die Dunkelziffer mutmaßlich weitaus größer ist. Die Erhebung basiert auf den Angaben von Unternehmen, die bereits einen Notfallplan haben. Das Vorhandensein eines Notfallplans ist ein elementarer erster Schritt, mit Angriffen effektiv umzugehen. Dies zwingt Unternehmen dazu, Gefährdungsszenarien zu betrachten, und stärkt den Aspekt der Geschäftsfortführung in einem Notfall oder einer Krise. Um einen Notfallplan in die Tat umzusetzen, sind allerdings noch weitere Anstrengungen zur maßgeschneiderten Operationalisierung erforderlich, die durch das obere Management gewürdigt und unterstützt werden müssen.

In 2020 führt das Allianz Risk Barometer 2020[4] Cybervorfälle nicht nur als das Top-Risiko weltweit, sie finden sich neben Deutschland auch in vielen anderen der untersuchten Länder unter den ersten drei Risiken. Bereits 2019 waren Angriffe über den digitalen Raum die am meisten gefürchteten Auslöser von Betriebsunterbrechungen. Als

[2] https://www.ibm.com/security/digital-assets/resilient/cyber-resilient-organization-repor t
[3] https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2020.pdf
[4] https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/Allianz-Risk-Barometer-2020.pdf

mögliche Ursachen werden technisches Versagen oder Fehler durch Mitarbeiter, Bedrohungen von Hackern, Ransomware sowie DoS-Angriffe (Denial of Service) genannt. Häufigste Auslöser sind laut des Barometers Datenklau bzw. die damit verbundenen IT-Sicherheitsvorfälle. Diese Bandbreite unterschiedlicher Gefährdungskategorien, der Anstieg von Angriffen sowie aktualisierte Angriffsmethoden verdeutlichen die zwingende Notwendigkeit für Unternehmen, das Thema Cyber-Sicherheit ernst zu nehmen.

Doch wie kann wirtschaftlich effizient, aber dennoch dem Risikoprofil angemessen vorgegangen werden? Im nächsten Abschnitt werden die fachlichen Schwerpunkte der Cyber-Sicherheit und -Resilienz gegenübergestellt und es wird mit den weiteren angrenzenden Themengebieten ein Gesamtbild zur Einsortierung der Cyber Resilience erschaffen.

2. Einsortierung der Cyber Resilience

Für viele Wirtschaftsunternehmen gehört das Management von Risiken aus unterschiedlichen Fachbereichen zur Tagesordnung, es ist ein grundlegender Teil angemessener Geschäftspraxis. Auch wenn der Vorgang in unregulierten Branchen meist in einer weniger formalen Form stattfindet, sind Finanz- und Reputationsrisiken sowie Informationssicherheits- und IT-Risiken doch oftmals feste Bestandteile des unternehmensweiten Risiko-Managements. Mitunter gibt es allerdings für die bestehenden Risikoarten eines Unternehmens oder einzelner Abteilungen unterschiedliche Ansätze und Methoden zur Bewertung bzw. Konsolidierung von einzelnen Risiken, die eine gewisse Unschärfe in der Gesamtsicht mit sich bringen. Eine echte Herausforderung, jedoch nicht Bestandteil dieser Ausarbeitung, ist die Notwendigkeit, die Risiken im Geschäftskontext zusammenzuführen, um eine Übersicht über die Gesamt-Risiko-Lage darstellen zu können. Die Fähigkeit von Unternehmen, sich auf schadenstiftende Cyber-Vorfälle einzustellen, diesen entgegenzuwirken und schnellstmöglich wieder in den Normalbetrieb zu kommen, wird im Allgemeinen als „Cyber Resilience“ bezeichnet. Die ursprüngliche Bedeutung „Resilienz“ kann als „Flexibilität“, aber auch als „Widerstandsfähigkeit“ und „Robustheit“ verstanden werden. Der Schwerpunkt eines sicheren und gleichzeitig resilienten Unternehmens liegt demnach auf einer Mischung von Sicherheitsmechanismen, die vorwiegend aus der Informationssicherheit stammen, kombiniert mit Methoden und Maßnahmen aus dem Bereich der Geschäftsfortführung (Business Continuity). Mit ausgewählten Maßnahmen kann somit die Basis für einen anzustrebenden Sicherheitsreifegrad festgelegt werden. Der Fokus aus Sicherheitsperspektive umfasst die Verteidigung, den präventiven Schutz und die Absicherung kritischer Unternehmenswerte und der zugehörigen IT-Landschaft sowie die mögliche Erkennung von Angriffen. Das Sicherheitsziel besteht demnach darin, Risiken und das Eintreten von Schäden zu reduzieren bzw. idealerweise sogar ganz oder teilweise zu vermeiden. Das unangefochtene Ziel der Geschäftsfortführung ist die schnellstmögliche Wiederherstellung des Betriebs. Da jegliche Anstrengungen zur Sicherheit als zeitlich begrenzt und in Bezug auf die aufzubringenden Kosten gegenüber ihrem letztendlichen Nutzen auch relativ anzusehen sind, kann eine hundertprozentige Sicherheit nicht wirtschaftlich sein. Bei einem eingetretenen Sicherheitsvorfall bleibt meistens wenig Zeit, um über die unmittelbar notwendigen

nächsten Schritte nachzudenken. Notfälle, zu denen auch Cyber-Angriffe führen können, sollten daher proaktiv vor dem Eintritt geplant und strukturiert werden. Dies umfasst das Bewusstsein der eigenen Verletzlichkeit, indem auch etablierte Sicherheitsmechanismen versagen können. Der Fokus von Aktivitäten rund um die Geschäftsfortführung beinhaltet unter anderem die Reaktion auf Sicherheits- oder Cybervorfälle sowie die Wiederherstellung der Geschäftstüchtigkeit nach einem Vorfall.

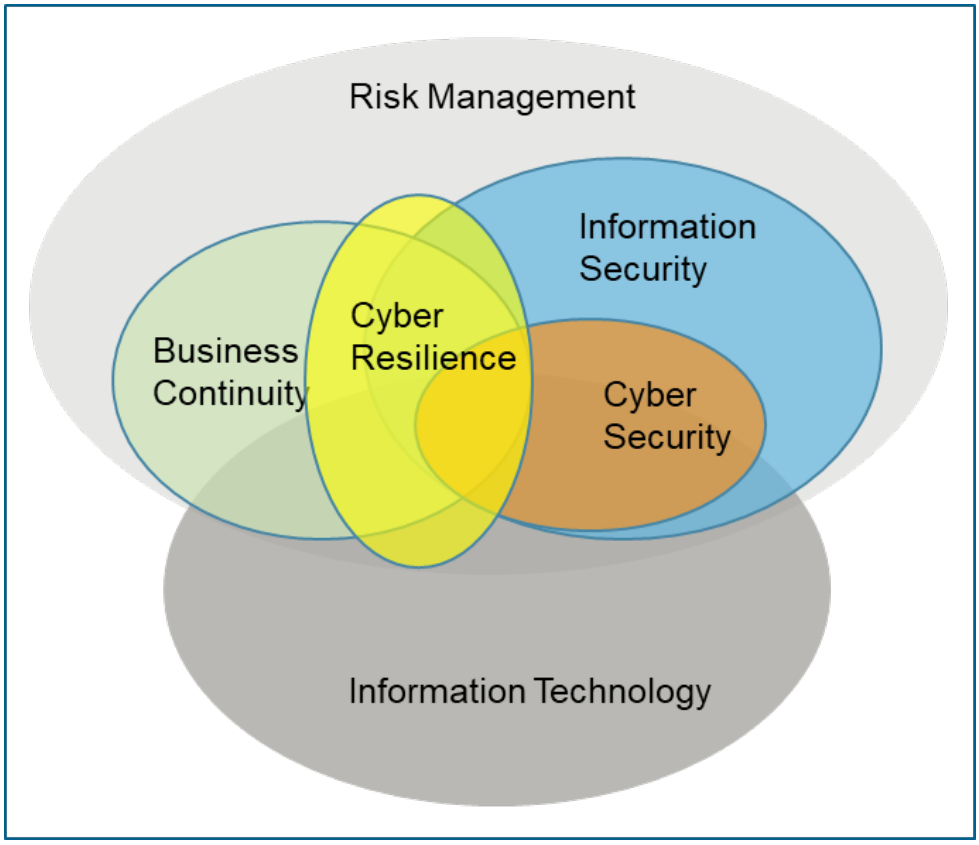


Abbildung 1: Einsortierung der Cyber Resilience

3. Grundlagen der Cyber Resilience

Nach der Einsortierung und Definition der (Cyber-)Sicherheit und Resilienz stellt sich die Frage nach einem möglichen Vorgehensmodell für eine Bestandsaufnahme des eigenen Unternehmens. Die gute Nachricht: Das Rad muss an dieser Stelle nicht neu erfunden werden. Es sind bereits standardisierte Ansätze verfügbar, die ein Lagebild zur Cyber-Sicherheit für ein Unternehmen effektiv ermöglichen. Näher eingegangen wird hierzu auf den Leitfaden zur Durchführung eines „Cyber-Sicherheits-Checks[5]“, der in Zusammenarbeit des BSI mit dem Berufsverband der IT-Revisoren, Wirtschaftsprüfer sowie Experten der Informationssicherheit und IT-Governance (ISACA) erstellt wurde. Zur Festlegung des Beurteilungsgegenstandes ist das komplette Unternehmen angedacht, inklusive Anbindungen an Partner, Dienstleister und Kunden. Die physischen Si-

[5] https://www.bsi.bund.de/SharedDocs/Downloads/ACS/leitfaden_CSC_v2.html

cherheitsaspekte und spezielle Sicherheitserfordernisse für Produktionsanlagen, Komponenten und Systeme der Operation Technologie (OT) stehen weniger im Fokus. Zunächst wird in dem Vorgehensmodell ermittelt, wie exponiert sich das Unternehmen gegenüber Cyber-Gefahren darstellt. Die Bandbreite möglicher Bedrohungen reicht von sogenannten Skript-Kiddies bis hin zu gezielter kriminell-organisierter Spionage. Selbstverständlich können bereits vorhandene Risikobetrachtungen einbezogen werden, sofern diese auf der klassischen Kombination von Schadenshöhe und Eintrittswahrscheinlichkeit basieren. Mithilfe dieser rudimentären Schutzbedarfs- und Risikoeinschätzung wird ein Snapshot der aktuellen Sicherheitslage erstellt. Dabei werden sowohl bestehende Dokumentationen gesichtet als auch Gespräche mit den jeweiligen Fachexperten sowie dem zuständigen Management geführt. Das abschließende Berichtsdokument beinhaltet neben den bereits erwähnten Aussagen zur Risikosituation und der Feststellung des konsolidierten Cyber-Sicherheitsstatus ebenfalls eine detaillierte Gegenüberstellung der verbindlich geforderten Maßnahmenziele. Dazu gehören auszugsweise die Absicherung von Netzübergängen, die Abwehr von Schadprogrammen, die Inventarisierung von IT-Systemen, die Vermeidung von ausnutzbaren Sicherheitslücken sowie die Logdaten-Erfassung und -Auswertung zur Erkennung aufgetretener Sicherheitsvorfälle. Insgesamt werden 14 Kategorien betrachtet, gegenüber den Basis-Maßnahmenzielen bewertet und in einem übersichtlichen Abschlussbericht dokumentiert. Dieses Vorgehen ermöglicht es, den aktuellen Maßnahmenstand der Cyber-Sicherheit den grundlegenden Vorschlägen zur Absicherung gegenüberzustellen.

4.1. Optimierung des technischen Sicherheitsstands

Die Ergebnisse eines durchgeführten Cyber-Sicherheits-Checks bilden das Sicherheitsniveau zum Zeitpunkt der Überprüfung ab und geben eine mögliche Richtung für die Weiterentwicklung sicherheitsrelevanter Maßnahmen vor. Je nach aktuellem Reifegrad und Risikoprofil besteht neben den Umsetzungsvorgaben von Basis-Maßnahmenzielen zusätzlich die Möglichkeit, die IT-Landschaft weiter zu optimieren. Dies bietet sich insbesondere für die hoch schützenswerten Unternehmenswerte im Gesamtkontext (Technik, Mensch, Prozesse) an.

Schwachstellenmanagement

Aus technischer Sicht werden beispielsweise frühzeitige IT-Schwachstellenerkennung und -behebung als Schlüsselkomponenten zur proaktiven Bekämpfung von Cyber-Angriffen und zur Absicherung der IT-Infrastruktur angesehen. Als wesentliches Einfallstor gilt hierbei die Vielzahl an Schwachstellen durch häufig eingesetzte Hard- und Software. Verwundbare Systeme ohne etabliertes Schwachstellen-Management werden in der Regel nicht rechtzeitig identifiziert und daher auch nicht hinreichend gepatcht. Vor allem in Bezug auf das Maßnahmenziel einer Inventarisierung wird bei der großen Menge an IT-Komponenten zunehmend eine Automatisierung benötigt, um den Bestand der Hard- und Software, die eingesetzten Versionen und Patch-Stände von Betriebssystemen sowie Anwendungen zu identifizieren. Ein dynamisches IT-Inventar ermöglicht zudem eine Übersicht von autorisierten Geräten. Denn als weitere wesentliche Herausforderung gilt die Zunahme einer unkontrollierten Schatten-IT. So sind neben der offi-

ziellen IT-Infrastruktur immer häufiger weitere Geräte innerhalb des Unternehmensnetzwerks zu finden, die weder technisch noch organisatorisch durch die zentrale IT betreut werden. In der Konsequenz stellen diese Geräte ein enormes Sicherheitsrisiko für die gesamte IT-Landschaft dar. Ein initialer Schwachstellen-Scan ist in der Lage, viele dieser „schwarzen Schafe" zu erkennen. Bekannte Sicherheitslücken können entweder kurzfristig durch Workarounds oder bereitgestellte Sicherheitsaktualisierungen geschlossen werden. Idealerweise findet allerdings eine Überprüfung auf Schwachstellen regelmäßig in Form eines definierten Security Services statt. Die Etablierung eines Schwachstellen-Management-Service erhöht das Sicherheitsniveau der gesamten IT-Landschaft nachhaltig, da so Assets fortwährend erkannt und vorhandene Schwachstellen reduziert werden.

Security Incident & Event Management (SIEM)

Ein weiterer Sicherheitsaspekt zur Vervollständigung und Optimierung des Maßnahmenziels der Logdatenerfassung und -auswertung ist die Implementierung eines Security Incident & Event Management (SIEM). Insbesondere unter Berücksichtigung der laufenden Überarbeitung des IT-Sicherheitsgesetzes[6] findet sich dort auch ein angestrebter, verpflichtender Einsatz von „Systemen zur Angriffserkennung". In der aktuell vorliegenden Bundestagsfassung[7] (Stand Januar 2021) wird die Kontinuität und die Automatisierung von geeigneten Parametern und Merkmalen des laufenden Betriebs angeführt. Des Weiteren sollen die vorgesehenen Systeme in der Lage sein, fortwährend Bedrohungen zu identifizieren und zu vermeiden. Grundsätzlich können SIEM-Systeme sowohl unautorisierte Zugriffe schnell und effektiv erkennen als auch die Auswirkungen eines erfolgreichen Angriffs bewerten und Folgeschäden minimieren. Diese kontinuierliche Prüfung auf mögliche Attacken ist ein wichtiger Sicherheitsgewinn, speziell bei Angreifern, die sich viel Zeit lassen und vorsichtig, gegebenenfalls über längere Zeiträume auch minimalinvasiv, tätig werden. Durch diese Angriffstechnik kann auch bei länger andauernden Zeitintervallen die Kontrolle über Systeme erlangt werden. Daher ist es sinnvoll, Verfahren zur Aufdeckung von nicht offensichtlichen Sicherheitsvorfällen und langfristig angelegten Angriffen zu etablieren. Neben der reinen Daten- und Logsammlung diverser Systeme ist es erforderlich, die SIEM-Plattform anzulernen und mit sogenannten Use Cases zu füttern. Damit ist das System in der Lage, das Angreiferverhalten und die eingesetzten Techniken in den unterschiedlichen Phasen eines Angriffs zu erkennen und in Zusammenhang zu bringen. Der qualifizierten Erstellung von Use Cases kommt in Bezug auf die Bewertung und den späteren Betrieb der Lösung eine besondere Bedeutung zu. Um aus den riesigen Datenmengen die „korrekten" Ereignisse herauszufiltern, ist ein kluges Vorgehen bei der Auswahl und Implementierung von Lösungen sowie das nötige Expertenwissen wichtig. Unzureichend präzise Use Cases können dazu führen, dass Zeit mit „falsch" erkannten Angriffen vergeudet wird. Auch wenn die technische Expertise an dieser Stelle eine große Rolle spielt, dürfen die

[6] https://www.bmi.bund.de/SharedDocs/gesetzgebungsverfahren/DE/entwurf-zweites-it-sicherheitsgesetz.html
[7] https://ag.kritis.info/wp-content/uploads/2021/01/20210101-IT-Sicherheitsgesetz-2.0.pdf

notwendigen prozessualen Anpassungen und die menschlichen Aspekte keinesfalls ignoriert werden.

4.2. Optimierung der organisatorischen Sicherheit

Prozessuale Unterstützung durch ein Framework
Die Ergebnisse einer durchgeführten und auch aktualisierten Bestandsaufnahme in der Cyber-Sicherheit sollen effizient weiterverarbeitet werden. Wichtig sind dabei die Auswahl und Etablierung eines standardisierten Vorgehens (Framework) zur regelmäßigen Erhebung des Cyber-Sicherheitsniveaus. Grundlage ist ein klares Verständnis der Bedrohungsszenarien, der operativen und technischen Prozesse sowie der eingesetzten Technologien. Eine festgelegte Durchführungspraxis mit wiederholbaren Ergebnissen schafft in der organisatorischen Sicherheit entscheidende Mehrwerte und ermöglicht beispielsweise eine risikoorientierte Priorisierung aus den identifizierten Mängeln, um das vorhandene Budget und die Ressourcen optimal einzusetzen. Mithilfe eines erweiterten Reifegrad-Modells lassen sich darüber hinaus bei einer regelmäßigen Durchführung auch entsprechende Kennzahlen für die Messbarkeit und für eine bessere Transparenz und Wirksamkeit von Maßnahmen für die Cyber-Sicherheit ableiten. Die bereits vorhandenen und zukünftig noch festzulegenden Sicherheitsmaßnahmen sind häufig mehrschichtig und komplex. Daher müssen diese gemäß der Risikosituation unter Berücksichtigung bestehender Vorgaben und Anforderungen betrachtet werden.

Orientierung an Standards
Unternehmen können sich an nationalen und internationalen Standards der Informationssicherheit und Geschäftsfortführung wie dem BSI 200-x, ISO/IEC 27001, ISO/IEC 22301 und dem NIST Cybersecurity Framework (CSF) orientieren. Im Anschluss an eine Analyse zur Cyber-Sicherheit, beispielsweise mithilfe des „Cyber-Sicherheits-Checks“, kann in Ergänzung mit den erwähnten Standards auf ein angepasstes Managementsystem der Cyber- oder Informationssicherheit hingearbeitet werden.

Besonders für den Aufbau oder die Weiterentwicklung der Cyber Resilience ist eine Kombination von Information Security Management (ISM) und Teilen aus dem Business Continuity Management (BCM) empfehlenswert. Beide Ansätze können kombiniert abgebildet werden und zu einem einheitlichen Sicherheits- und Notfallverständnis beitragen. Das NIST Cyber Security Framework (CSF)[8] verdeutlicht diesen Ansatz durch die Einteilung von Sicherheitsmaßnahmen in sogenannte Funktionen. Es handelt sich dabei um ein Rahmenwerk aus den USA zum Umgang mit Cyber-Risiken, das durch die Zusammenarbeit zwischen Industrie und Regierung geschaffen wurde. Die Anforderungen sind mit einer generischen Grundstruktur versehen und ermöglichen technologie- und branchenübergreifend ein gemeinsames Verständnis.

Die Funktionen konsolidieren darin die folgenden Inhalte (Auszug):

[8] https://www.nist.gov/cyberframework

- Identifizieren – Herstellen eines organisatorischen Verständnisses. Dies beinhaltet Regelwerke, Vorgaben und Governance, die Identifikation geschäftskritischer Assets, Systeme, Daten und Perspektiven des Risikomanagements (strategisch und operativ).
- Schützen – Schutzmaßnahmen zur sicheren Bereitstellung der Services. Umfasst sowohl die technologischen Maßnahmen als auch die Sicherheitsprozesse sowie Anforderungen an die benötigte Weiterbildung und Awareness-Maßnahmen.
- Erkennen – Identifizieren von Cyber Security Events. Neben den zugehörigen Prozessen beinhaltet das Modul mögliche Techniken zur Erkennung von Anomalien und ein kontinuierliches Überwachen der Wirksamkeit von Schutzmaßnahmen.
- Reagieren – rechtzeitiges Agieren und Reagieren. Die Reaktion auf Cyber Security Events durch angemessene Maßnahmen zur Eindämmung des Schadens.
- Wiederherstellen – die Fähigkeit, widerstandsfähig zu bleiben und die notwendigen Abläufe im Unternehmen so schnell wie möglich wiederherzustellen.

Organisationsstruktur & „Faktor Mensch“

Die notwendige Struktur der Cyber-Sicherheit wird durch eine Sicherheitsorganisation im Unternehmen festgelegt, die klare Aufgaben, Verantwortungen und Kompetenzen definiert. Die Sicherheitsorganisation unterstützt die Geschäftsleitung, alle notwendigen Ressourcen bereitzustellen. Zur Erfüllung der anstehenden Aufgaben – in Zusammenarbeit mit unterschiedlichen Unternehmensbereichen – müssen auch die Kompetenzen konkret benannt und entsprechend aufgebaut werden. Dabei spielen nicht nur die reinen Prozess- oder Technik-Optimierungen eine Rolle, sondern auch die ausführenden Personen – also der „Faktor Mensch“. Die Mitarbeiter sind ein elementarer Baustein bei der Implementierung einer angemessenen Cyber-Strategie. Durch menschliche Fehlhandlungen, sei es vorsätzlich oder unbeabsichtigt, können technische Vorkehrungen ausgesetzt und damit wirkungslos werden. In diesem Zusammenhang sind je nach aktuellem Reifegrad Maßnahmen sinnvoll, die aus einem Awareness-Konzept abgeleitet werden. Unter Einbeziehung bestehender Vorgaben und Schulungen beinhaltet ein solches Konzept die Awareness-Ziele des Unternehmens hinsichtlich einer kulturellen Weiterentwicklung zum Thema Sicherheit. Die Individualisierung auf das eigene Unternehmen stellt das Erfolgsrezept dar. Es ist notwendig, komplex anmutende Regelwerke einfach zu erklären. Voraussetzung dafür ist allerdings eine inhaltliche Aktualität. Häufig sind weiterführende Erklärungen hinsichtlich Grund und Mehrwert der eingesetzten Sicherheitsmaßnahmen der Schlüssel, um die Aufmerksamkeit und das Verständnis der Mitarbeiter zu erhöhen. Dadurch können sie die Hintergründe der meist einschränkenden Maßnahmen nachvollziehen, sich beteiligen und damit den Grundstein einer wachsenden Sicherheitskultur legen. Klassische Formate wie E-Learnings, Intranetbeiträge, Mitarbeitermagazine, Plakate, Live-Hacking, Gamification oder Phishing-

Tests etc. können Mitarbeiter für die Cyber-Sicherheit begeistern und geben ihnen die Gelegenheit, sich an sichere Verhaltensweisen zu gewöhnen.

4. Fazit

Die Ausweitung der Digitalisierung hat zwangsläufig auch größere Angriffsflächen bekannter und neuer Risiken zur Folge. Die Cyber-Sicherheit ist eine wesentliche Voraussetzung für das Gelingen der digitalen Transformation und Grundlage zur Entwicklung einer dem Risikoprofil entsprechenden Widerstandsfähigkeit. Effektive Cyber Resilience besteht aus vielen Facetten mit organisatorischen sowie technischen Lösungsschwerpunkten, die in ein Gesamtkonzept überführt und umgesetzt werden müssen. Wichtig ist, dass sich nur durch das Zusammenwirken der eingesetzten Technik, der Prozesse und menschlichen Aspekte optimale Ergebnisse erzielen lassen. Bei der Konzeption eines wirkungsvollen Gesamtkonzeptes ist die professionelle Unterstützung von Experten ratsam, damit zukünftige Faktoren und Entwicklungen von Anfang an einbezogen werden.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Identitätsdiebe in die Schranken weisen - Account Security erhöhen mit Identity-Guard

Timo Malderle[1], Dr. Matthias Wübbeling[2], Pascua Theus[1], Prof. Dr. Michael Meier[2]

Kurzfassung:
Als Folge von digitalen Einbrüchen bei Online-Diensten werden deren Betreiber und ihre Nutzer regelmäßig Opfer krimineller Aktivitäten mithilfe von Identitätsdiebstahl. Die Abwicklung solcher Schadensfälle ist für die Dienstanbieter aufwendig und teuer. Die weitere Verfolgung eines einzelnen Falls lohnt sich daher häufig nur bei hohen Schadenssummen. Existierende Informationssysteme über Identitätsdiebstahl, die sich an die Nutzer von Online-Diensten wenden, setzen deren Bekanntheit und regelmäßige Verwendung voraus. Das Forschungsprojekt *Effektive Information nach digitalem Identitätsdiebstahl* (EIDI)[3] sucht und verarbeitet, unter Wahrung der Belange des Datenschutzes, öffentlich verfügbare Identitätsdaten-Leaks, um mithilfe dieser Daten Nutzer von Online-Diensten proaktiv zu warnen. Während der Projektlaufzeit konnten bereits mehr als 25 Milliarden Identitätsdaten gesammelt werden und über eine Million Benutzer gewarnt und so vor Identitätsdiebstahl geschützt werden.

Stichworte: Account Security, Credential-Stuffing, EIDI (Effektive Information nach digitalem Identitätsdiebstahl), Identitätsdiebstahl, Leakchecker, Passwörter, Zugangsdaten

1. Einleitung

Täglich etablieren sich neue und innovative Onlinedienste am Markt. So wächst die Anzahl verfügbarer Dienste kontinuierlich. Bei vielen Onlinediensten müssen sich Benutzer ein eigenes Benutzerkonto anlegen, um den Dienst nutzen zu können. Der Zugriff auf dieses Benutzerkonto wird in der Regel mit einer Kombination aus E-Mail-Adresse und Passwort abgesichert. Aufgrund der Vielzahl verfügbarer Onlinedienste sammelt jeder Benutzer im Laufe der Zeit eine größere Anzahl von Benutzerkonten. Da viele Benutzer mit dem eigenen Passwortmanagement bei der Menge an Zugangsdaten überfordert sind, verwendet ein Großteil der Benutzer zur Authentifikation bei unterschiedlichen Onlinediensten häufig dieselben Anmeldeinformationen.

Die gesamten Mengen an Benutzerkonten und die Mehrfachverwendung von Passwörtern weckt bei Kriminellen das Interesse für solche Zugangsdaten. Deshalb werden Zugangsdaten regelmäßig auf den verschiedensten Wegen von Kriminellen entwendet und für weitere illegale Aktivitäten missbraucht. In den meisten Fällen bemerken betroffene Personen nicht unmittelbar, dass sie bereits Opfer einer Straftat geworden sind. Je nach krimineller Aktivität erfahren die Betroffenen erst Jahre später davon, dass sich die Täter unter Verwendung ihrer persönlichen Daten bereichert haben. Das Problem wird darüber hinaus noch massiv verstärkt, da viele Benutzer ihre Passwörter mehrfach bei unterschiedlichen Onlinediensten wiederverwenden. Wird ein Benutzer-

[1] Universität Bonn.
[2] Universität Bonn, Fraunhofer FKIE.
[3] Das Forschungsvorhaben EIDI wurde vom Bundesministerium für Bildung und Forschung unter dem Förderkennzeichen 16KIS0696K gefördert.

konto eines solchen Benutzers kompromittiert, dann sind indirekt auch die Benutzerkonten bei den anderen Onlinediensten betroffen.

Um Benutzer vor der Problematik des Identitätsdatendiebstahls zu schützen, fordern verschiedene Institutionen geeignete Maßnahmen von Dienstbetreibern. Der *IT-Grundschutz* des BSI fordert in *ORP.4.A23*: „Es MÜSSEN Maßnahmen ergriffen werden, um die Kompromittierung von Passwörtern zu erkennen“ [1]. Ebenfalls besagt die *Digital Identity Guidelines* des *National Institute of Standards and Technologie*, dass Passwörter beim erstmaligen Setzen oder beim anschließenden Verändern dahingehend überprüft werden müssen, ob sie bereits in einem Identitätsdaten-Leak enthalten sind [2]. Auch wenn diese Überprüfung sinnvoll ist, wird sie auf Seiten der Dienstanbieter selten umgesetzt. Ein Grund dafür liegt darin, dass aktuell keine fertige Sicherheitskomponente existiert, welche die Aufgabe der Überprüfung nach kompromittierten Passwörtern erledigen kann.

Im Rahmen des Forschungsprojekts *Effektive Information nach digitalem Identitätsdiebstahl (EIDI)* [3] entwickelte das interdisziplinäre Konsortium ein System, mit dem sich die genannten Anforderungen, also eine Überprüfung der Zugangsdaten, umsetzen lassen. Die Ergebnisse dieses Forschungsprojektes werden in dieser Arbeit vorgestellt.

Dazu beschreibt Kapitel 2 zunächst die Ursache von Identitätsdatendiebstahl und die Verbreitung gestohlener Datensätze. Kapitel 3 stellt die Vor- und Nachteile ähnlicher, bereits existierender Dienste dar. Kapitel 4 beschreibt den Aufbau des Warndienstes, der die proaktive Warnung Betroffener ermöglicht und so die Kontosicherheit maßgeblich erhöht. Anschließend wird in Kapitel 5 die technische Funktionsweise im Detail erläutert. In Kapitel 6 werden die Ergebnisse präsentiert, die aus der Anwendung des Warndienstes resultieren. Hier wird detailliert auf die umfangreichen Testergebnisse bei den Anwendungspartnern eingegangen. Im Anschluss werden die Potenziale für wirtschaftliche Unternehmen in Kapitel 7 diskutiert. Zum Schluss werden die wichtigsten Ergebnisse zusammengefasst (Kapitel 8).

2. Identitätsdaten-Leaks

Seit dem Doxxing von etwa 1.000 Persönlichkeiten des öffentlichen Lebens, Politikern und Journalisten im Dezember 2018, erfährt der Diebstahl von Identitätsdaten insgesamt eine große öffentliche Aufmerksamkeit. Immer häufiger werden auch Zugangsdaten und Kreditkartennummern nicht prominenter Personen als Identitätsdaten-Leaks verbreitet. Werden solche Identitätsdaten von Datendieben erbeutet und anschließend verkauft oder veröffentlicht, besteht die Gefahr, dass eine Vielzahl individueller Benutzerkonten gefährdet ist. Benutzerdaten von Onlinediensten, die von Kriminellen erbeutet werden, werden als Identitätsdaten-Leak bezeichnet. Leaks enthalten oft nicht nur die Anmeldeinformationen von Benutzerkonten, sondern auch darüber hinaus gehende persönliche Daten, etwa Postadressen, Telefonnummern aber auch Bankkonto- oder Kreditkartennummern. Leaks werden von Kriminellen in Untergrundforen getauscht, verkauft oder sogar öffentlich zugänglich gemacht. Diese Leak-Daten können dann von Betrügern für Identitätsdiebstahl online wie offline genutzt werden. Durch

eine Veröffentlichung von Leaks werden nicht nur professionelle Kriminelle sondern auch kriminelle Laien und sogenannte Scriptkiddies in die Lage versetzt, Identitätsbetrug zu begehen. Öffentlich verfügbare Identitätsdaten bergen aber im Gegenzug auch die Möglichkeit, betroffene Verbraucher frühzeitig zu warnen und so Identitätsbetrug zu verhindern, wenn kritische Informationen, wie Anmeldeinformationen von Betroffenen, rechtzeitig geändert werden können.

3. Existierende Informationsdienste

Bisher gibt es keine erprobte Methode, die Opfer zuverlässig und proaktiv vor dem Missbrauch ihrer gestohlenen Zugangsdaten schützt. Es existieren lediglich reaktive Systeme wie *have i been pwned* [4], *HPI-Leakchecker* [5], *Uni-Bonn Leakchecker* [6]. Diese Leak-Informationsdienste können von interessierten Personen genutzt werden, um zu überprüfen, ob ihre Zugangsdaten in der Vergangenheit gestohlen und veröffentlicht wurden. Jeder dieser Dienste basiert auf einem eigenen Datenbestand gestohlener Identitätsdaten. Auf der Website des jeweiligen Dienstes können Personen eigene E-Mail-Adressen eingeben. Der Dienst überprüft anschließend, ob er die eingegebene E-Mail-Adresse in den gespeicherten Leaks findet. Anschließend bekommt der Benutzer eine Rückmeldung über das Analyseergebnis. Die Rückmeldung erfolgt beim *HPI-Leakchecker* und beim *Uni-Bonn Leakchecker* per E-Mail an die eingegebene Adresse. Lediglich *have i been pwned* stellt die Informationen direkt als Ergebnis auf der Website dar. Das ist aus datenschutzrechtlicher Sicht bedenklich und mit europäischen Gesetzen vermutlich nur schwer vereinbar. Der Grund dafür ist, dass damit unter Umständen jeder für beliebige E-Mail-Adressen herausfinden kann, welche kompromittierten Dienste der Eigentümer einer E-Mail-Adresse verwendet. Das Konzept dieser Informationsdienste ist zwar sinnvoll, jedoch müssen solche Dienste zum einen den Benutzern bekannt sein und zum anderen regelmäßig verwendet werden. Das führt dazu, dass ein solcher Informationsdienst nur eine geringe Reichweite besitzt und deshalb viele Benutzer nicht schützen kann.

4. Ein proaktiver Warndienst

Im Rahmen des in diesem Beitrag vorgestellten Forschungsprojekts wurde ein Frühwarnsystem entwickelt und bereits mit einigen Partnern für deren Benutzerkonten getestet. Dabei wurden viele Probleme thematisiert und hinterfragt, unter anderem der Datenschutz bei der Analyse und Übertragung von Identitätsdaten, die Legitimation bzw. der öffentliche Auftrag zum Betrieb eines solchen Frühwarnsystems, die technische Umsetzung und die Kommunikation mit den Betroffenen.

Das Hauptproblem eines proaktiven Warndienstes liegt in der Kommunikation mit den betroffenen Verbrauchern. Für diese Kommunikation mit den Betroffenen wird ein geeigneter Kommunikationskanal benötigt, um die Warnung zu übermitteln. Ein trivialer Ansatz ist, die in den Leaks enthaltene E-Mail-Adresse zu nutzen, um an die Betroffenen eine E-Mail zu senden. Allerdings ist dieses Vorgehen nicht zielführend, da viele Betroffene eine E-Mail mit einem solchen Warninhalt von einem für sie unbekannten Absender für SPAM halten werden.

Besser wäre es, wenn ein Benutzer von jemandem gewarnt wird, dem er mehr Vertrauen entgegenbringt als einem unbekannten Absender. Dafür in Frage kommen gerade jene Unternehmen, die einem Benutzer bereits Dienstleistungen anbieten und denen der Nutzer zur Erbringung dieser Dienstleistung seine Identitätsdaten bereitgestellt hat. Geeignet sind somit also sämtliche Dienste, die Benutzerkonten verwalten: E-Mail-Anbieter, Soziale Netzwerke, Web-Shops, Banken und viele mehr. Diese Onlinedienste besitzen selbst ein großes Interesse daran, kompromittierte Benutzerkonten zu erkennen und abzusichern, damit über diese kein Missbrauch des angebotenen Onlinedienstes möglich ist und keine Schäden an der eigenen Infrastruktur oder an den Benutzerkonten des Onlinedienstes angerichtet werden können.

Die Kommunikation mit den Betroffenen kann dann zu unterschiedlichen Zeitpunkten erfolgen. Unmittelbar bei Bekanntwerden der Betroffenheit können Betroffene durch eine E-Mail des Onlinedienstes gewarnt werden. Aber auch beim nächsten Besuch oder dem nächsten Login eines Benutzers kann dem Benutzer diese Warnung angezeigt werden. Dabei besteht natürlich die Gefahr, dass auch Kriminelle bei einem Login mit denselben Benutzerdaten die angezeigte Warnung quittieren und zukünftig ausblenden. Um dies zu verhindern, kann der Onlinedienst das betroffene Benutzerkonto in einen eingeschränkten Modus versetzen, der nur noch unkritische Aktivitäten erlaubt. Der betroffene Benutzer hat so die Möglichkeit, relevante Informationen zu erhalten und die Kontowiederherstellung zu starten. Ein Missbrauch des Kontos und der somit erfolgreiche Identitätsbetrug sind aber nicht mehr möglich.

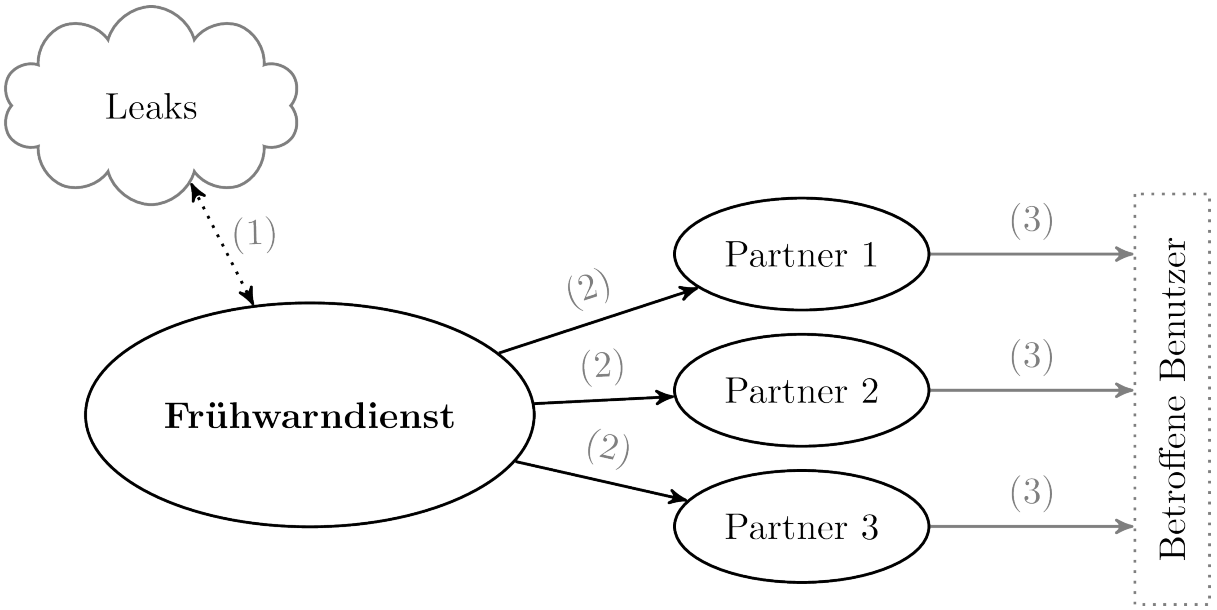


Abbildung 1: Aufbau eines Frühwarndienstes [7], [8].

Das im Rahmen des Forschungsprojektes entwickelte Warnkonzept sieht vor, dass ein zentraler Frühwarndienst betrieben wird, der Identitätsdaten-Leaks sammelt und auswertet (siehe Abbildung 1, Punkt 1). Die Analyseergebnisse übermittelt der Frühwarndienst im nächsten Schritt an seine kooperierenden Onlinedienste und Unternehmen (siehe Abbildung 1, Punkt 2). Abschließend überprüft jeder kooperierende Partner, ob Identitätsdaten seiner eigenen Benutzer in den empfangenen Analyseergebnissen ent-

halten sind. Wenn dies zutrifft, wird der betroffene Benutzer über einen vom Onlinedienst ausgewählten Kommunikationskanal benachrichtigt (siehe Abbildung 1, Punkt 3). Der genaue technische Aufbau wird im nachfolgenden Kapitel erläutert.

5. Technischer Aufbau

In diesem Kapitel wird der technische Aufbau des proaktiven Warnsystems genauer dargestellt. Der technische Aufbau wird erläutert, damit der interessierte Leser ein genaueres Verständnis über den tatsächlichen Nutzen dieses Systems entwickeln kann. Auch müssen technische Details erläutert werden, damit das hohe Maß des erreichten Datenschutzes nachvollzogen werden kann.

Vom Identitätsdaten-Leak bis zur Warnung durch die Partner gibt es drei Teilprozesse zu berücksichtigen: 1. Identitätsdaten-Leaks sammeln, 2. Daten analysieren, 3. Datenübermittlung an kooperierende Unternehmen. Zur besseren Übersicht werden die drei Teilprozesse in eigenen Abschnitten beschrieben. Aus Platzgründen kann dabei nicht jeder Teilprozess in vollem Umfang dargestellt werden. Die angegebenen Referenzen enthalten jeweils detailliertere Beschreibungen der einzelnen Teilprozesse.

5.1. Leak-Daten sammeln

Das vorgestellte Warnsystem für kompromittierte Identitätsdaten basiert darauf, dass kontinuierlich nach Identitätsdaten-Leaks gesucht wird und diese Daten zur Überprüfung der Betroffenheit von Benutzern genutzt werden können. Aus diesem Grund wurde ein Verfahren verwendet, das möglichst effizient neue Identitätsdaten-Leaks aufspüren kann. Aus rechtlichen Gründen kann hierbei im Forschungskontext nur auf öffentlich verfügbare Identitätsdaten-Leaks eingegangen werden. In Abbildung 2 ist das Vorgehen zum Sammeln von Identitätsdaten-Leaks zu sehen. Während des Forschungsprojektes konnten so bereits mehr als 25 Milliarden gestohlene Identitätsdaten gesammelt werden. Die analysierten Datensätze verteilen sich dabei auf über 80.000 Dateien. Einige dieser Dateien sind in eigenen Sammlungen, sogenannten Collections, zusammengefasst.

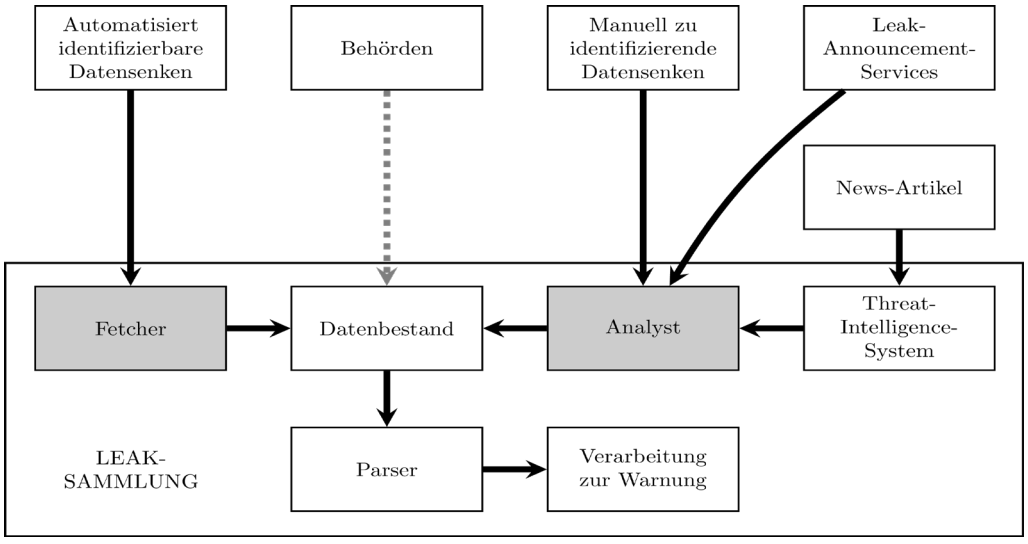


Abbildung 2: Prozess zum Sammeln von Identitätsdaten-Leaks [9].

Der Prozess zum Sammeln besteht aus zwei Hauptkomponenten. Der *Fetcher* durchsucht vollautomatisiert geeignete Quellen und speichert Dateien mit relevanten Inhalten im *Datenbestand* ab. Quellen, die nicht automatisiert verarbeitet werden können, werden durch einen Mitarbeiter, den *Analysten*, durchsucht und Dateien manuell heruntergeladen. Damit der Analyst weiß, wonach er suchen muss, nutzt er vorhandene Leak-Announcement-Services [10] oder ein eigens entwickeltes Threat-Intelligence-System [11], das auf aktuelle Identitätsdaten-Leaks hinweisen kann. Darüber hinaus ist es denkbar, dass Personen oder Einrichtungen weitere Identitätsdatensammlungen zur Verfügung stellen. So können etwa Polizeibehörden, nach einer entsprechenden Untersuchung, die Warnung betroffener Verbraucher über den Warndienst sicherstellen.

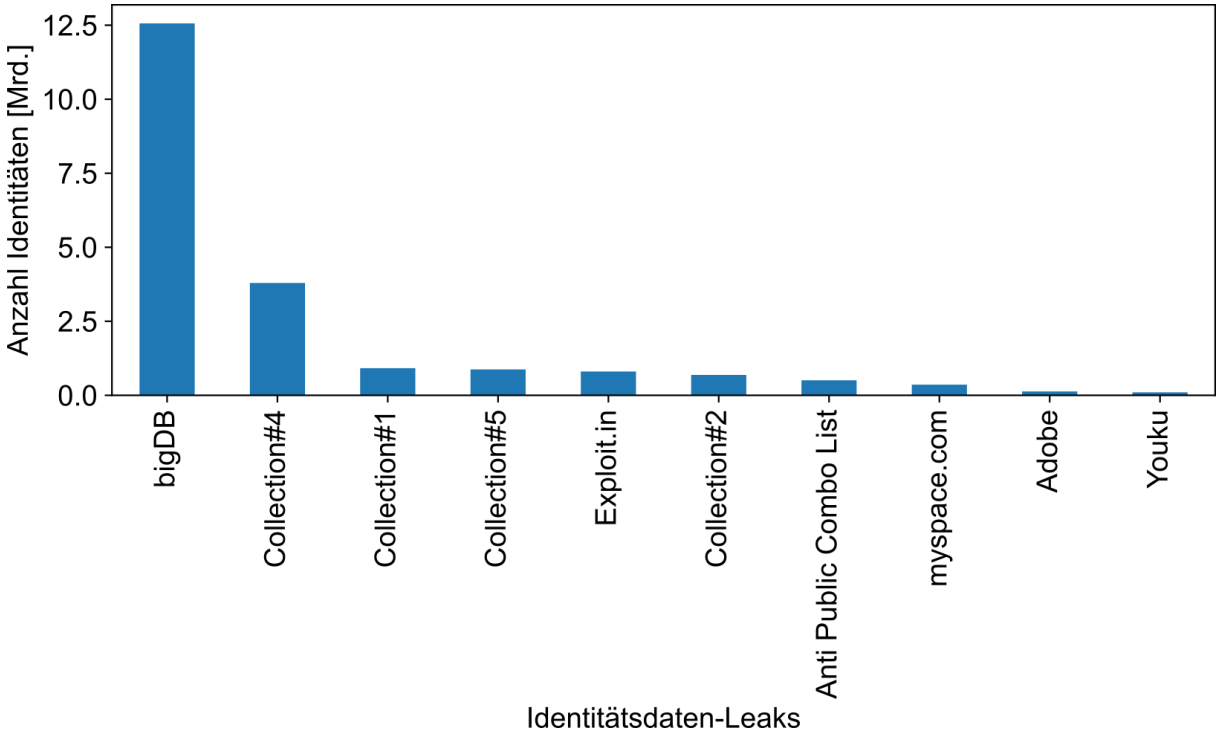


Abbildung 3: Anzahl der Identitäten in den größten Identitätsdaten-Leaks.

Die zehn größten Identitätsdaten-Leaks und die Anzahl enthaltener Datensätze sind in Abbildung 3 dargestellt. Zusammen enthalten die zehn größten Identitätsdaten-Leaks bereits mehr als 21 Milliarden Identitätsdaten.

5.2. Daten verarbeiten

Die automatische Analyse von Identitätsdaten-Leaks ist nicht trivial. Ein maßgeblicher Grund ist, dass Identitätsdaten-Leaks keine einheitliche Syntax für die enthaltenen Datensätze verwenden. Das führt dazu, dass verschiedene Identitätsdaten-Leaks auch unterschiedliche Datenformatierungen aufweisen, dies erschwert die automatisierte Verarbeitung. Die am häufigsten verwendete Formatierung von Identitätsdaten-Leaks sind Textdateien, bei der jede Zeile einen eigenen Identitätsdatensatz eines Benutzers enthält. Jede Zeile besteht aus mehreren Datenbestandteilen, wie E-Mail-Adresse, Passwort, Name und Geburtsdaten. Diese Bestandteile werden mit unterschiedlichen Trennzeichen voneinander getrennt. Zur automatisierten Verarbeitung dieser Dateien

muss für jede einzelne Zeile sowohl die Syntax als auch die Semantik der einzelnen Datenfelder zuverlässig erkannt werden. In Abbildung 4 ist der im Forschungsprojekt entwickelte Prozess zur automatisierten Verarbeitung von Identitätsdaten-Leaks dargestellt. Identitätsdaten-Leaks werden mithilfe des *Input-Moduls* eingelesen und Archivdateien, falls notwendig, vorher entpackt. Das *Block-Detektions-Modul* trennt eine Datei in einzelne Blöcke auf, um Veränderungen der Syntax oder Semantik innerhalb einer einzelnen Leak-Datei zu erkennen. Das *Seperator-Modul* erkennt die in den einzelnen Blöcken verwendeten Trennzeichen, mit denen die einzelnen Attribute eines Datensatzes innerhalb einer Zeile voneinander getrennt sind. Danach wird die Semantik der einzelnen Attribute mittels *Semantisierungs-Modul* erkannt. Abschließend speichert das *Output-Modul* die verarbeiteten Daten in einer Datenbank ab.

Vor der Speicherung in der Datenbank werden zwingende Maßnahmen zur Realisierung des Datenschutzes durchgeführt. Nicht zur sinngemäßen Überprüfung benötigte Attribute werden gelöscht und alle übrigen Attribute eines Datensatzes nach dem aktuellen Stand der Technik pseudonymisiert. Eine genauere Darstellung des Verfahrens ist in [10], [12] erfolgt.

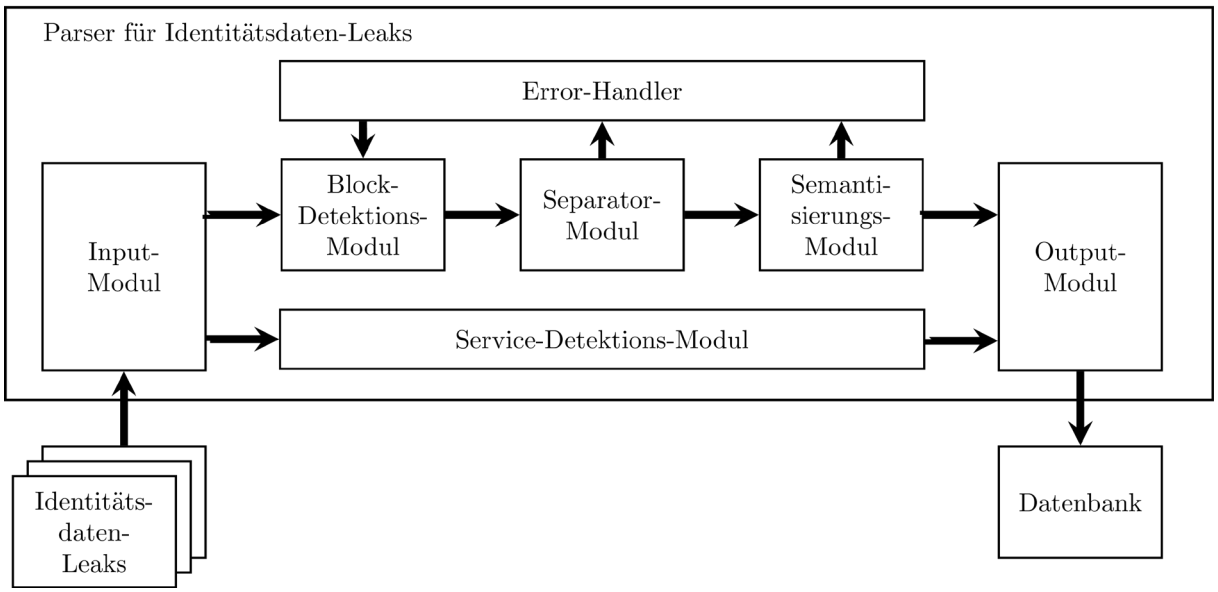


Abbildung 4: Prozess zum Verarbeiten von Identitätsdaten-Leaks [13].

5.3. Datenübermittlung an kooperierende Unternehmen

Die pseudonymisierten Daten müssen im letzten Schritt zur Überprüfung an die kooperierenden Partner übermittelt werden. Dazu betreibt jeder Kooperationspartner eine eigene Schnittstelle, an die der Frühwarndienst die Datensätze übermitteln kann. Sobald ein Kooperationspartner eine gewisse Anzahl an Datensätzen empfangen hat, kann er diese mit den Daten der eigenen Benutzerdatenbank vergleichen. Aufgrund der zur Pseudonymisierung verwendeten kryptografischen Verfahren bekommt der Kooperationspartner dabei nur Einsicht in die Datensätze der eigenen Benutzer. Mit den in einem Datensatz enthaltenen Klartextpasswörtern kann anschließend überprüft werden, ob ein Login am Dienst mit den vorliegenden Datensätzen möglich ist. Ist dies der

Fall, muss der Kooperationspartner geeignete Maßnahmen zum Schutz des Benutzerkontos und damit auch zum Schutz der eigenen Infrastruktur umsetzen. Vorgesehen ist hier, dass der betroffene Benutzer über einen geeigneten Kommunikationskanal informiert und das Konto bis zur vollständigen Absicherung in seiner Funktion eingeschränkt wird. Damit werden sowohl die eigene technische Infrastruktur, als auch der angebotene Dienst und der betroffene Benutzer gleichermaßen geschützt. Das genaue Verfahren zur Datenübertragung ist in [7] erklärt.

6. Ergebnisse im Praxiseinsatz

Bereits während des Forschungsprojektes wurden die entwickelten Systeme mit den Anwendungspartnern getestet. Dazu wurden die mit dem Vorgehen aus Abschnitt 5.1 gesammelten Identitätsdaten-Leaks mit dem in Kapitel 4 vorgestellten Parser analysiert. Abgespeichert und übermittelt wurden dabei nur pseudonymisierte Datensätze.

Im Testzeitraum von sechs Monaten wurden ca. 20 % aller gesammelten Datensätze an die Anwendungspartner versendet. Die Anwendungspartner haben anschließend überprüft, ob gültige Zugangsdaten der eigenen Benutzer in den übermittelten Leak-Datensätzen enthalten sind. War dies der Fall, so wurden die betroffenen Benutzer gewarnt und weitere technische Sicherheitsmaßnahmen seitens der Anwendungspartner umgesetzt. Die übermittelten Daten enthielten dabei Leak-Daten, die nicht bei den Anwendungspartnern selbst abhandengekommen sind. In der Vergangenheit ist weder über Leaks bei den Partnern berichtet worden, noch waren Hinweise in den übermittelten Daten zu erkennen, die auf einen Leak bei den Partnern hinweisen. Das heißt unmittelbar, dass die Benutzer der Anwendungspartner die gleichen Zugangsdaten bei unterschiedlichen Diensten verwenden.

Nur wenn die Zugangsdaten für einen Anwendungspartner gültig waren, also ein Login bei dem Dienst damit möglich war, wurden weitere Maßnahmen zum Schutz der betroffenen Benutzerkonten ergriffen. Bereits während der Projektlaufzeit konnten mit dem vorgestellten Verfahren mehr als eine Million akut gefährdete Benutzerkonten geschützt werden. Das bedeutet, dass in den übermittelten Datensätzen mehr als eine Million gültige Zugangsdaten für die Dienste der Projektpartner enthalten waren. Das bedeutet zugleich, dass mehr als eine Million Benutzer vor Schäden durch Identitätsdiebstahl geschützt werden konnten. Das Projekt hat dazu geführt, dass ein enormer Sicherheitsgewinn bei den Anwendungspartnern erzielt wurde. Die nachvollziehbaren erfolgreichen Credential-Stuffing-Angriffe gegen diese Onlinedienste haben sich seitdem merkbar reduziert.

7. Potenziale für die Wirtschaft

Identitätsdiebstahl verursacht Schäden in unterschiedlichen Bereichen. Da Kriminelle zumeist das Ziel der eigenen Bereicherung verfolgen, stehen finanzielle Schäden durch Identitätsdiebstahl im Vordergrund für betroffene Verbraucher und Dienstanbieter. Aber auch psychische oder soziale Beeinträchtigungen der Opfer müssen bei der Betrachtung berücksichtigt werden. Ebenso erleiden Dienstanbieter einen Schaden bei

der öffentlichen Wahrnehmung des Dienstes, wenn der Identitätsdiebstahl von Kunden medial thematisiert wird.

Es lassen sich also sowohl die Verbraucher als auch die Anbieter von Onlinediensten als Geschädigte durch Identitätsdiebstahl ausmachen. Bereits vor dem Forschungsprojekt bestehende Lösungen, wie die Leakchecker, werden heute auch in Passwortmanagern genutzt, um die gespeicherten Zugangsdaten der Verbraucher zu überwachen. Damit gibt es für die Seite der Verbraucher bereits erprobte Werkzeuge, um die eigene Betroffenheit drastisch zu reduzieren und zuverlässige Gegenmaßnahmen zu treffen. Damit diese Produkte großflächig, möglichst von allen Benutzern verwendet werden, müssen sie in die Standardsoftware der Betriebssysteme, vor allem in die dort genutzten Internetbrowser, integriert werden.

Betrachtet man die Haftungsfrage im Nachgang zu vollendetem Identitätsbetrug, etwa bei Bestellungen im Namen der Betroffenen bei einem Online-Shop, so tragen zumeist nicht die Verbraucher, sondern die Anbieter der Online-Shops das Risiko eines Zahlungsausfalls. Größere Online-Shops kalkulieren dieses Risiko bereits umfänglich in die Preise der angebotenen Produkte mit ein und verfolgen solche Fälle bei kleineren Summen, etwa bis 500 Euro, gar nicht weiter. Bei größeren Schadenssummen wird entweder eine eigene Abteilung für Missbrauchserkennung und Betrugsmanagement aktiv oder ein externer Dienstleister für die weitere Abwicklung beauftragt. Man kann davon ausgehen, dass in diesen Fällen am Ende auch der Anbieter den Schaden hat, zumindest solange der Täter nicht gefasst werden kann.

Ein weiteres Haftungsrisiko ist die öffentliche Wahrnehmung eines Dienstanbieters, wenn dort vermehrt Identitätsdatendiebstahl offenkundig wird. Benutzerkonten von Online-Diensten mit der Funktion öffentliche oder private Nachrichten zwischen den Benutzern auszutauschen, werden von Kriminellen regelmäßig zu Spam-Zwecken genutzt. Erhalten Benutzer dieser Dienste vermehrt Spam, wirft das ein schlechtes Licht auf den Betreiber und seine Fähigkeiten, die eigenen Benutzer vor Spam zu schützen. Auch hier ist der Betreiber also unmittelbar der Geschädigte durch Identitätsdatendiebstahl seiner Benutzerkonten. Reputations- oder Imageschäden lassen sich häufig nicht in einer finanziellen Dimension ausdrücken. Anbieter sind aber bestrebt, solche Schäden vom eigenen Unternehmen fernzuhalten.

Wenn Dienstanbieter einen solchen Warndienst einsetzen, schützen sie damit zunächst sich selbst. Durch die Warnung der Benutzer schützen sie aber mittelbar auch alle von Identitätsdaten-Leaks betroffenen Verbraucher und möglicherweise sogar Anbieter, die den Warndienst selbst gar nicht einsetzen. Der hier vorgestellte Warndienst ist in der Art einzigartig, dass er datenschutzkonform auch von europäischen Dienstanbietern genutzt werden kann. Hier ergibt sich ein relevantes Nutzungspotenzial für die gesamte Online-Wirtschaft in Europa. Sogenannte Cyber-Versicherungen ermöglichen die finanzielle Absicherung gegen Schäden unterschiedlicher Art, also auch gegen Schäden durch Identitätsdiebstahl oder Imageschäden in Folge von Social-Media-Aktivitäten. Durch den Einsatz des Warndienstes kann jedoch ein großer Teil dieser Schäden bereits im Vorfeld verhindert werden. Versicherungen selbst können, um

günstigere Versicherungspolicen anzubieten, den Einsatz eines solchen Warndienstes in Anspruch nehmen.

Die Betrachtung möglicher Schäden durch Identitätsbetrug im Internet fokussiert sehr ein Anbieter-Kunden-Verhältnis und die damit verbundene Haftungsfrage. Auch Unternehmen, die kein Dienstanbieter in diesem Sinne sind, erleiden immer wieder Schäden durch kompromittierte Zugangsdaten ihrer Mitarbeiter. Diese Zugangsdaten öffnen häufig für Kriminelle den Zugang zu internen Netzwerken und damit zu Betriebs- oder Geschäftsgeheimnissen. Mit diesem Vorgehen kann eine umfassende Industriespionage durchgeführt werden. Betrachtet man mögliche Schäden durch illegale Zugriffe auf solche Informationen, ist man schnell in mindestens gleichwertigen Bereichen, die auch Erpressungs- oder Crypto-Trojaner erreichen. Es gilt auch, die Zugangsdaten der eigenen Mitarbeiter im Blick zu haben und regelmäßig deren Kompromittierung und die Veröffentlichung in Identitätsdaten-Leaks zu überprüfen.

8. Zusammenfassung

Passwörter sind für den Schutz von Nutzerzugängen zu Onlinediensten nach wie vor sehr verbreitet. Über verschiedene Wege gelingt es Angreifern immer wieder, Passwörter zu kompromittieren und damit Identitätsdaten zu stehlen, um den Onlinedienst missbräuchlich nutzen zu können. Die schadenbegrenzende Sicherheitsvorgabe, für jeden Onlinedienst ein eigenes Passwort zu verwenden, stellt viele Nutzer vor dem Hintergrund der großen Anzahl heute genutzter Onlinedienste vor eine Herausforderung, an der sie scheitern. Im Ergebnis wird die Sicherheitsvorgabe regelmäßig verletzt und einzelne Passwörter werden für mehrere Onlinedienste verwendet, und das Missbrauchs- und Schadenspotenzial durch eine Passwortkompromittierung vervielfacht sich. Da Identitätsdatendiebstahl zunächst unbemerkt bleibt, fehlt die Möglichkeit geeignete Reaktionen einzuleiten. Hier setzt der in diesem Beitrag vorgestellte Warndienst an. Systematisch werden kontinuierlich Identitätsdaten-Leaks gesammelt und analysiert, um von Identitätsdatendiebstahl Betroffene proaktiv ohne deren Aktivwerden zu warnen. Durch die Einbindung von kooperierenden Onlinediensten in den Warnprozess kann mit großer Reichweite gewarnt werden. Außerdem besteht dadurch und im Unterschied zu bisherigen Lösungen die Möglichkeit, zunächst eine Validierung der konkreten Bedrohung vorzunehmen. Damit kann eine Warnung auf Fälle begrenzt werden, in denen eine akute Bedrohung besteht, also ein aktuelles (und nicht irgendein altes) Passwort kompromittiert wurde. Durch Umsetzung kryptografischer Pseudonymisierungstechniken wird dabei hohen Vertraulichkeitsschutzanforderungen und den umfangreichen europäischen Datenschutzanforderungen entsprochen. Tests mit dem entwickelten Frühwarnsystem zeigten, dass bereits 25 Milliarden gestohlene Identitätsdatensätze gesammelt und analysiert werden konnten. Bei den am Test beteiligten Anwendungspartnern konnten bereits über eine Million akut betroffene Nutzer gewarnt und geschützt werden. Für die im Frühwarnsystem umgesetzten Technologien ergeben sich weitere sinnvolle Einsatzszenarien und Wirkpotenziale im Unternehmenskontext. Einerseits können berufliche Onlinedienst-Zugänge von Mitarbeitern einmalig, regelmäßig oder kontinuierlich auf mögliche Kompromittierung durch gestohlene ggf. mehrfach verwendete Identitätsdaten überprüft werden, um so

den Schutz der Unternehmensinfrastruktur zu verbessern und entsprechenden Vorgaben z.B. des BSI-Grundschutzes gerecht zu werden. Andererseits können Identitätsdaten von Kunden durch den Onlinedienst-Betreiber auf Kompromittierung überprüft werden und aktuelle Kundentransaktionen, z.B. Bestellungen in Online-Shops, hinsichtlich möglicher Betrugsrisiken bewertet und verarbeitet werden, z.B. durch risikoabhängige Einschränkung der Zahlungs- oder Versandoptionen. Entsprechende Überprüfungs-Dienstleistungen bietet das vom Projektteam gegründete Unternehmen Identeco[4] an.

[4] Identeco GmbH & Co. KG: https://identeco.de.

Literaturverzeichnis

[1] Bundesamt für Sicherheit in der Informationstechnik, „IT-Grundschutz - ORP.4 Identitäts- und Berechtigungsmanagement“, 2021. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Kompendium_Einzel_PDFs_2021/02_ORP_Organisation_und_Personal/ORP_4_Identitaets_und_Berechtigungsmanagement_Editon_2021.html (neue URL ab 1. Februar 2021).

[2] National Institute of Standards and Technologie, „Digital Identity Guidelines - Authentication and Lifecycle Management“, 2021. https://pages.nist.gov/800-63-3/sp800-63b.html (zugegriffen Jan. 05, 2021).

[3] Universität Bonn - Institut für Informatik IV, „Effektive Information nach digitalem Identitätsdiebstahl“, 2020. https://itsec.cs.uni-bonn.de/eidi/ (zugegriffen Jan. 05, 2020).

[4] T. Hunt, „have i been pwned?“, 2017. https://haveibeenpwned.com (zugegriffen Jan. 05, 2021).

[5] Hasso Plattner Institut, „HPI-Leakchecker“, 2020. https://sec.hpi.de/ilc/ (zugegriffen Jan. 05, 2021).

[6] Universität Bonn, „Identity Leakchecker - Uni Bonn“, 2020. https://leakchecker.uni-bonn.de (zugegriffen Jan. 05, 2021).

[7] T. Malderle, M. Wübbeling, S. Knauer, und M. Meier, „Warning of Affected Users About an Identity Leak“, in *Proceedings of the Tenth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2018)*, 2019, S. 278–287.

[8] T. Malderle, M. Wübbeling, und M. Meier, „Effektive Warnung bei Identitätsdiebstahl an Hochschulen“, in *Sicherheit in vernetzten Systemen - 27. DFN-Konferenz*, 2020.

[9] T. Malderle, M. Wübbeling, und M. Meier, „Sammlung geleakter Identitätsdaten zur Vorbereitung proaktiver Opfer-Warnung“, in *Multikonferenz Wirtschaftsinformatik 2018 : Data driven X - Turning Data into Value*, 2018, S. 1381–1393.

[10] T. Malderle, M. Wübbeling, S. Knauer, A. Sykosch, und M. Meier, „Gathering and Analyzing Identity Leaks for a Proactive Warning of Affected Users“, in *Proceedings of the 15th ACM International Conference on Computing Frontiers*, 2018, S. 208–211, doi: 10.1145/3203217.3203269.

[11] T. Malderle, S. Knauer, M. Lang, M. Wübbeling, und M. Meier, „Track Down Identity Leaks Using Threat Intelligence“, in *ICISSP 2020 - Proceedings of the 6th International Conference on Information Systems Security and Privacy*, 2020.

[12] T. Malderle, M. Wübbeling, S. Knauer, und M. Meier, „Ein Werkzeug zur automatisierten Analyse von Identitätsdaten-Leaks“, in *SICHERHEIT 2018*, 2018, S. 43–54.

[13] T. Malderle, „Bedrohung durch Identitätsdatendiebstahl - Datenerhebung, Analyse und Mitigation“, Bonn, 2020.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Angriffserkennungssysteme in ICS Netzwerken

Jörg Kippe[1], Markus Karch[1]

Kurzfassung:
Das vorliegende Papier fasst eine vom Bundesamt für Sicherheit in der Informationstechnik (BSI) beauftragte Evaluierung von kommerziell verfügbaren Systemen zum Monitoring und zur Erkennung von Cyber-Angriffen in den Netzen industrieller Automatisierungssysteme (Industrial Control Systems – ICS) zusammen. Die Untersuchung wurde durch die Referate DI23 und TK15 des BSI im Rahmen des Projektes „Sicherheits- und Funktionsanalysen / Proof of Concepts für Industrie 4.0“ beauftragt und vom Fraunhofer IOSB in Karlsruhe durchgeführt. In einer realen Unternehmensnetzen und Automatisierungssystemen nachempfundenen Test- und Demonstrationsumgebung wurde eine Auswahl von vier am Markt verfügbaren Systeme installiert und in Betrieb genommen. Anschließend wurde das Verhalten der Systeme in zwei voneinander getrennten Phasen untersucht: In der ersten Phase wurden „normale“ Änderungen in der Testumgebung, die während des normalen Netzbetriebes vorkommen können, beispielhaft nachempfunden. In der zweiten Phase wurden Cyber-Angriffe gegen industrielle Komponenten durchgeführt. In beiden Phasen wurde das Verhalten der Systeme evaluiert. Zusätzlich wurden im Rahmen der Evaluation sowohl Funktionalitäten und typische Eigenschaften der Systeme identifiziert und verglichen als auch eine Einschätzung bezüglich Ihrer Einsatzmöglichkeit unter Berücksichtigung potenziell auftretender Probleme gegeben.

Stichworte: Angriffserkennung, Anomalieerkennung, ICS-Netzwerke, Intrusion Detection, Industrielle Automatisierungssysteme

1. Einleitung

Das vorliegende Papier präsentiert eine Reihe von Schlussfolgerungen hinsichtlich des Einsatzes von Systemen zur Detektion von Cyber-Angriffen in den Netzwerken industrieller Automatisierungssysteme (Industrial Control Systems – ICS). Diese Schlussfolgerungen sind das Ergebnis einer Studie über das Verhalten von vier kommerziell verfügbaren Angriffserkennungssystemen (Intrusion Detection Systems – IDS), die als beauftragtes Projekt von uns durchgeführt wurde. Die Evaluierung der Angriffserkennungssysteme wurde in einer gemeinsam mit dem BSI als Auftraggeber konzipierten und von uns realisierten Test- und Demonstrationsumgebung durchgeführt. Aus naheliegenden Gründen können die untersuchten Produkte an dieser Stelle nicht genannt und konkrete Produkteigenschaften nicht dargestellt werden; vielmehr sollen

- typische Produkteigenschaften herausgearbeitet,
- typische Umstände der Installation und des Betriebs solcher Systeme beschrieben,
- das konkrete Evaluationsvorgehen vorgestellt,
- und Schlussfolgerungen aufgrund der Evaluationsergebnisse erläutert werden.

[1] Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung, Abteilung Informationsmanagement und Leittechnik (ILT), 76131 Karlsruhe

2. Kategorisierung von Angriffserkennungssystemen

Sicherheitsmaßnahmen lassen sich in drei Kernfunktionalitäten einteilen:

- **Prevention** umfasst die Schutzmaßnahmen, die das Eindringen eines Angreifers oder den Erfolg von Angriffen verhindern sollen.
- **Detection** bedeutet das Erkennen des versuchten oder erfolgreichen Eindringens oder der Durchführung eines Angriffs. Unter der Grundannahme des *assume breach* (d.h. die Prevention Maßnahmen sind niemals perfekt und ein Eindringen des Angreifers lässt sich nicht vermeiden), ist dieses die wichtigste Cyber Security Funktion. Ohne Detection bleibt der Angreifer oftmals unsichtbar.
- **Response** bedeutet die Reaktion auf einen Angriff, die wiederum leichter wird, je besser die Detection ausfällt und je genauer der Angriff und seine Umstände erkannt werden.

Bezüglich der Angriffserkennungsverfahren lassen sich vier Grundtypen [1] unterscheiden:

- **Konfigurationsbasierte Erkennung** beruht auf der Erkennung der Abweichungen von einer Basis (Baseline). Für die Definition einer solchen Basis gibt es viele Möglichkeiten: Hosts im Netzwerk, verwendete Dienste, verwendete Port-Nummern, typische Parameterbelegungen in Netzwerkprotokollen, aktive User, etc. Alles Neue und Andere ist eine Abweichung von dieser Basis. Konfigurationsbasierte Erkennung beruht auf dem Wissen über die Architektur und das Design einer Umgebung. Leider ist die Konfiguration einer solchen Umgebung aber oftmals Änderungen unterworfen. Konfigurationsbasierte Erkennung kann in einer hoch-statischen ICS Umgebung effektiv sein, jedoch sind die meisten industriellen Umgebungen nicht wirklich statisch [1]. Selbst in moderat dynamischen Umgebungen kann die Pflege einer korrekten konfigurationsbasierten Erkennung eine mühevolle Aufgabe sein. Die Erkennung jeglicher Abweichung führt im Fall von legitimen Änderungen zu einer hohen Anzahl von Meldungen, denen kein Cyber-Angriff zugrunde liegt. Mangelndes Wissen und mangelnde Pflege der Konfigurationsdaten kann diese Art der Erkennung daher für den praktischen Einsatz unbrauchbar machen.
- **Modellbasierte Erkennung** basiert auf der Annahme, dass der Erkennungsmechanismus zwischen legitimen und illegitimen Aktivitäten unterscheiden kann. Modellbasierte und konfigurationsbasierte Erkennung sind sehr ähnlich, der Unterschied liegt darin, dass die Konfiguration von Experten erstellt wird, während die Modellbildungsverfahren auf dem Erlernen einer Baseline beruhen und versuchen, ohne Expertenwissen auszukommen. Das Erlernen eines Modells erfordert einen signifikanten Aufwand, um alle Kommunikationsaspekte abzudecken, und nach Änderungen der Umgebung muss das Modell in der Regel neu generiert werden. Es besteht die Möglichkeit, dass bereits vorhandene bösartige Aktivitäten mit in das Modell eingehen. Dies kann ein reales Problem sein, wenn Bedrohungen in einer Umgebung hunderte von Tagen unbemerkt aktiv sein können. Weil der Erkennungsmechanismus auf einem erlernten Modell beruht, mangelt es diesem Ansatz an Transparenz und Kontextinformationen hinsichtlich der

Gründe eines Alarms, der nicht direkt auf einen Angriff zurückgeführt werden kann.

- **Indikatorbasierte Erkennung** beruht auf einfachen Informationselementen (indicators of compromise), die bestimmte Zustände und Kontexte bezüglich schädlicher Aktivitäten identifizieren. Diese Informationselemente werden in der Regel aus der Analyse von Angriffen bestimmt. Sie sind ein schneller und weit verbreiteter Weg zur Angriffserkennung. Von Vorteil ist die direkte und sofort verfügbare Beziehung zum Angriff. Der Nutzen ist allerdings dadurch begrenzt, dass viele Indikatoren durch den Angreifer leicht modifiziert werden können und dadurch wertlos werden.
- **Verhaltensbasierte Erkennung** beruht auf der Identifikation der von einem Angreifer durchgeführten Verhaltensweisen oder Methoden. Dabei wird von technischen Attributen, wie beispielsweise Indikatoren, abstrahiert. Die Erkennung arbeitet meistens mit Folgen von Ereignissen und einer komplexen Menge von Regeln. Diese sogenannten Analytics können nicht durch Modellierungsverfahren generiert werden, sondern müssen durch Analysten durch die Analyse realer Angriffe entwickelt werden.

Eine weitere wichtige Kategorisierung der Angriffserkennung erfolgt hinsichtlich der herangezogenen Datenquellen.

- **Netzwerkbasierte Daten** (Paketdaten) werden durch geeignete Instrumentierung (Mirror Ports an Switches, Netzwerk Taps) als Packetstrom gewonnen, auf den verschiedene Erkennungsmechanismen angewendet werden können. An erster Stelle ist hier die Packetanalyse mittels vordefinierter Erkennungsmuster (Signaturen) zu nennen (technologische Beispiele hierzu sind Snort[2] und Suricata[3]). Auf einer höheren Abstraktionsebene arbeiten Systeme, die Kommunikationsverbindungen und Protokollelemente extrahieren können (als Beispiel: Zeek[4]). Die Analyse alle Protokollelemente einschließlich der Payload bezeichnet man als *Deep Packet Inspection*.
- **Hostbasierte Daten** (Ereignisdaten) entsprechen Meldungen von Ereignissen, die innerhalb des Betriebssystems sowie in Prozessen und Applikationsprogrammen erkannt werden. Neben dem reinen Weiterleiten von Logmeldungen sind Agenten üblich, die neben dieser Weiterleitung auch Sicherheitsüberprüfungen vornehmen und darüber hinaus Meldungen erzeugen.

3. Untersuchte Systeme

Bei den untersuchten Systemen handelt es sich um kommerziell verfügbare Produkte zur Angriffserkennung in den Netzwerken industrieller Automatisierungssysteme. Im Sinne der zuvor getroffenen Kategorisierung lassen sich diese Systeme primär beschreiben als modellbasierte Erkennungssysteme (d.h. Anomalieerkennungssysteme) unter

[2] https://snort.org
[3] https://suricata-ids.org
[4] https://zeek.org

Verwendung von netzwerkbasierten Daten. Daneben beinhalten manche Systeme aber auch Elemente indikatorbasierter und modellbasierter Erkennung. Neben der Erkennungsfunktionalität bieten die Systeme in der Regel als weitere Funktionalitäten Asset Inventory, Vulnerability Management und Risk Management, diese sollen hier aber nicht weiter betrachtet werden.

Gemeinsame Eigenschaften der untersuchten Systeme sind:

- Passives Monitoring des Netzwerkverkehrs und Analyse der Paketdaten auf den Protokollschichten 2, 3, 4 und 7
- Verwaltung, statistische Aufbereitung und Visualisierung von Kommunikationsbeziehungen
- Zwei unterschiedliche Betriebsmodi:
  - In der *Lernphase* werden die beobachteten Paketdaten verwendet, um ein Modell des zu überwachenden Netzwerks aufzubauen
  - In der *Detektionsphase* werden die beobachteten Paketdaten mit dem erlernten Modell verglichen und Abweichungen als Anomalien gemeldet

4. Testumgebung und Evaluationsvorgehen

Die untersuchten Systeme wurden in eine Testumgebung integriert, welche gemeinsam mit dem Auftraggeber skizziert und durch das Fraunhofer IOSB realisiert wurde. Bei

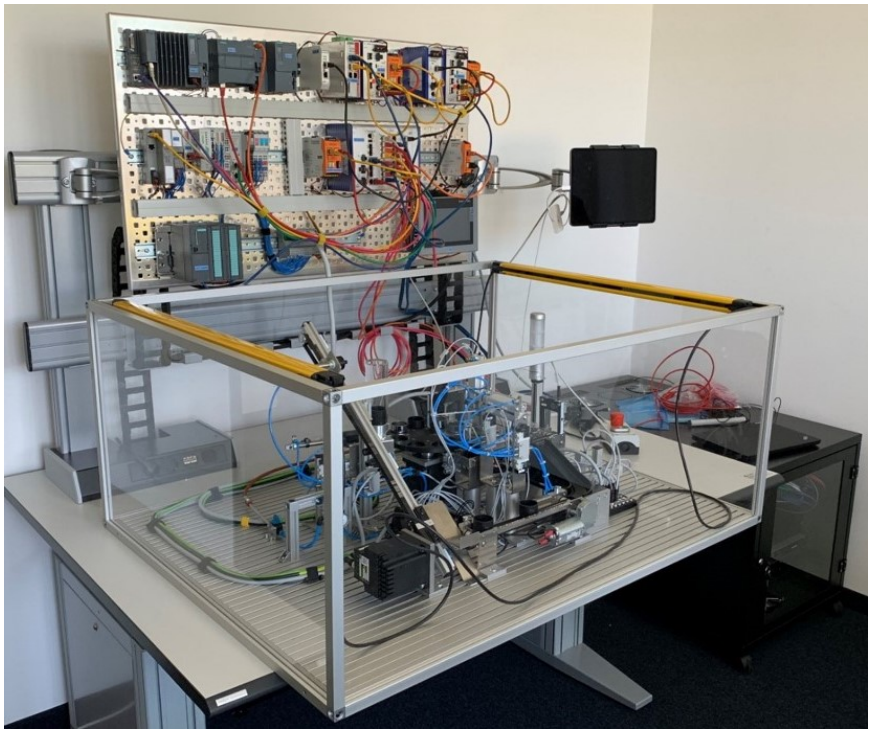


Abbildung 1: Testumgebung zur Evaluation der untersuchten Systeme, bestehend aus einem auf der Prozessplatte angebrachten und im Plexiglaskasten befindlichen Fertigungsprozess, gängigen auf der Prozesstafel angebrachten Automatisierungskomponenten zur Steuerung des Fertigungsprozesses und einem schwarzen Serverschrank, welcher die Virtualisierungsumgebung sowie Office-typische Geräte beinhaltet.

der Testumgebung, welche die Beobachtung des gesamten Netzwerkverkehrs erlaubt, handelt es sich um einen vereinfachten und miniaturisierten Fertigungsablauf. In Kombination mit einem auf die Testumgebung anwendbaren Portfolio von Cyber-Angriffen wurden die untersuchten Systeme evaluiert.

4.1. Testumgebung und Integration der untersuchten Systeme

Die Testumgebung ist in Abbildung 1 dargestellt und besteht aus einem modellhaften Fertigungsprozess, gängigen industriellen Automatisierungskomponenten und einem Serverschrank. Der Fertigungsprozess wird mittels mechanischer Komponenten auf der Prozessplatte durchgeführt. Die zur Steuerung des Prozesses dienenden industriellen Automatisierungssysteme sind auf einer Prozesstafel angebracht. Der Serverschrank beinhaltet eine Virtualisierungsumgebung, die eine Netzhierarchie entsprechend dem Purdue Modell [3] realisiert und eine Reihe virtueller Maschinen enthält, die Automatisierungs- und Officefunktionen erfüllen.

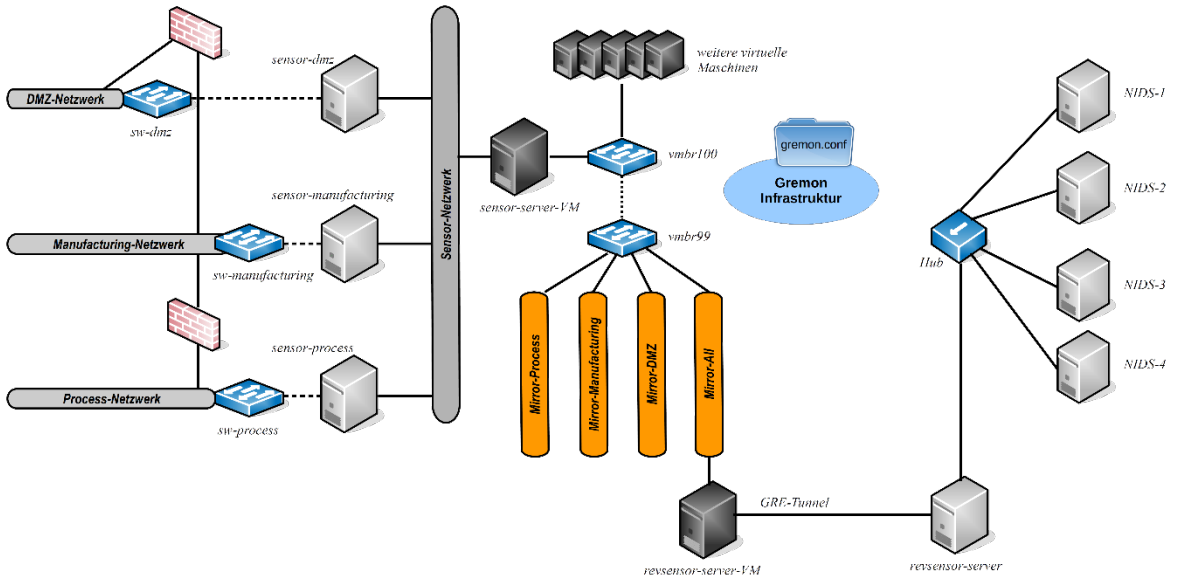


Abbildung 2: Architektur zur Bereitstellung des gesamten Netzwerkverkehrs der Testumgebung und zur Integration der untersuchten Angriffserkennungssysteme.

Die Testumgebung stellt eine Infrastruktur zur Beobachtung des gesamten Netzwerkverkehrs bereit. Abbildung 2 zeigt die Architektur, sowie die Integration der untersuchten Systeme (NIDS-1 bis NIDS-4): Auf der linken Seite veranschaulicht die Abbildung die physischen Teile der drei Automatisierungsnetze *Process*, *Manufacturing* und *DMZ*. Über die drei Industrie-Switche dieser Netze (*sw-dmz*, *sw-manufacturing*, *sw-process*) sind die Verbindungen von Mirrorports zu physischen Sensoren durch gestrichelte Linien gekennzeichnet. Bei diesen Sensoren (*sensor-dmz*, *sensor-manufacturing*, *sensor-process*) handelt es sich um Minirechner (Revolution Pi), die jeweils den empfangenen Mirrorstream über GRE-Tunnel in die Virtualisierungsumgebung weiterleiten. Eine virtuelle Maschine (*sensor-server-VM*) terminiert die drei GRE-Tunnel und vereint den Mirrorstream der drei physischen Switche mit dem Mirrorstream weiterer, für den Fertigungsprozess relevanter und den entsprechenden Netzwerken zugehöriger virtueller

Maschinen. Die Verarbeitung des Mirrorstreams erfolgt mittels virtueller Switche (Open vSwitch), die mit Hilfe des Tools *Gremon* konfiguriert werden. *Gremon* ist eine am Fraunhofer IOSB entwickelte Python-Implementierung zur flexiblen Konfiguration von Mirroringaufgaben mittels Open vSwitchen. Der gesamte gespiegelte Netzverkehr wird in einer Reihe von VLANs (*Mirror-Process*, *Mirror-Manufacturing*, *Mirror-DMZ*, *Mirror-All*) bereitgestellt. Für die hier durchgeführte Untersuchung von Angriffserkennungssystemen wurde der gesamte Netzwerkverkehr der drei Automatisierungsnetze in dem VLAN *Mirror-All* bereitgestellt. Die virtuelle Maschine *revsensor-sever-VM* leitet diesen Mirrorstream über einen GRE-Tunnel aus der Virtualisierungsumgebung an einen weiteren Minirechner (*revsensor-server*) weiter, der den GRE-Tunnel terminiert und den Mirrorstream über ein physikalisches Interface bereitstellt. Über einen Hub wurden die untersuchten Angriffserkennungssysteme angeschlossen. Somit wird sichergestellt, dass alle evaluierten Systeme den gleichen Traffic zeitgleich erhalten.

4.2. Evaluationsvorgehen

Die zu untersuchenden Systeme wurden mit dem vollständigen Netzwerkverkehr der Zielumgebung versorgt. Auf dieser Basis wurden die folgenden Tests durchgeführt:

1. **Initiale Lernphase**: Während der Lernphase wurden die zu untersuchenden Systeme über die Dauer von ca. einer Woche dem gesamten Netzwerkverkehr ausgesetzt. Anschließend wurden die erlernten Modelle einer Inspektion unterzogen und manuelle Anpassungen vorgenommen.
2. **Erkennung *normaler* Änderungen in der Zielumgebung**: Die Gruppe dieser Tests umfasste Änderungen in dem zu schützenden System, die während des normalen Netzbetriebes vorkommen können. Beispielsweise wurde ein Hardwareaustausch eines Buskopplers durchgeführt. Ebenfalls wurden neue Netzwerkteilnehmer in die Zielumgebung integriert und neue Kommunikationsbeziehungen aufgebaut.
3. **Erkennung von Cyber-Angriffen in der Zielumgebung**: In der letzten Testphase wurden *bösartige* Anomalien und Angriffe in der Testumgebung erzeugt und die Reaktion der zu untersuchenden Systeme analysiert. Es wurden *Netzwerkscans*, *Man-In-the-Middle-Angriffe* mit anschließender Datenmanipulation und die Exfiltration von sensiblen Daten aus dem Prozessnetzwerk durchgeführt. Zusätzlich wurde ein *Download-Angriff*, welcher die Programmierung einer speicherprogrammierbaren Steuerung (SPS) zu Lasten von Safety-Funktionalitäten verändert, und ein *Rename-Angriff*, welcher durch Veränderung des Gerätenamens eines Motors die Kommunikation zwischen der zuständigen SPS und dem Motor unterbricht, durchgeführt.

5. Grundsätzliche Eigenschaften

5.1. Passive Discovery

Eine initiale passive Lernphase ist den untersuchten vier Systemen gemeinsam. Ziel dieser Lernphase ist die Generierung eines Modells des zu schützenden Zielsystems, wel-

ches dessen Normalverhalten widerspiegelt. Hierbei werden die Hosts und Kommunikationsbeziehungen (Links, Protokolle, Payload) bestimmt.

Die Qualität des erlernten Modells hängt unter anderem davon ab, dass alle Kommunikationsvorgänge während der Lernphase auch vorkommen. Unsere Erfahrung ist, dass auch in einer einfachen Testumgebung Bedienhandlungen, beispielsweise über ein HMI (Human Machine Interface), oder Managementaktivitäten, wie Zugriffe auf Weboberflächen zur Konfiguration von industriellen Komponenten, während der Lernphase nicht oder nur teilweise durchgeführt werden. In größeren Umgebungen mit detaillierten Bedienmöglichkeiten wird auch bei sorgfältigem Vorgehen eine Vollständigkeit nur schwer zu erreichen sein. Ein weiteres typisches Merkmal bei Automatisierungsprotokollen ist die lange Dauer von Verbindungen; im Extremfall wird eine Verbindung beim Systemstart aufgebaut und bis zur Abschaltung beibehalten. Dies ist für IT Protokolle und IT Betrieb ungewöhnlich und führt im Zusammenhang mit passiven Lernverfahren dazu, dass solche Verbindungen während der Lernphase oftmals nicht erkannt werden. Ein Ein/Ausschalten des gesamten Prozesses ist daher während der Lernphase notwendig.

Ferner sei darauf hingewiesen, dass natürlich vorausgesetzt wird, dass während der Lernphase keine Angriffe stattfinden.

5.2. Rückwirkungen auf das Zielsystem

Nach Abschluss der Lernphase sollte das erlernte Modell hinsichtlich seiner Korrektheit inspiziert werden. Ein entscheidendes Kriterium ist hier die Dokumentation des Modellschemas und die Transparenz der Modelldarstellung.

Das Ziel sollte ein Normalbetrieb sein, der im angriffsfreien Zustand keine Alarme erzeugt. Fehlerhafte Netzwerk- oder Gerätekonfigurationen können hingegen dazu führen, dass dieser stabile Normalzustand nicht erreicht wird. So suchen Windows Systeme in ihrer Normalkonfiguration den Kontakt zu Microsoft Servern. Hintergrund hierfür kann der Download von Updates oder die Übermittlung von Telemetriedaten sein. Diese Funktionalitäten sind in Automatisierungssystem in der Regel unerwünscht und diesen wird oftmals einzig durch das Blockieren der Verbindungen durch Firewalls begegnet.

Solche Störungen des Normalbetriebs müssen durch die Systemadministration beseitigt werden, bevor das Angriffserkennungssystem sinnvoll genutzt werden kann. Werden diese Störungen nicht beseitigt, führt dies zur Generierung zahlreicher Meldungen (engl. Alerts) in den Angriffserkennungssystemen, obwohl keine Bedrohung zugrunde liegt.

5.3. Deployment

Die Kombination von Asset Inventory und Angriffserkennung in einem Gerät ist bei der Positionierung des Gerätes in der Netzwerkarchitektur zu beachten. Während üblicherweise die Überwachung des Netzwerkverkehrs an kritischen Servern oder an Netzübergabepunkten (u.U. inline bei einer Firewall) empfohlen wird, erfordert ein Asset Inventory allerdings ein vollständiges Bild des Netzwerkverkehrs (z.B. über den Mirrorport eines Switches).

5.4. Anomalie- und Angriffserkennung

Als primäres Detektionsverfahren ist den vier untersuchten Systemen die Anomalieerkennung auf der Basis eines durch passive Beobachtung des Netzwerkverkehrs gelernten Modells gemeinsam. Als Anomalie werden bei allen untersuchten Systemen neue bzw. vom erlernten Modell abweichende Kommunikationsbeziehungen, Operationen, Protokolle, Netzwerkteilnehmer, Parameterbelegungen oder Antwortzeiten verstanden.

Neben der Anomalieerkennung sind indikatorbasierte Verfahren gängige Praxis zur Detektion von Angriffen. Hierbei wird nach Informationselementen gesucht, die im Zusammenhang mit Cyber-Angriffen beobachtet wurden, sogenannten *Indicators of Compromise.* Nur zwei der vier untersuchten Systeme setzen indikatorbasierte Verfahren ein und können somit als hybride Systeme eingestuft werden. Allerdings operieren die beiden Verfahren unabhängig voneinander, also ohne übergreifende Integration.

Untersucht wurden „gutartige“ und „bösartige“ Anomalien und Angriffe im Netzwerk. Hinsichtlich der Ergebnisse gibt es bei den untersuchten Systemen zum Teil große Unterschiede:

- Die Analyse der Protokolldaten auf der Protokollschicht 7 (Deep Packet Inspection - DPI) weist erhebliche Unterschiede auf. So verfügen manche Systeme für bestimmte Netzwerkprotokolle über gar keine DPI-Engine, wobei andere Systeme neben einer DPI-Engine darauf aufbauende statistische Analysen zur Anomalieerkennung ermöglichen.
- Eine Abstraktion (Korrelation) der Alarmmeldungen wird nicht unterstützt.
- Die Varianz der generierten Meldungen zwischen den untersuchten Systemen ist erheblich. So erzeugten manche Testsysteme bis zu 20 000 Meldungen bei einem einzigen Angriff, wobei andere Systeme bei demselben Angriff nur eine Meldung generierten.

5.5. Adaptability

Die Voraussetzungen hinsichtlich der statischen Natur des ICS Netzes sind in der Praxis nicht immer gegeben. Deshalb bieten die Systeme Möglichkeiten, die erlernten Modelle auch während der Detektionsphase weiter zu ergänzen. Die verschiedenen Systeme unterscheiden sich jedoch hinsichtlich der Möglichkeiten manueller Nachbearbeitung.

Während das Einfügen neuer Hosts und neuer Kommunikationsbeziehungen relativ einfach möglich ist, erfordert die Berücksichtigung von nicht mehr genutzten Kommunikationsbeziehungen und Hosts einen höheren Aufwand. Nicht mehr genutzte Objekte müssen explizit hinsichtlich ihrer letzten beobachteten Aktivität gesucht und gelöscht werden. Dies gilt für alle untersuchten Systeme.

5.6. Actionability

Anomalieerkennungssysteme haben grundsätzlich das Problem des „semantic gap“. Damit wird ausgedrückt, dass das System kein Wissen um die Bedeutung einer Anomalie besitzt, alle Anomalie-Meldungen sind gleichwertig. Wenn z.B. ein neuer Teilnehmer

im Netzwerk 50 Anomalien-Meldungen erzeugt, ist ein Operator gefordert, die letztendlich Ursache (*root cause*) herauszufinden (neue IP Adresse und neue MAC Adresse).

5.7. Active Defense

Ein vollständiges Spektrum von Verteidigungsmaßnahmen bei IT Infrastrukturen beschreibt R. Lee [2] mit „The sliding scale of cyber security“, dabei werden 5 Kategorien in aufsteigender Reihenfolge unterschieden: Architektur, Passive Defense, Active Defense, Intelligence und Offense. In die ersten Kategorien gehören Maßnahmen wie Netzwerk-Separierung, Firewalls, Antivirenprogramme und Härtungsmaßnahmen, die alle ohne kontinuierliche Betreuung auskommen. Intrusion Detection Systeme, wie sie hier betrachtet werden, fallen dabei in die Kategorie Active Defense, da sie kontinuierliche Betreuung erfordern.

Der Betrieb erfordert IT Security Fachpersonal, welches z.B. kontinuierlich Meldungen analysiert und entsprechend reagiert. Dazu kann auch die Analyse von PCAP Files mit Wireshark[5] gehören. Während größere Unternehmen über die nötigen Ressourcen und Infrastrukturen (z.B. SIEM) verfügen, bietet sich für kleinere Unternehmen die Nutzung eines entsprechenden Dienstleisters (Managed Security Service Provider MSSP) an, der über die nötigen Fachkenntnisse und die nötige Infrastrukturausstattung verfügt.

6. Zusammenfassung

Das Sicherheitsmanagement in ICS Netzwerken ist eine komplexe Aufgabe, deren Lösung noch nicht wirklich absehbar ist. Monitoring- und Anomalieerkennungssysteme, wie die hier betrachteten, sind möglicherweise nur ein Teil der Lösung. Von einer Vision bezüglich der Gesamtlösung hängt die Bewertung aktueller Systeme ab. Grundsätzlich zeichnen sich drei Alternativen ab, denen die untersuchten Systeme in unterschiedlichem Maße entsprechen:

- Stand-alone Lösungen: Systeme wie die hier vorstellten werden als primäres und möglichweise auch einziges ICS Angriffserkennungssystem eingesetzt. Dies setzt Funktionaliäten voraus, die über eine reine Anomalieerkennung hinausgehen und Ereigniskorrelation und Ereignismanagement bieten.
- OT SOC/SIEM: Systeme wie die hier vorgestellten werden als ein Angriffserkennungssystem neben anderen eingesetzt und die gewonnenen Daten werden in einem OT spezifischen SIEM zusammengefasst und dort analysiert. Dieses Szenario erfordert zur Integration eine entsprechende und hinreichend reichhaltige Datenschnittstelle, gleichzeitig wird eine meistens vorhandene Web GUI des Angriffserkennungssysteme weniger wichtig.
- Enterprise SOC: Systeme wie die hier vorgestellten Angriffserkennungssysteme leiten ihre Daten an ein Enterprise SIEM weiter. Dort werden sie mit allen anderen Security Ereignissen zusammengefasst und ausgewertet. Hier gilt die obige Bemerkung hinsichtlich Integration in gleicher Weise.

[5] https://www.wireshark.org

Die Alternativen betreffen die Organisationsstruktur eines Unternehmens hinsichtlich seines Security Managements und werden sicher nicht einheitlich ausfallen. Entsprechend unterschiedlich werden auch die Präferenzen für die einzelnen technischen Hilfsmittel ausfallen.

Literaturhinweise

[1] S. Caltagirone and R. Lee: The four types of threat detection. Dragos Whitepaper, 2018.

[2] R. Lee: The Sliding Scale of Cyber Security. SANS Institute, 2015.

[3] T.J. Williams: The Purdue enterprise reference architecture: a technical guide for CIM planning and implementation. Research Triangle Park, NC, 1992.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Certification Path Validation Test Tool (CPT) – Ein Tool zur Überprüfung der X.509-Zertifizierungspfadvalidierung

Dr. Heike Hagemeier[1], Dr. Evangelos Karatsiolis[2], Dr. Falko Strenzke[2]

Kurzfassung:
In dieser Arbeit stellen wir ein neues Testwerkzeug zum Testen der X.509-Zertifizierungspfadvalidierung, das Certification Path Validation Test Tool (CPT), vor, welches für das Bundesamt für Sicherheit in der Informationstechnik (BSI) entwickelt wurde. Wir erläutern vor dem Hintergrund bestehender Testwerkzeuge für diesen Zweck die Notwendigkeit eines neuen Testwerkzeugs, welches in der Lage ist, Testzertifikate für bestimmte Fehlerfälle flexibel und dynamisch zu erzeugen. Außerdem legen wir die grundlegende Funktionsweise des CPT zur Erzeugung von Testzertifikaten, Sperrlisten und OCSP-Antworten dar. Ferner beschreiben wir zusammen mit dem CPT veröffentlichte Erweiterungen zum Testen von TLS- und IPsec-Anwendungen. Wir erläutern die mit dem CPT ausgelieferte Testspezifikation und die damit an zehn Testgegenständen gefundenen Fehler.

Stichworte: BSI TR-02103, Public-Key Infrastrukturen, Softwaretest, TR-02103, X.509-Zertifikate, Zertifizierungspfadvalidierung

1. Einführung

In der digitalen Kommunikation werden Zertifikate zur Authentifizierung und zur Verifizierung von öffentlichen Schlüsseln verwendet. Diese Zertifikate binden den öffentlichen Schlüssel an die Identität seines Eigentümers innerhalb einer Public-Key-Infrastruktur (PKI). Der gebräuchlichste Standard für digitale Zertifikate ist X.509v3 [1]. Er definiert ein Framework für Public-Key-Infrastrukturen und digitale Zertifikate. RFC 5280 [2] spezifiziert die X.509-Zertifikats- und Sperrlistenprofile zur Verwendung im Internet. Dort sind die jeweiligen Datenformate und ein detaillierter Algorithmus zur Validierung von Zertifizierungspfaden (Zertifizierungspfadvalidierung oder kurz „Pfadvalidierung“) beschrieben. Dabei stellt ein Zertifikatspfad die lückenlose Kette von einem als Vertrauensanker fungierenden Wurzel-Aussteller-Zertifikat (Zertifikat der Root Certification Authority, kurz „Root-CA“) über gegebenenfalls mehrere Zwischen-CA- oder Sub-CA-Zertifikate bis zu einem Nutzerzertifikat dar. Abbildung 1 veranschaulicht eine PKI-Struktur mit mehreren solchen Zertifikatspfaden. Darin sind die Zertifikate dargestellt als gerichtete Kanten zwischen den als Rechtecken dargestellten Entitäten. Der Pfad von der Root CA als Vertrauensanker zum Zertifikat von Alice enthält beispielsweise die Zertifikate #2, #3 und #5. In dieser Kette ist bis auf den Vertrauensanker die Signatur eines jeden Zertifikats jeweils durch den öffentlichen Schlüssel aus dem vorhergehenden Zertifikat verifizierbar. Zur Validierung eines Zertifikats muss im Allgemeinen zuerst der Pfad zu einem Vertrauensanker mit Hilfe von verfügbaren Zwischen-CA-Zertifikaten konstruiert werden. Anschließend muss der Pfad basierend auf den allgemeinen Vorgaben für Zertifikate nach RFC 5280 validiert werden.

[1] Bundesamt für Sicherheit in der Informationstechnik, Bonn
[2] MTG AG, Darmstadt

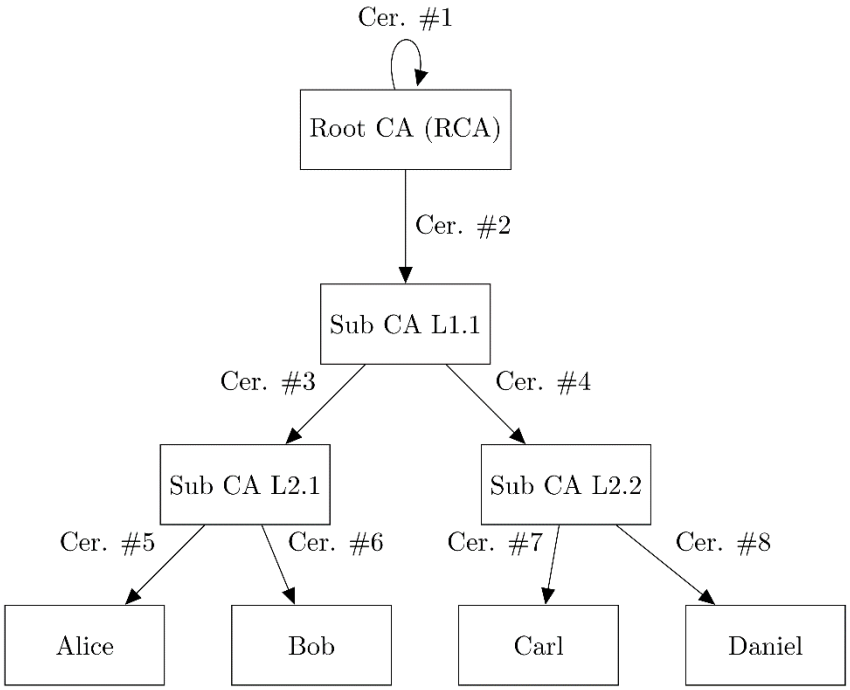


Abbildung 1: Beispiel einer PKI-Hierarchie

Ferner können je nach Anwendung noch weitere protokollspezifische Vorgaben dazukommen, wie z.B. für TLS, IPsec oder S/MIME. Schließlich muss noch eine Revokationsprüfung erfolgen, d.h. es muss geprüft werden, ob das Zertifikat eventuell gesperrt wurde.

Obwohl die entsprechenden Standards schon lange etabliert sind und der Einsatz von X.509 Zertifikaten inzwischen weit verbreitet ist, wurden in den letzten Jahren viele Fehler in Implementierungen der Pfadvalidierung in kryptografischen Bibliotheken und Anwendungen gemeldet. Diese Fehler traten aufgrund falscher Interpretationen des Standards oder aufgrund von Programmierfehlern auf. Forscher der Stanford University und der University of Texas untersuchten die Implementierung der Pfadvalidierung in kryptografischen Bibliotheken (z.B. OpenSSL [3]) und Datentransport-Bibliotheken (z.B. Apache HttpClient [4]) und kamen zu dem Ergebnis "Our main conclusion is that SSL certificate validation is completely broken in many critical software applications and libraries." [5].

Daraus leitet sich die Notwendigkeit des systematischen Testens von Implementierungen der Pfadvalidierung ab. Öffentlich verfügbare Testwerkzeuge für diesen Zweck existierten bereits vor der Erstellung des CPT, diese sind jedoch, wie in Abschnitt 2 dargelegt wird, nicht universell einsetzbar.

Vor dem Hintergrund dieser Situation haben wir im Rahmen eines Projekts des Bundesamts für Sicherheit in der Informationstechnik (BSI) das Certification Path Validation Test Tool (kurz CPT) entwickelt, um ein universelles Testwerkzeug bereitzustellen, mit dem der Fehleranfälligkeit von Implementierungen der Pfadvalidierung begegnet werden kann. Dieses Tool erzeugt X.509 Zertifikate, Sperrlisten und OCSP-Antworten,

welche gültige und ungültige Zertifizierungspfade darstellen, anhand einer sehr flexiblen und leicht erweiterbaren Testspezifikation. Das Tool ist mit einer umfangreichen Sammlung von Standard-Testfällen ausgestattet, welche Aspekte bezüglich Sicherheit, Interoperabilität und der kryptografischen Stärke abdecken. Es unterstützt auch das automatisierte Testen von TLS-Client- und Server-Anwendungen. Details zum CPT beschreiben wir in Abschnitt 3.

In Abschnitt 4 beschreiben wir die mit dem CPT ausgelieferte Testsuite. Wir haben deren Testfälle und Testdaten auf mehrere Bibliotheken und Anwendungen angewendet. In Abschnitt 5 stellen wir die dabei gewonnenen Testergebnisse vor. In Abschnitt 6 geben wir eine Zusammenfassung und einen Ausblick.

2. Vorherige Arbeiten

Im Rahmen des BSI-Projekts wurde eine Übersicht über bereits existierende Tools erstellt, mit denen sich Implementierungen der Zertifizierungspfadvalidierung überprüfen lassen. Dabei hat sich gezeigt, dass lediglich eine gute Handvoll von Tools zu Verfügung stand, nämlich x509test [6], Public Key Interoperability Test Suite (PKI-Testsuite von NIST) [7], Certificate Fuzzer [8], Frankencerts [9], TLSPretense [10] sowie generische Fuzzing Tools wie beispielsweise Go Fuzz [11]. Die beiden vielversprechendsten Ansätze waren dabei Frankencerts sowie die NIST PKI-Testsuite. Beide Tools sind aufgrund gewisser Einschränkungen nicht universell in allen Projekten für Entwicklungstests einsetzbar.

In der Frankencerts Veröffentlichung [9] schlagen die Autoren eine Methode zum automatisierten Testen der Routinen zur Pfadvalidierung in kryptografischen Bibliotheken und Anwendungen vor. Die Hauptidee ist, Teile gültiger X.509-Zertifikate zufällig auszuwählen und diese anschließend ebenso zufällig zu kombinieren, um neue Zertifikate zu erzeugen, sogenannte „Frankencerts“. Der somit erzeugte Testdatensatz enthält sowohl gültige als auch ungültige Zertifikate. Diese Zertifikate werden dann von einem bestimmten Satz an zu testenden Implementierungen validiert. Wenn für ein bestimmtes Testzertifikat abweichende Ergebnisse bzgl. dessen Gültigkeit bei der Pfadvalidierung innerhalb der getesteten Implementierungen auftreten, dann wird eine nähere Analyse durchgeführt, um festzustellen, ob ein Fehler in einer der Implementierungen vorliegt. Dies ist eine Form des sogenannten differenziellen Testens.

Der am weitesten verbreitete Testdatensatz zum Testen der Pfadvalidierung ist die PKI-Testsuite von NIST [7]. Viele Bibliotheken und Anwendungen testen ihren Validierungsalgorithmus bereits gegen diese Testsuite (z.B. Mozilla, Android und GnuTLS). Sie spezifiziert über einhundert Testfälle und enthält einen Pool von etwa 400 Zertifikaten mit den zugehörigen geheimen Schlüsseln und 170 Sperrlisten. Als kryptografische Verfahren kommen dabei nur RSA und DSA vor. Es werden somit rein statische Testdaten in Form einer Dateisammlung zur Verfügung gestellt. Die Testfälle sind durch textuelle Beschreibungen gegeben. Um eine Anwendung damit zu testen, müssen die Testdaten für jeden einzelnen Testfall zusammengestellt werden und in geeigneter Weise der Validierungsroutine der Anwendung zugeführt werden.

Beide Testwerkzeuge sind aufgrund ihrer Einschränkungen nicht universell durch Softwareentwickler einsetzbar. Der Frankencerts-Ansatz des differenziellen Testens erfordert eine zu hohe fachliche Versiertheit des Testers. Ferner sind beim differenziellen Testen die Ergebnisse immer abhängig von den eingesetzten Referenzimplementierungen. Die NIST PKI-Testsuite dagegen ist zwar vom Grundansatz genau auf den Einsatz als Testdaten in Entwicklungsprojekten ausgerichtet, allerdings mangelt es hier an Flexibilität. So würden sich beispielsweise damit keine modernen Anwendungen testen lassen, die nur Elliptische Kurven Kryptografie unterstützen. Ferner ist das Arrangement der Testdaten zu den einzelnen Testfällen aus der textuellen Beschreibung mit einem nicht unerheblichen Aufwand verbunden, welcher den Einsatz dieser Testsuite erschwert.

3. Das CPT

Vor dem Hintergrund dieser Einschränkungen der existierenden Werkzeuge wurde das CPT entwickelt. Der Grundgedanke war es, eine flexible Beschreibung der Testdaten für jeden Testfall bereitzustellen, aus welcher die Testdaten durch das Tool dynamisch erzeugt werden. Auf diese Weise können die Testdaten auf die Erfordernisse eines Entwicklungsprojektes angepasst werden, indem beispielsweise andere kryptografische Algorithmen oder spezielle Zertifikatserweiterungen für sämtliche Zertifikate spezifiziert werden. Das CPT bietet eine Basiskomponente, welche dafür zuständig ist, basierend auf den formalen Testdatenspezifikationen im XML-Format die verschiedenen PKI-Objekte wie Zertifikate und Sperrlisten zu erstellen, damit diese als Eingabe für Testgegenstände benutzt werden können. Die grundlegende Funktion der Basiskomponente ist es, aus einer Verzeichnisstruktur unterhalb ihres Hauptverzeichnisses die Testdatenspezifikationen einzulesen und darauf basierend in einem Ausgabeverzeichnis in je einem Unterordner für jeden einzelnen Testfall die entsprechenden PKI-Objekte zu erzeugen.

Die Basiskomponente des CPT wird in Abschnitt 3.1 beschrieben. Neben der Basiskomponente bietet das Tool Erweiterungen, um das Testen von Anwendungen wie TLS und IPsec zu unterstützen. Diese werden in Abschnitt 3.2 beschrieben.

3.1. Das CPT Basistool

Die Hauptfunktion des CPT ist es, Zertifikate, Sperrlisten und OCSP-Antworten zu Testzwecken zu erstellen. Die erzeugten Objekte bilden eine PKI-Struktur. Ein Beispiel einer solchen Struktur ist in Abbildung 1 dargestellt. Eine derartige Struktur enthält mehrere Zertifizierungspfade. Die vom CPT erstellten Pfade enthalten potenziell sowohl gültige Zertifikate, deren Zertifikatspfad erfolgreich geprüft werden kann, als auch Zertifikate, welche Fehler aufweisen oder aufgrund der Inhalte anderer Zertifikate im Pfad nicht erfolgreich validiert werden können. Beispiele für solche Fehler sind eine falsche Kodierung eines Zertifikats, eine fehlende vorgeschriebene Zertifikatserweiterung oder ein abgelaufenes Zertifikat. Auf diese Weise realisiert das CPT sowohl Positiv- als auch Negativtests zur Überprüfung der Zertifizierungspfadvalidierung.

Um die Möglichkeit zu haben, verschiedene Testfälle abzubilden und die Gestaltung der Pfade und der Zertifikate, Sperrlisten und OCSP-Antworten als PKI-Objekte möglichst

flexibel zu halten, werden die Zertifikatsinhalte in einer XML-basierten Beschreibung dargestellt. Diese Darstellung erlaubt es, alle in diesen Objekten enthaltenen Datenfelder zu spezifizieren, wie beispielsweise die Namen von Zertifikatsinhaber und -aussteller, enthaltene Zertifikatserweiterungen, usw. Die Gültigkeitszeitpunkte werden dabei relativ zum Zeitpunkt der Erzeugung angegeben. Somit ist es möglich, immer aktuelle Testdaten zu generieren. Um geeignete Testdaten zu erzeugen, können Benutzer:innen neben gültigen Werten auch ungültige Werte wie zum Beispiel eine ungültige Versionsnummer oder eine falsche Kodierung für ein Datumsfeld konfigurieren. Alle gängigen kryptografischen Algorithmen wie ECDSA, ECDH und RSA werden dabei als Signaturalgorithmen unterstützt. Ferner können in den Testfällen Variablen benutzt werden, deren konkreter Wert in einer zentralen Konfigurationsdatei gesetzt werden kann. Somit ist es beispielsweise möglich, auf einfache Weise die Signaturalgorithmen aller Testfälle zu ändern. Es ist auch möglich, bestehende Zertifikatsdefinitionen aus anderen Testfällen als Referenz zu verwenden, um neue Zertifikate zu spezifizieren. In diesem Fall müssen nur die jeweils abweichenden Felder explizit angegeben werden.

Listing 1 gibt die XML-Beschreibung des Pfades vom Vertrauensanker zum Zertifikat von Alice aus Abbildung 1 wieder. Ein PKIObjects-Element ist für die Aufnahme verschiedener Objekte verantwortlich, die innerhalb einer PKI verwendet werden, wie z.B. Zertifikate (XML-Element Certificate) und Sperrlisten (XML-Element CRL). Innerhalb jedes Elements können dessen Basisfelder und Erweiterungen spezifiziert werden; z.B. wird einem Zertifikat eine Version, eine Seriennummer oder eine Aussteller-DN zugewiesen. Dabei ist es möglich, diese Felder so zu spezifizieren, dass sie in Bezug auf die geltenden Standards fehlerhaft sind.

Neben den Zertifikaten und Sperrlisten ist es auch möglich, OCSP-Antworten als XML-Elemente zu definieren. Abgesehen von der Erzeugung der verschiedenen PKI-Objekte, erzeugt das CPT auch signierte E-Mails basierend auf dem S/MIME Protokoll [12], um das Testen von E-Mail-Anwendungen zu unterstützen. Außerdem bietet es die Funktion, die zu den Testfällen gehörenden OCSP-Antworten mittels eines OCSP-Servers bereitzustellen. Zusätzlich stellt es die ausgestellten Sperrlisten über HTTP und LDAP zur Verfügung. Das erlaubt es Anwendungen, über ihre Standardwege die Sperrlisten und OCSP-Antworten einzuholen. Damit wird ein realistisches Testszenario ermöglicht. Eine ausführliche Beschreibung des CPT ist in der Dokumentation auf der CPT-Seite [11] des BSI zu finden.

Das CPT hat einige Vorteile gegenüber den bisher vorhandenen Testwerkzeugen. Im Gegensatz zur NIST-Suite ist das Tool in der Lage, Zertifikate mit anderen kryptografischen Algorithmen als RSA und DSA zu erstellen. Es bietet im Gegensatz zu den beiden anderen Testwerkzeugen auch Unterstützung für das OCSP-Protokoll. Außerdem enthält es Tests, um die Verwundbarkeit einer Implementierung der Pfadvalidierung gegenüber kryptografischen Angriffen wie dem Low-Exponent Angriff von Bleichenbacher [13] oder der „null prefix attack“ [14] zu überprüfen. Darüber hinaus ist es für den Nutzer möglich, zusätzliche Testfälle zu spezifizieren. Da das Tool die Daten dynamisch aus der XML-Testspezifikation erzeugt, können bei Bedarf immer aktuelle Zertifikate,

```
<PKIObjects>
  <Certificate id="TEST_001_CP_001_ROOT_CA" type="TA">
    <VerifiedBy>TEST_001_CP_001_ROOT_CA</VerifiedBy>
    <Version>2</Version>
    <SerialNumber>1</SerialNumber>
    <Signature>1.2.840.113549.1.1.11</Signature>
    <IssuerDN encoding="UTF-8">CN=Root CA, C=DE</IssuerDN>
    <SubjectDN encoding="UTF-8">CN=Root CA, C=DE</SubjectDN>
    <NotBefore>-3D</NotBefore>
    <NotAfter>+5Y</NotAfter>
    <PublicKey>RSA,2048</PublicKey>
    <Extension oid="2.5.29.35" critical="false" name="AKI" type="pretty"></Extension>
    <Extension oid="2.5.29.14" critical="false" name="SKI" type="pretty"></Extension>
    <Extension oid="2.5.29.15" critical="true" name="KU" type="pretty">keyCertSign</Extension>
    <Extension oid="2.5.29.32" critical="true" name="CP" type="pretty">1.2.3.4</Extension>
    <Extension oid="2.5.29.19" critical="true" name="BC" type="pretty">true,2</Extension>
  </Certificate>
  <Certificate id="TEST_001_CP_001_SUB_CA_L1.1">...</Certificate>
  <Certificate id="TEST_001_CP_001_SUB_CA_L2.1">...</Certificate>
  <Certificate id="TEST_001_CP_001_EE_ALICE" type="TC">...</Certificate>
  <CRL>
    <VerifiedBy>TEST_001_CP_001_SUB_CA_L2.1</VerifiedBy>
    <Version>1</Version>
    <Signature>1.2.840.113549.1.1.11</Signature>
    <IssuerDN encoding="UTF-8">CN=Sub CA L2.1, C=DE</IssuerDN>
    <ThisUpdate>-1D</ThisUpdate>
    <NextUpdate>+6D</NextUpdate>
    <RevokedCertificate refid="TEST_001_CP_001_EE_ALICE" />
    <Extension oid="2.5.29.20" critical="false" name="CRL Number" type="pretty">11</Extension>
    <Extension oid="2.5.29.35" critical="false" name="AKI" type="pretty"></Extension>
  </CRL>
</PKIObjects>
```

Listing 1: Modellierung des Pfads zum Zertifikat von Alice mittels der XML-Beschreibung von CPT.

Sperrlisten und OCSP-Antworten erzeugt werden. Im Gegensatz zum Ansatz des Frankencerts-Forschungsteams können mit den Testfällen die Anwendungen auf spezifische Fehler getestet werden und das Testergebnis ist direkt verfügbar. Darüber hinaus bieten die mit dem CPT ausgelieferten Testfälle auch anwendungsspezifische Tests, was bei den bisher existierenden Testwerkzeugen nicht gegeben ist. Die Testfälle sind durch das leicht verständliche XML-Format durch Anwender:innen leicht erweiterbar. Dies ermöglicht es, die Testfälle anzupassen, um z.B. verschiedene kryptografische Algorithmen zu unterstützen und komplexe Public-Key-Infrastrukturen zu modellieren.

3.2. Die CPT Erweiterungen

Um spezielle Anwendungen mit den durch das CPT erstellten Testfällen zu testen, ist es notwendig, die erstellten Zertifikatsketten der Zertifikatspfadvalidierung der entsprechenden Anwendung zuzuführen. Da diese Zertifikatsketten im Allgemeinen Fehler enthalten, ist es nicht ohne weiteres möglich, beispielsweise eine existierende Webserver-Implementierung mit TLS-Unterstützung mit einer fehlerhaften Zertifikatskette zu konfigurieren, um damit einen TLS-Client zu testen. Im Normalfall wird der Server die fehlerhafte Zertifikatskette nicht akzeptieren. Aus diesem Grund wurde sowohl ein Werkzeug für das Testen von TLS-Client und -Server Anwendungen als auch für IPsec-Anwendungen erstellt, um zwei der weitverbreitetsten Protokolle für sichere Kommunikation in Netzwerken abzudecken.

Die CPT-Erweiterungen sowie die zugehörige Dokumentation können von der CPT-Seite des BSI [15] bezogen werden.

3.2.1. Die Testwerkzeuge für TLS-Anwendungen

Die TLS-Testwerkzeuge wurden auf Basis der in der Bibliothek Botan [16] integrierten Kommandozeilen-basierten TLS-Client und -Server Implementierung erstellt. Bei der Testausführung liest die Implementierung die von der Basiskomponente für den entsprechenden Testfall erzeugte Zertifikatskette ein und präsentiert diese während des TLS-Handshakes. Anhand des Ablaufs des Handshakes wird bewertet, ob die Gegenseite die präsentierte Zertifikatskette akzeptiert hat oder nicht. Wird der Handshake abgebrochen bzw. dabei eine entsprechende Fehlernachricht (ein TLS-Alert in der Stufe „Fatal“) von der Gegenseite geschickt, so wird dies als Abbruch des Handshakes gewertet.

Der TLS-Test-Server unterstützt ferner OCSP-Stapling nach [17].

3.2.2. Das Testwerkzeug für IPsec Anwendungen

Das Testwerkzeug für IPsec Anwendungen basiert auf einer modifizierten Variante der IPsec Implementierung strongSwan in der Version 5.5.3. Die Modifikationen bestehen in der Deaktivierung von einigen Prüfungen der eingesetzten Zertifikate und sind notwendig, um die Verwendung von fehlerhaften X.509-Zertifikaten im Rahmen der Tests zuzulassen. Die Testausführung mit dem IPsec Testwerkzeug erfolgt nach ähnlichen Prinzipien wie beim TLS Testwerkzeug.

4. Die mitgelieferte Testsuite des CPT

Die mit dem CPT mitgelieferte Testsuite enthält insgesamt 94 Testfälle. Um einen Testfall zu beschreiben, wird die XML basierte Testspezifikation aus Abschnitt 5 der TR-03124-2 [18] benutzt. Die Testsuite ist unterteilt in verschiedene Module, von welchen beispielsweise eines die generelle, anwendungsunspezifische Verarbeitung von Zertifikaten prüft; weitere decken die Verarbeitung von Sperrlisten und OCSP-Antworten ab. Für Tests von kryptografischen Aspekten existieren zwei weitere Module. Des Weiteren gibt es je ein Modul für TLS-Client, TLS-Server, IPsec und S/MIME.

5. Testergebnisse

Die im CPT integrierten Testfälle wurden gegen zehn Testgegenstände ausgeführt. Darunter waren fünf kryptografische Bibliotheken, und zwar Botan in einer Vorversion des Releases 2.2.0, identifiziert durch den git commit 5b2fe4a6d4dfdb28af364eec86a407327e64d1d7, mbedTLS in der Version 2.4.2, OpenSSL in der Version 1.1.0e, Bouncy Castle in der Version 1.57 und OpenJDK in der Version 1.8.0_121. Zudem wurden die fünf Anwendungen Apache HTTP Server in der Version 2.4.18 (Ubuntu), der Firefox Webbrowser in der Version 55.0.2, die StrongSwan IPsec Implementierung in der Version 5.5.3, der KMail Mailclient in der Version 5.1.3 mit Kleopatra in der Version 2.2.0, und die VPN-Implementierung OpenVPN in der Version 2.3.4-5+deb8u2 getestet.

Die ersten drei der genannten Bibliotheken sind in C/C++ implementiert, die letzten beiden in Java. Es wurde für beide Sprachen jeweils eine Testapplikation erstellt, welche die entsprechenden Bibliotheken einbindet, die Zertifikatsvalidierung mit den vom CPT generierten Testfällen anstößt und die Rückgabewerte auswertet.

Als einziger Testgegenstand wies OpenJDK keinerlei Fehler auf. Dessen Implementierung der Zertifizierungspfadvalidierung orientiert sich strikt an dem formalen Algorithmus aus [2].

5.1. Erkannte Fehler

Tabelle 1 listet alle Fehler und Risiken der untersuchten Testgegenstände auf, die durch die in das CPT integrierte Testsuite gefunden wurden. Die jeweiligen Hersteller wurden vor der Veröffentlichung über alle gefundenen Fehler informiert. In den folgenden Unterabschnitten werden diese thematisch geordnet vorgestellt.

5.1.1. Fehler im Zusammenhang mit Zertifikatsgültigkeitszeiträumen

VCM_08: Ein schwerwiegender Fehler trat in KMail auf. Die getestete Version akzeptiert Zertifikate für S/MIME Signaturen, deren Gültigkeitszeitraum abgelaufen ist. Damit ist auch das Risiko verknüpft, gesperrte Zertifikate zu akzeptieren, denn diese können nach Ablauf ihres Gültigkeitsdatums von der Sperrliste genommen werden.

5.1.2. Fehler in Bezug auf Zertifikatserweiterungen

VEX_06: Apache und OpenVPN akzeptierten Zwischen-CA-Zertifikate, in welchen die für diesen Zweck notwendige Basic Constraints Zertifikatserweiterung fehlt. Dies ist ein klarer Verstoß gegen die Vorgaben. Allerdings wird bei beiden Implementierungen mindestens sichergestellt, dass die Zertifikate die Key Usage keyCertSign gesetzt haben. Somit wird ausgeschlossen, dass Nutzerzertifikate als CA-Zertifikate verwendet werden können.

VEX_11: strongSwan akzeptierte Zertifikate ohne die Key Usage keyCertSign als Zwischen-CA-Zertifikate. Dies ist ebenfalls eine Verletzung der Spezifikationen. Allerdings stellt die Implementierung mindestens sicher, dass die Basic Constraints Erweiterung vorhanden ist und das Zertifikat als CA-Zertifikat ausweist, so dass auch in diesem Fall keine Nutzerzertifikate missbräuchlich als CA-Zertifikate verwendet werden können.

5.1.3. Sonstige Fehler in Bezug auf Zertifikate

VCM_11: Die Implementierungen Bouncy Castle, OpenSSL und OpenVPN akzeptierten zukünftige, noch nicht definierte Zertifikatsversionen. Dies stellt insofern einen Fehler dar, als dass die Verarbeitungsregeln für neuere Zertifikatsversionen von denen für v3-Zertifikate abweichen können. Somit müssten unbekannte Zertifikatsversionen korrekterweise abgelehnt werden.

VIP_04: strongSwan verletzte eine der Vorgaben für IPsec aus [19]. Und zwar wird dort für den IKEv1 oder IKEv2 Key Exchange bei IPsec gefordert, dass mindestens eine der beiden Key Usages digitalSignature oder nonRepudiation gesetzt ist. strongSwan akzeptierte jedoch auch Zertifikate ohne eine der beiden Key Usages.

5.1.4. Fehler im Zusammenhang mit der Verarbeitung von Sperrlisten

VCR_01 und VCR_15: Bezogen auf Sperrlisten wurde eine Vielzahl an Fehlern gefunden. Der bedeutsamste wurde bei mbedTLS entdeckt. Diese Bibliothek nimmt zwar über den Funktionsaufruf zur Zertifikatspfadvalidierung Sperrlisten entgegen, ignoriert jedoch die Sperrlisten einfach, wenn diese als ungültig eingestuft werden und akzeptiert das Zertifikat, statt es folgerichtig abzuweisen. Dass dies tatsächlich das intendierte Verhalten der Bibliothek ist, zeigt folgende Aussage in der API-Dokumentation zu der Funktion mbedtls_x509_crt_verify(): „It is your responsibility to provide up-to-date CRLs for all trusted CAs. If no CRL is provided for the CA that was used to sign the certificate, CRL verification is skipped silently, that is without setting any flag.”[3] Damit ist die Überprüfung des Sperrstatus in der getesteten mbedTLS-Version völlig unbrauchbar, da die Prüfung der Eignung der Sperrliste gerade die Aufgabe dieser Funktion wäre. Nach Mitteilung des Fehlers bestätigte der Hersteller uns gegenüber, dass es sich hierbei um ein intendiertes Verhalten der Bibliothek handelt und dieses daher nicht angepasst wird. In strongSwan wurde der gleiche Fehler festgestellt, allerdings nur in Bezug auf Zwischen-CA-Zertifikate, während bei mbedTLS alle Zertifikate betroffen sind.

VCR_06: Sperrlisten können ebenso wie Zertifikate Erweiterungen, die sog. CRL-Extensions, enthalten. Ebenfalls wie bei Zertifikaten können diese als kritisch markiert sein, was besagt, dass eine Anwendung, welche die Sperrliste verarbeitet, diese Erweiterung verarbeiten können muss, um die Sperrliste bei der Prüfung zu berücksichtigen. Sowohl Botan als auch mbedTLS akzeptierten Sperrlisten mit unbekannten Sperrlistenerweiterungen, die als kritisch markiert waren.

VCR_08: Ein weiterer Fehler, der in strongSwan auftrat, ist die Akzeptanz von noch nicht gültigen Sperrlisten. Dies ist allerdings in der Praxis als mit geringer Gefahr verbunden anzusehen.

VCR_13: In Botan, mbedTLS und strongSwan trat ein Fehler bezüglich der Verarbeitung der CRL Distribution Point Erweiterung im Zertifikat und der Issuing Distribution Point Sperrlistenerweiterung auf. Diese müssen bei der Verarbeitung einer Sperrliste für

[3] https://tls.mbed.org/api/group__x509__module.html#ga98ed4504e4f832b735a230acf54fcde3, zugegriffen am 15.12.2020.

das entsprechende Zertifikat auf den gleichen Distribution Point verweisen. Bei den genannten Implementierungen fand diese Prüfung nicht statt, so dass es bei diesen prinzipiell vorkommen könnte, dass eine nicht-geeignete Sperrliste zur Prüfung des Revokationsstatus des Zertifikats verwendet wird.

5.1.5. Nicht-sicherheitskritische Fehler und risikobehaftete Verhaltensweisen

Neben den oben aufgeführten Fehlern, die sämtlich dazu führen, dass nach den entsprechenden Standards abzuweisende Zertifikatspfade akzeptiert werden, wurden weitere Klassen von Problemen an den Testgegenständen gefunden. Diese stellen zwar keine Schwachstellen im eigentlichen Sinn dar, sind aber dennoch von Bedeutung. Bei der ersten Klasse handelt es sich um risikobehaftete Verhaltensweisen der Implementierungen. Diese sind durch einen initialen Buchstaben „R“ kenntlich gemacht. Die zweite Klasse stellt Kompatibilitätsprobleme dar, d.h. Verletzungen von Standards, die zwar kein Sicherheitsrisiko im Sinne der Akzeptanz ungültiger Zertifikatspfade darstellen, wodurch aber die korrekte Funktion der Implementierungen nicht erfüllt ist. Diese sind durch einen führenden initialen Buchstaben „C“ gekennzeichnet.

RCM_14: Die Testdurchführung an den kryptografischen Bibliotheken beinhaltet den Schritt der Pfadkonstruktion, d.h. das Bauen der korrekten Kette vom Nutzerzertifikat bis zu einem Vertrauensanker unter der Verwendung einer beliebigen Anzahl von Zwischen-CA-Zertifikaten ohne bestimmten Vertrauensstatus aus einem entsprechenden Zertifikatspool. Einer der beim CPT mitgelieferten Tests überprüft, ob die Implementierung eine erschöpfende Suche innerhalb der möglichen Zertifikatspfade durchführt, bis ein gültiger Pfad gefunden wurde, oder nur eine Untermenge davon betrachtet wird. Dies wird in dem betreffenden Test dadurch realisiert, dass für das Nutzerzertifikat ein gültiges Zwischen-CA-Zertifikat bereitgestellt wird, welches die Verbindung zu dem gültigen Vertrauensanker herstellt. Daneben wird auch ein ungültiges Zwischen-CA-Zertifikat bereitgestellt, in welchem der gleiche SubjectDN und Public Key wie im gültigen Zwischen-CA-Zertifikat gesetzt sind, welches aber im IssuerDN auf einen nicht existierenden Aussteller verweist. Bei der Testausführung wird der Test mehrfach wiederholt, wobei die Zwischen-CA-Zertifikate jeweils in zufälliger Reihenfolge zu dem Zertifikatspool hinzugefügt werden. Dadurch wird potenziell bei der Suche nach gültigen Pfaden von der Implementierung in manchen Testdurchführungen zuerst das gültige Zwischen-CA-Zertifikat gefunden, in anderen das ungültige. Eine Implementierung, welche die Pfadsuche nach dem Verarbeiten des ungültigen Zwischen-CA-Zertifikats abbricht, ist dann daran zu erkennen, dass das Ergebnis der Pfadvalidierung bei den einzelnen Testausführungen variiert. Dieses Verhalten ist bei manchen Implementierungen offenbar intendiert. Das OpenSSL Team bestätigte dies beispielsweise uns gegenüber.

Ein potenzielles Problem, was aus einer Implementierung mit einer solchermaßen nicht erschöpfenden Pfadsuche resultieren kann, ist, dass ein Angreifer, der in der Lage ist, ungültige Zwischen-CA-Zertifikate in den Zertifikatspool einzuschleusen, auf diese Weise Denial-of-Service Angreife auf die Pfadkonstruktion für bestimmte Zertifikate durchführen kann.

Fehler	Beschreibung	Botan	Bouncy Castle	mbedTLS	OpenSSL	Apache	Firefox	KMail	OpenVPN	strongSwan
VCM_08	Akzeptanz von abgelaufenen Zwischen-CA- oder Nutzerzertifikaten							X		
VCM_11	Akzeptanz einer ungültigen Versionsnummer		X		X				X	
VEX_06	Akzeptanz eines Zwischen-CA-Zertifikats ohne Basic Constraints Erweiterung					X			X	
VEX_11	Akzeptanz von Zwischen-CA-Zertifikaten ohne keyCertSign Key Usage									X
VIP_04	Akzeptanz von Nutzerzertifikaten mit keyAgreement als einziger Key Usage									X
VCR_01	Ignorieren des Fehlens einer gültigen Sperrliste für jegliche Zertifikate			X						
VCR_15	Ignorieren des Fehlens einer gültigen Sperrliste für ein Zwischen-CA-Zertifikat									X
VCR_06	Akzeptanz von unbekannten kritischen Sperrlistenerweiterungen	X		X						
VCR_08	Akzeptanz von noch nicht gültigen Sperrlisten									X
VCR_13	Akzeptanz von Zertifikaten mit nicht-übereinstimmenden CRL-DP des Zertifikats und IDP der Sperrliste	X		X						X
RCM_14	Durchführung einer nicht-erschöpfenden Pfadsuche	X		X	X					
CCR_00	Probleme mit dem Sperrlisten-Cache							X		
CCM_01	Zu restriktive Behandlung der Pfadlängenbeschränkung	X								
CCM_13	Zu restriktive Behandlung der Pfadlängenbeschränkung bei selbstausgestellten (self-issued) Zertifikaten		X	X			X	X		X

Tabelle 1: Übersicht über die mit dem CPT aufgedeckten Schwachstellen und Problem in Bibliotheken und Anwendungen. Ein „X“ in einem Feld bedeutet, dass der betreffende Testgegenstand von dem Fehler bzw. Problem betroffen ist.

CCR_00: KMail hält Sperrlisten mit Hilfe von dirmngr [20] in einem Cache. Bei den CRL-Testfällen des CPT wird stets der gleiche Aussteller (IssuerDN) verwendet aber unterschiedliche Distribution Points. Dies konnte von dem Cache offenbar nicht korrekt gehandhabt werden, so dass es nicht möglich war, zuverlässige Testergebnisse für irgendwelche der Sperrlisten-Testfälle für KMail zu erreichen.

CCM_01: In Botan wurde die Pfadlängenbeschränkung, die in der Basic Constraints Zertifikatserweiterung spezifiziert werden kann, konsequent zu restriktiv ausgelegt. Somit wurden gültige Zertifikatspfade abgewiesen.

CCM_13: Eine ganze Reihe von Testgegenständen hat sog. selbstausgestellte (self-issued) Zertifikate bei der Bestimmung der Pfadlängenbeschränkung durch die Basic Constraints Zertifikatserweiterung nicht korrekt berücksichtigt und dadurch gültige Zertifikatspfade abgewiesen.

6. Fazit und Ausblick

In dieser Arbeit haben wir ein neues Werkzeug und eine Testspezifikation zum Testen der Zertifizierungspfadvalidierung in Bibliotheken und Anwendungen vorgestellt. Unser Testwerkzeug zeichnet sich durch nützliche Eigenschaften aus, die von keinem der bisher für diese Aufgabe existierenden Testwerkzeuge erfüllt werden. Mittels dieses Werkzeugs wurden mehrere Fehler in einer Reihe von weit verbreiteten Implementierungen der X.509 Pfadvalidierung identifiziert.

Um Entwickler:innen für zukünftige Implementierungen Hilfestellungen zu geben, sind die Ergebnisse in eine Technische Richtlinie [21] des BSI eingeflossen, die Empfehlungen zu X.509-Zertifikaten und Zertifizierungspfadvalidierung gibt. Dabei handelt es sich teilweise um Vorgaben für verschiedene Anwendungskontexte (z.B. TLS, IPsec, E-Mail) aus den entsprechenden Standards sowie um weitergehende Empfehlungen, die im Rahmen des Projekts entstanden sind. Die Technische Richtlinie wurde Ende 2020 vom BSI veröffentlicht.

Literaturhinweise

[1] ITU-T, „ITU-T: SERIES X: DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY / Directory, Information technology – Open Systems Interconnection – The Directory: Public-key and attribute certificate frameworks, X.509 (10/2016),“ 2016.

[2] D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley and W. Polk, „Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile“ (RFC 5280), 2008.

[3] OpenSSL, https://www.openssl.org/.

[4] Apache HTTP Client, https://hc.apache.org/.

[5] M. Georgiev, S. Iyengar, S. Jana, R. Anubhai, D. Boneh und V. Shmatikov, „The most dangerous code in the world: Validating SSL certificates in non-browser software,“ In proceedings of ACM CCS '12, pp. 38-49, 2012.

[6] x509test, https://github.com/yymax/x509test

[7] NIST - Computer Security Resource Center, „Path Validation Testing Program, https://csrc.nist.gov/projects/pki-testing.“.

[8] Certificate Fuzzer, http://cryptosource.de/certificate_fuzzer_de.html

[9] C. Brubaker, S. Jana, B. Ray, S. Khurshid und V. Shmatikov, „Using Frankencerts for Automated Adversarial Testing of Certificate Validation in SSL/TLS Implementations,“ http://www.cs.cornell.edu/~shmat/shmat_oak14.pdf.

[10] tlspretense, https://github.com/iSECPartners/tlspretense

[11] Go Fuzz, https://github.com/dvyukov/go-fuzz

[12] J. Schaad, B. Ramsdell, S. Turner, „Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 4.0 Message Specification“ (RFC 8551), April 2019

[13] D. Bleichenbacher, „Chosen Ciphertext Attacks Against Protocols Based on the RSA Encryption Standard PKCS #1,“ CRYPTO 1998.

[14] M. Marlinspike, „Null Prefix Attacks Against SSL/TLS Certificates,“ https://moxie.org/papers/null-prefix-attacks.pdf.

[15] Bundesamt für Sicherheit in der Informationstechnik „Certification Path Validation Test Tool“, http://www.bsi.bund.de/CPT

[16] Botan: Crypto and TLS for Modern C++. https://botan.randombit.net/

[17] D. Eastlake 3rd: „Transport Layer Security (TLS) Extensions: Extension Definitions“ (RFC 6066), January 2011

[18] Bundesamt für Sicherheit in der Informationstechnik, BSI TR-03124 www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr03124/TR-03124_node

[19] B. Korver, „The Internet IP Security PKI Profile of IKEv1/ISAKMP, IKEv2, and PKIX“ (RFC 4945), August 2007

[20] Klarälvdalens Datakonsult, g10 Code GmbH, „Using Dirmngr“, https://www.gnupg.org/documentation/manuals/dirmngr/

[21] Bundesamt für Sicherheit in der Informationstechnik, BSI TR-02103 https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr02103/TR-02103_node

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Virtualisation-based Security: Looking beyond strong isolation properties

Werner Haas[1]

Abstract:
Virtualisation technology allows replacing direct hardware access with a virtual machine environment. This provides for strong isolation of critical data and enables system monitoring from an elevated privilege level. The latter, however, is complicated because virtualisation is concerned with hardware properties such as CPU instructions or memory addresses. In the following we will present recent improvements in bridging this semantic gap for the Windows OS, thus opening the door for dynamic supervision algorithms. We also make a case for a security-oriented hypervisor architecture in order to limit the risk from highly privileged applications.

Keywords: memory forensics, security by design, semantic gap, Virtualisation, VMI

1. Introduction

The concept of virtualisation is known for roughly half a century and originated in the desire for increased hardware resource utilisation via time sharing. This implies robust isolation properties that are nowadays leveraged to improve system security via defence-in-depth strategies. In 2003, Virtual Machine Introspection extended the horizon for security features beyond basic partitioning as it described fully transparent monitoring of a running system by means of virtualisation technology. The next section briefly explains VMI‘s basic concept and its three core ingredients: isolation, inspection, and interposition.

It is followed by section 3 on digital forensics. Virtual Machines (VMs) are heavily used by analysts and defenders in various settings. Sandboxes allow testing of suspicious software in an isolated environment and VM state is easier to evaluate than a running system when searching for artefacts from malicious activities. However, VMI comes with its own share of baggage due to the inherent loss of semantic information resulting from the chosen abstraction level. The so-called semantic gap between hardware properties, such as CPU instructions or memory addresses, and the context of executed code must be bridged, for example via operating system (OS) profile data, in order to be able to draw meaningful conclusions.

In Section 4 we dive deeper into VMI-based analysis and describe our contributions for bridging the semantic gap for closed-source Windows systems. An important goal of our research is to reduce the amount of manual reverse engineering to a bare minimum and we show how this can be achieved with readily available debugging information. Furthermore we explain how working in a live system can improve the basic algorithms used for forensics and apply them to real-time system monitoring. From our point of

[1] Cyberus Technology GmbH

view, interposition is underappreciated although it can elegantly side-step a couple of hurdles.

While VMI offers unique advantages with respect to detection and protection capabilities, it also increases the value at risk in case of compromise. This is especially true when relying on virtualisation's strict isolation for protecting sensitive information. Section 5 is dedicated to this often overlooked aspect and we elaborate how the risk can be mitigated by means of a security-oriented hypervisor architecture. Capability-based access control provides the framework for fine-granular management of information flow which in turn allows minimising the trusted compute base.

We conclude in section 6 with a summary of our findings and present an outlook over our planned follow-up research.

2. Virtual machine introspection (VMI)

Information technology has become an integral part of our modern society and it is still growing in importance. Hence, like other precious goods, it requires adequate protection techniques and, as 100 percent security is unlikely, appropriate mechanisms to discover attacks. VMI first appeared in a paper on a novel architecture for Intrusion Detection Systems (IDS) [1]. An IDS residing on the same system it is designed to protect has obviously excellent insight into what is going on but risks getting compromised in case of successful attacks, hence defeating the purpose of intrusion detection. Network-based IDSs achieve greater attack resistance by relocating the detection mechanism to a different entity. The drawback, however, is significantly reduced visibility. VMI promises the best of both worlds by leveraging virtualisation technology to pull the IDS out of the system it is designed to monitor while keeping it on the same hardware host.

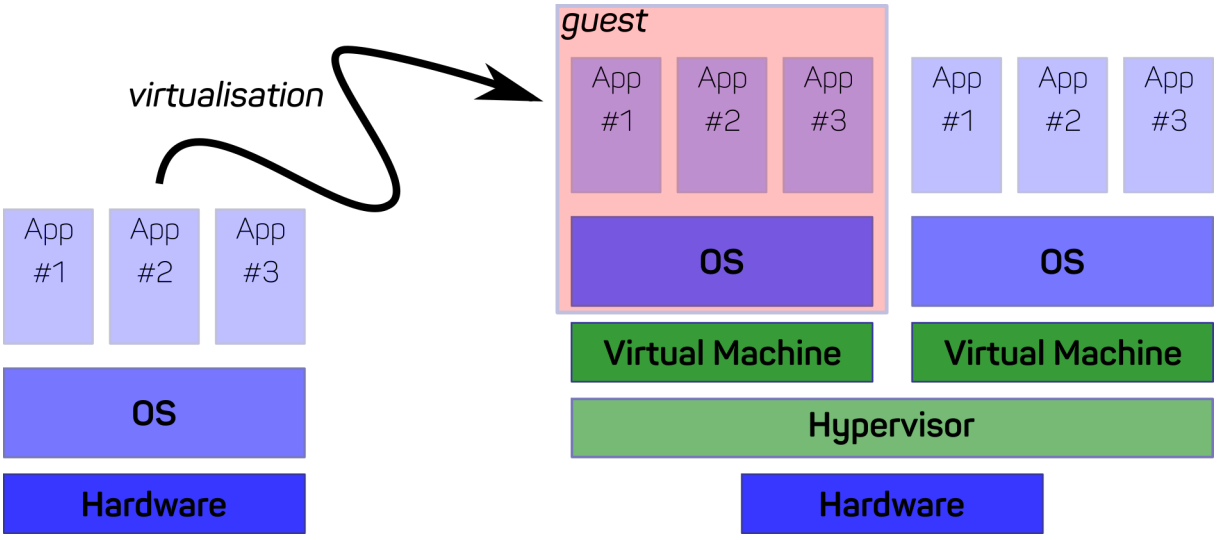


Figure 1: Traditional virtualisation use case. A hypervisor layer (green) creates virtual machines where the original software stack runs unmodified as guest on the host. Hardware utilization increases by running several guests in parallel.

Virtualisation [2] adds a software layer between the computer hardware and the software stack designed to run on a physical machine, thus creating a so-called virtual machine (VM) which is "an efficient, isolated duplicate of the real machine." [3]. The main driver behind virtualisation technology was, and still is, more efficient resource utilisation which in turn enabled today's cloud computing paradigm as a single *host* computer can provide several *guest* VMs. In the following, we will use the term hypervisor layer to denote the software creating a VM environment in general while its parts controlling VM operation will be called *Virtual Machine Monitor (VMM)*. *Hypervisor* refers to the software running at the highest privilege level[2]. VMI leverages three basic virtualisation properties:

1. Isolation
   Stringent isolation properties follow directly from the basic requirements for virtualisation as guest software should see a lifelike replica of physical hardware. This in turn requires that the hypervisor must be in complete control of all available resources. An attacker successfully compromising a guest VM thus has not automatically taken over the full system.
2. Inspection
   Being in full control of the hardware implies that the hypervisor can access all system state. This makes it very difficult for an attacker to hide from monitoring because all VM state is accessible from the hypervisor layer, an ideal starting point for intrusion detection.
3. Interposition
   In order to create an illusion of a physical system, the VMM has to interpose, that is to intercept and modify, certain guest actions, for example if code execution would violate VM isolation properties. This feature allows VMI to react to guest events in a dynamic manner.

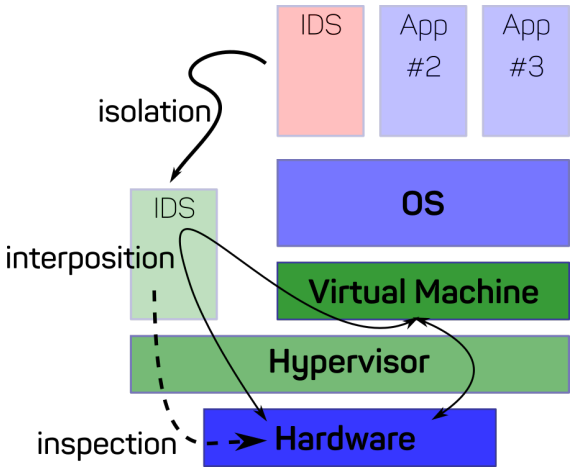


Figure 2: Intrusion detection via VMI. The IDS logic is removed from the original system but still has access to the same hardware resources. It can also be interposed on selected hardware accesses. The hypervisor does not have to multiplex multiple guests in this case i.e. most of the hardware devices can be passed through directly to the virtual machine.

[2] Note that hypervisor and VMM are often used synonymously. We prefer making a distinction in order to underline the fact that code executed in the hypervisor layer does not necessarily require maximum access rights.

To summarize: by means of virtualisation technology a full computer environment can be moved to a virtual machine and other programmes can be run in parallel on the same hardware. An IDS can be implemented on the same host while keeping it out of the realm of the original software stack. Thus, even if an attacker manages to fully control the VM, it does not automatically imply a compromised intrusion detection capability.

3. Digital forensics

Forensics typically describes the scientific techniques in connection with crime. While there are many different flavours of computer crime, we will focus on the detection of attacks on system integrity. VMs have become pretty much indispensable for defending against attacks e.g., during automatic testing of email attachments for malicious activities. Suspicious files are brought into a sandbox environment, first, and the corresponding VM can be quickly restored to a known good initial state after each test run without having to go through a full installation process.

In order to cover the tracks, fileless attacks are increasingly popular. Virtualisation is of great assistance in this case, too, as the hypervisor is in full control of the hardware, including memory. Hence, it can provide a full snapshot of the RAM content independent from the software running in a VM and any code to be executed by a processor has to be present in memory. However, this kind of computer analysis comes with its own set of problems. The hypervisor layer's task is to provide a duplicate of a real machine i.e., it is guest-software agnostic. [4] coined the name *semantic gap* for the delta between the bits and bytes in a raw memory image, and reading content from within an OS, where one can e.g. easily focus on the data belonging to a specific process.

Very basic detection mechanisms can safely ignore the semantic gap by scanning directly for known malicious patterns. YARA ([5]) is a tool commonly used for static analysis. A simple VMI-based IDS would simply ask the hypervisor for a memory snapshot, run a set of YARA rules, and report suspicious addresses. The obvious caveat is the static nature of the search. It requires malware analysts to develop specific detection rules, first, and cannot capture more complex indicators of compromise (IOC) like manipulation of OS-kernel data structures.

The latter can only be achieved with semantic information. As a simple example, the Windows task manager shows active applications on a system. Support processes running in the background, however, are often not listed and malicious actors have an interest in operating under the radar. If the specific OS version is known, one can check whether a program has removed itself from the list displayed by the task manager which could be an IOC.

Memory forensic tools bridge the semantic gap and allow incident responders and malware analysts to dissect memory images. The Volatility Framework ([6]) is probably best known toolkit and actively developed by the community. For example, last year an extension was made publicly available allowing researchers to extract SSH session keys from memory which render encrypted communication completely transparent to

network traffic monitors[3]. One possible usage scenario would run the malware in a VM, take a system snapshot once it is up and running, carve out the crypto configuration with Volatility, and then continue the analysis by monitoring the network traffic in clear text.

One caveat of Volatility is the knowledge about the observed system that is required for proper operation. An OSs' internal data structure, such as the layout of its process lists, can change from revision to revision, so an incorrect assumption can quickly lead to nonsensical results. Rekall ([7]) was forked from Volatility and it innovated the way different versions get handled[4]. First, instead of baking OS knowledge into the source code of the forensic tool, it leveraged debug information published by Microsoft in order to interpret memory data structures. This still requires the precise OS version but Rekall figures it out automagically via scanning the memory content.

Rekall had been used by the GRR incident response framework but in the end, relying on memory analysis, only, turned out to be too fragile. Memory acquisition from within the system to analyse may have blind spots in case of a successful compromise. Using alternative access methods to physical RAM is getting more and more cumbersome because of increasing sizes, tighter access control, and increased use of memory encryption. Nevertheless, memory forensics plays an important role and the next section will explain how VMI can sidestep some of the hurdles.

4. VMI-based analysis

4.1. Microsoft pdb data

The previous section introduced the semantic gap, which has to be bridged in order to get full value from VMI's inspection capability. Simply put, one has to locate the memory regions where relevant data like process information or crypto keys can be found. Rekall replaced manual reverse engineering with debug information publicly disclosed by Microsoft. This data (cf. [8] for further details) comprises for example the definition of Windows-internal datatypes, and relevant symbols, such as function names and commonly used kernel structures.

As alluded to earlier, internal data is highly version-specific, that is it depends on the concrete binary file. Since the debug information is not needed during regular operation, it is published in the form of so-called pdb (program database) files. All binaries hold a globally unique identifier (GUID) which can be used to fetch the corresponding pdb data from Microsoft's public symbol server. The GUID itself is part of the metadata of any code loaded by the OS. Specifically, it is a field in the PE header which is the format of all Windows executables. By searching the PE header in memory and fetching the pdb file corresponding to its GUID field we get the required semantic information for interpreting raw memory data.

---

[3] The SSHKeys plugin won 2nd place in the 2020 Volatility Plugin Contest. Further information can be found at https://blog.fox-it.com/2020/11/11/decrypting-openssh-sessions-for-fun-and-profit/

[4] Note that Rekall is no longer actively maintained but features like OS auto-detection can be found in the upcoming Volatility 3 release

4.2. Windows kernel address space

The downside of system hardening mechanisms is that they also affect legitimate use cases. For a long time, code like the OS kernel was loaded at a static address. Since it got more and more difficult for attackers to make their own code executable with highest system privileges, they reverted to stitching together existing code snippets in an unplanned manner[5]. This requires knowledge of the addresses of useful code which in turn was countered by so-called address space layout randomisation (ASLR). It is supposed to turn finding usable code into searching the needle in a haystack[6].

All modern OSs randomise the locations where code is loaded into memory and the kernel base address, that used to be directly accessible, is now hidden. Memory forensic tools like Volatility apply heuristics in order to find the location of kernel structures but with VMI in a live system we can evaluate the processor state directly for this purpose. Regular applications have to interact with the OS, after all, hence there has to be a well-defined interface between user- and kernel-space. It consists of a pointer to the so-called _KPCR structure which is stored in a processor register. Being in full control of the hardware implies that inspection from the hypervisor layer is not limited to memory and also includes CPU state. Hence, the VMI engine need not guess the _KPCR[7] address.

The _KPCR structure itself is part of the Windows kernel and its layout is defined in the corresponding pdb file. Since the kernel is loaded under hypervisor supervision when starting a VM, one can parse the corresponding header information during the process and tell VMI the kernel GUID. The pdb file, however, only provides us with relative offsets within the corresponding executable (= the kernel to get things started). Hence, we still need to figure out the kernel base address.

An integral part of the processor state is information about how to handle irregular events. Such events can be triggered externally, for example when it is time to switch to another task, or by code execution itself, for example when accidentally attempting to divide by 0. The generic term is interrupt and _KPCR holds the descriptor table (IDT) which holds the address of the corresponding handler function, that is how the CPU should proceed in case such an event occurs. Division by 0 is handled by the OS in the #DE (division error) handler function whose absolute address can be found in the IDT.

The #DE function symbol, however, is defined in the kernel's pdb which we know from starting the VM. By subtracting the relative offset of #DE function (from the pdb file) from the absolute value of the #DE handler address in the IDT (from the gs register and _KPCR knowledge), VMI can calculate kernel_base directly i.e., without reverting to heuristics. Because all further code is loaded under kernel supervision, base addresses of subsequent modules can be determined by parsing kernel data structures which can be located via kernel_base.

---

5 Return Oriented Programming (ROP) manipulate function return addresses. Since the CPU does not check whether such addresses correspond to function calls, a cleverly crafted return-address-stack can implement arbitrary algorithms.

6 ASLR raises the bar for attackers but it should be seen more like a stop-gap technique while we are waiting for more effective protection primitives. There are simply too many holes to resolve the randomisation.

7 _KPCR refers to the Processor Control Region which can be accessed via the gs register in x64 Windows

1. Loading of guest kernel → GUID
2. Fetching of corresponding pdb information
3. Runtime hardware inspection → _KPCR address
4. _KPCR layout (from #2) + _KPCR address (from #3) = IDT location
5. #DE address from corresponding IDT entry (from #4)
6. #DE address (from #5) = kernel_base + #DE offset (from #2)
   → kernel_base

Figure 3: Algorithm for closing the semantic gap for memory forensic analysis via VMI.

4.3. Memory access

Symbol information, that is the mapping between memory addresses and their meaning, is not sufficient, though. Forensics tools operating with memory images face the challenge of mapping the physical address space of the data to the virtual address space[8] that is used during code execution. Furthermore, an OS might have offloaded (=swapped) unused data to disk. Thus, offline processing has to stitch together the relevant memory regions before it can start the analysis.

VMI in a live system reduces the complexity significantly. First, inspection can operate directly with the virtual addresses used by the code running in the VM as the address translation information is readily available. This eliminates the task of rearranging regions in physical memory according to their logical order. Second, interposition can be used to fake program behaviour. Swapping data to disk is a performance optimisation, only, that is accessing the corresponding addresses is perfectly legitimate. In case of a memory dump, Volatility has to fetch the data from the pagefile. A VMI-based engine can simply pretend the currently executing code wanted to read a value by injecting a (bogus) page fault[9] and let the OS itself restore the desired data.

While VMI provides direct access to a guest‘s physical memory, virtual memory regions might actually be paged out to disk. Memory forensics tools thus combine memory dumps with pagefile content in order to reconstruct the relevant address regions. In a live system, however, one can simply inject a (bogus) page fault and let the OS itself restore the desired data, thus simplifying the access procedure significantly.

4.4. Windows kernel assists

Taking this idea one step further, the VMI engine can also inject entire system calls. One straight forward application with respect to forensic analysis is the de-obfuscation of kernel-internal data structures. For example, Microsoft has recently changed the meaning of the TypeIndex value of Windows objects. Instead of being an index into an array that is constant across all processes, it is now combined with run-time data. The transfer function is irrelevant for normal operation, hence it can be changed by Microsoft at-will[10].

---

8 Note: this virtual address is independent from virtualisation with a hypervisor! It provides the mechanisms for memory access control and is used also when operating on hardware directly.

9 Memory is managed by the CPU/OS in fixed-size chunks called pages. Hence the name page fault when such a chunk is not accessible and pagefile for the storage location on disk.

10 The curious reader can find a detailed discussion at https://medium.com/@ashabdalhalim/a-light-on-windows-10s-object-header-typeindex-value-e8f907e7073a

Instead of reverse engineering this function (and checking/repeating the procedure for every new Windows version), a VMI algorithm can pretend the application wanted to fetch the corresponding information by faking the corresponding system call. So instead of looking up the array content in memory directly (which requires deobfuscation of the index), it can simply query the OS.

5. Endpoint security software

5.1. Attack surface considerations

So far we have looked at virtualisation and how VMI can be used to simplify memory forensics. Instead of reacting to a known incident, the methods can also be applied to detect an attack in its early stages. This would turn an analysis instrument into a protection mechanism. A quick look-up of „antivirus“ in the CVE database returns almost 600 matches, though. It is obvious that the software meant to increase system security represents a considerable security risk by itself.

One of the value propositions of VMI was the isolation property which moved the engine out of reach of the VM. Of course, this is perspective is too narrow as it requires the presence of a hypervisor layer, which can also be attacked. It is important to note, however, that IDS functionality in the host itself relies on the OS for its integrity protection which comprises millions of lines of code as the OS is responsible for managing the entire hardware.

In general, a hypervisor would have to provide somewhat similar functionality as it has to create the VM environment. But if one is interested in VMI of a single VM, only, this task can be simplified significantly. Instead of multiplexing hardware access between different guests, most devices can be assigned directly to the one guest VM where the OS controls their operation as before. Thus the hypervisor layer basically has to manage memory, only. The code for this task is orders of magnitude smaller, compared to legacy OSs like Linux or Windows.

5.2. Hypervisor architecture

Another vector to improve resilience is the architecture of the virtualisation solution itself. Commonly used virtualisation platforms are similar in design to monolithic operating systems. Intel Architecture, however, provides the same different privilege levels in VMX-root mode as for guests running in a VM. Thus, design principles from the microkernel world can be applied to the hypervisor domain.

In case of a monolithic system, management mechanisms and the corresponding policies are implemented at the same privilege level. In our system, the highest privileged code implements mechanisms, only, and policy decisions are deferred to less privileged code running basically as user-space process in the virtualisation layer. This raises the bar for an attacker because even if she found a vulnerability, for example in the VMI memory analysis, the effects would not automatically compromise the memory isolation of different components in the hypervisor layer.

Furthermore, by using capabilities for access control it is possible to implement fine-grained information flow management. If one was only interested in periodic YARA

scans, the VMI code could be granted read-only access to the VM's memory, only. Neither a bug in the YARA engine, nor a compromised YARA rule update[11] would allow an attacker to modify its VM's memory content, effectively preventing spreading of an infection via the security tool.

5.3. Virtualisation-based security

We envision a system where virtualisation is a first class citizen with respect to security considerations i.e., it is not reduced to its strong isolation guarantees. We have developed a lightweight base infrastructure based on the Principle Of Least Privilege (POLP) where our Hedron hypervisor ([9]) requires on the order of 10.000 lines of code, making its implementation amenable to manual code review.

Runtime policies are implemented as less privileged processes. This avoids costly VM switches as would be required in off-the-shelf virtualisation solutions where the VMI engine is isolated via a VM boundary. Isolation and protection is controlled via capabilities similar to the seL4 microkernel ([10]) which calls itself "the world's most highly assured OS kernel."

By following a pass-through philosophy and leveraging latest hardware support we both reduce the implementation complexity and make the performance impact of virtualisation barely noticeable. The VMI overhead is essentially a function of the desired IDS capabilities. Even in the simple case of memory scanning we achieve significant speed-up because we can focus on the relevant memory regions in a live system instead of having to process an image of the entire memory space. Furthermore, we extended the scope of interposition found in classic virtualisation solutions, allowing us to perform event-driven, on-demand analysis.

6. Summary

Virtualisation technology enables the operation of multiple VMs on a single piece of hardware. Additionally, it opens the door for VMI, which, in its simplest form, enables advanced digital forensics by means of analysing static system snapshots. VMI allows removing the IDS from a monitored system without having to move it to a different physical machine, as in case of network-based intrusion detection.

However, there is still a semantic gap between the data available inside an OS environment and the data seen on hypervisor layer. By leveraging readily available debug information we have developed an elegant way to apply standard forensic analysis to a live system. An additional reduction of implementation complexity is achieved by making use of interposition, giving us access to OS functionality in a fully transparent way. This opens the door to dynamic VMI which can be leveraged to improve forensic algorithms by running detection algorithms on-demand.

Finally, we have explained how designing the hypervisor architecture with security in mind can mitigate potential risks stemming from an increased code base. We follow the principle of least privilege by using a microkernel architecture for the virtualisation

[11] cf. 2020's SolarWinds attack

layer, together with fine-grained access control via capabilities. This serves as example of how appropriate architecture design of the underlying virtualisation platform can address concerns regarding standard security applications and opens the door for innovative software solutions.

So far our VMI work has been at the user/system interface which makes it relatively easy to define intercept points for transitioning into the hypervisor domain. Moving forward we expect to extend our analysis capabilities further into user space as Windows performs a lot of data processing in system DLLs before actually entering the kernel. Another task is implementing nested virtualisation which will allow us to run Windows 10 with Hyper-V enabled inside our VM.

Literaturhinweise

[1] Garfinkel, Tal, and Mendel Rosenblum. "A virtual machine introspection based architecture for intrusion detection." Ndss. Vol. 3. No. 2003. 2003.

[2] https://www.ibm.com/cloud/learn/virtualization-a-complete-guide

[3] Popek, Gerald J., and Robert P. Goldberg. "Formal requirements for virtualizable third generation architectures." Communications of the ACM 17.7 (1974): 412-421.

[4] Chen, Peter M., and Brian D. Noble. "When virtual is better than real [operating system relocation to virtual machines]." Proceedings eighth workshop on hot topics in operating systems. IEEE, 2001.

[5] https://virustotal.github.io/yara/

[6] https://www.volatilityfoundation.org/

[7] http://www.rekall-forensic.com/

[8] https://www.wintellect.com/pdb-files-what-every-developer-must-know/

[9] https://www.cyberus-technology.de/posts/2020-11-13-hedron-hypervisor.html

[10] https://sel4.systems/

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Mikrohypervisor als Basis für eine hochsichere Cloud

Steffen Liebergeld[1], Dr.-Ing. Michael Hohmuth[1], Dr.-Ing. Adam Lackorzynski[1]

Kurzfassung:
Mikrokerne haben sich als Basis von sicherheitskritischen Anwendungen insbesondere für Aufgaben mit erhöhtem Schutzniveau bis GEHEIM bewährt. Die Verheißungen der Cloud sind auch für Behördenanwendungen von Interesse. Die dazu nötigen Anforderungen, insbesondere für eine Evaluierung können etablierte Hypervisorlösungen architekturbedingt nicht bieten. In diesem Beitrag betrachten wir wie ein mikrokernbasierter Hypervisor die Basis für eine hochsichere Cloud bilden kann.

Stichworte: Behördencloud, Cloud Computing, Common Criteria, Hypervisor, Mikrohypervisor, Mikrokern, Security by Design, Virtualisierung, Virtuelle Maschine, VM, Zertifizierung, Zulassung

1. Einleitung

In der heutigen Unternehmenspraxis haben sich Cloudlösungen weitgehend durchgesetzt. Zentrale Vorteile sind dabei zum einen Kostenersparnisse, die durch die zentralisierte Administration und durch das Zusammenlegen von Rechenzentren entstehen. Für die Nutzer ist von Vorteil, dass man agil auf Lastsituationen reagieren kann. Des Weiteren kann eine Cloudlösung durch das verwendete Datacenter eine höhere Sicherheit bieten, als das einzelne Unternehmen für sich finanzieren könnten. Ein weiterer Gesichtspunkt ist die Ausfallsicherheit, die schon allein dadurch erhöht werden kann, dass man Redundanz schafft, indem man Workloads auf mehrere Datacenter verteilt.

Diese Vorteile machen die Cloud auch für Behörden interessant. Gerade mit der Corona-Pandemie haben wir erlebt, wie auch größere Behörden in kürzester Zeit ihre Anwendungen online verfügbar machen mussten, damit ihre Mitarbeiter im Home-Office arbeiten können. Die schnelle Verfügbarkeit von sicheren Rechenressourcen kann gerade in einem solchen Fall die Situation retten.

Allerdings spielen im Bereich Behörden-IT – insbesondere in Bezug auf Daten unter besonderem Schutz – die Zulassung und Zertifizierung der verwendeten IT-Produkte eine sehr viel größere Rolle als in der Privatwirtschaft.

Die zentrale Motivation von Cloudlösungen ist die gleichzeitige Nutzung von Systemressourcen durch mehrere Nutzer zur gleichen Zeit. Je mehr Ressourcen geteilt werden, desto besser ist die Systemauslastung und desto geringer sind die Kosten für den Einzelnen.

Eine zentrale Komponente von Cloudlösungen stellt der Hypervisor dar. Aufgabe des Hypervisors ist es, virtuelle Maschinen (VM), also virtuelle Abbilder von physischen Systemen auszuführen. In diesen virtuellen Maschinen laufen die Anwendungen der Kunden. Auf den Servermaschinen werden mehrere virtuelle Maschinen gleichzeitig ausgeführt, die sich die vorhandenen Ressourcen teilen. Einzelne Kunden haben typi-

[1] Kernkonzept GmbH

scherweise keine Kontrolle darüber, auf welchen Servermaschinen ihre virtuellen Maschinen zur Ausführung kommen. Auch haben die Kunden kein Wissen über andere Cloudkunden und können daher den Anwendungen in anderen virtuellen Maschinen nicht trauen. Das bedeutet, dass das Vertrauen der Kunden in die Sicherheit der Cloudlösung inhärent davon abhängt, wie gut die Cloud virtuelle Maschinen voneinander isolieren kann. Wir möchten uns daher in diesem Beitrag explizit darauf konzentrieren, wie Hypervisoren Isolation umsetzen und wie vertrauenswürdig die Isolation ist.

Etablierte Hypervisor-Lösungen etwa von VMware (VMware ESX), Microsoft (HyperV), Citrix (Xen) oder der Linux Foundation (KVM) benutzen monolithische Architekturen, was bedeutet, dass eine große Menge an essenzieller Funktionalität, d.h. Code, in einer Softwarekomponente vereint sind. Schätzt man die Sicherheit einer Software ein, so ist die Menge an Code, der man vertrauen muss, eine brauchbare Metrik. Wir nennen diese Menge die Trusted Computing Base (TCB) der Anwendung. Der Satz „Code enthält Fehler“ ist eine der Grundannahmen der heutigen IT. Wenn das gilt, so kann man ableiten: „Viel Code enthält viele Fehler“. Untersuchungen haben eine durchschnittliche Fehlermenge von zwei Fehlern pro 1000 Zeilen Code gefunden [2]. Fehler in der TCB können die Sicherheit der Anwendung kompromittieren. Es ist daher erstrebenswert, die TCB zu minimieren.

In monolithischen Softwarearchitekturen existiert durch die Konzentration von essenziellen Funktionen in nur einer Softwarekomponente keine Grenze, die die Auswirkungen eines Fehlers eindämmen kann. Das bedeutet, dass man zur Zulassung oder Zertifizierung eines solchen monolithischen Hypervisors möglicherweise die Gesamtheit seines Codes evaluieren müsste. Bei Codeumfängen von mehreren Millionen Zeilen ist dies praktisch unmöglich.

Für nachweislich sichere Systeme ist darum ein anderer architektureller Ansatz nötig. Das System muss in kleine Module aufgeteilt werden, die klar definierte Schnittstellen bieten und einzeln auditiert und verifiziert werden können. Die kleinen Module müssen nachweislich sicher isoliert werden, so dass sie sich nicht ungewollt gegenseitig beeinflussen können. Dies erlaubt es zudem, nur genau die Softwarefunktionen einzusetzen, die zwingend nötig sind, um eine Funktion umzusetzen, und damit die TCB der Anwendung minimal zu halten.

Mikrokernsysteme sind genau nach dieser Architektur aufgebaut. Hier wird eine Systemtrennung durch den Mikrokern durchgesetzt. Der Mikrokern selbst besteht aus nur wenigen tausend Zeilen Code, eine Menge, die überschaubar ist und sich für Evaluierungen eignet; es besteht sogar die Möglichkeit, die Qualität durch Anwendung formaler Methoden abzusichern. Andere Komponenten des Systems laufen voneinander getrennt in eigenen Adressräumen. Mit einem solchen Systemsetup ist es möglich, die Abhängigkeiten von Systemteilen voneinander zu entkoppeln. Beispielsweise ist eine Anwendung, die keine Netzwerkfunktionalität verwendet, auch nicht von einem Netzwerktreiber oder gar einem Netzwerkstack abhängig.

Auf Mikrokernen basierte Systeme kommen bereits heute in Systemen zum Einsatz, welche bis zur Geheimhaltungsstufe GEHEIM freigegeben wurden. Es bietet sich also

an zu untersuchen, ob es möglich ist, einen Cloud-Hypervisor auf Basis eines Mikrokerns zu gestalten und damit einen Weg in Richtung eines Hypervisors freizumachen, welcher zulassungsfähig und zertifizierbar ist.

2. Stand der Technik

In diesem Kapitel möchten wir zuerst die Softwarearchitektur einer typischen Cloudlösung vorstellen. Wir werden danach eine zentrale Komponente der Cloud, den Hypervisor, vorstellen und die Softwarearchitektur typischer Lösungen skizzieren und analysieren.

Danach führen wir die Mikrokernarchitektur ein, welche als Grundlage für das Folgekapitel dienen soll. Als Beispiel dient uns der Open-Source-Mikrokern L4Re, der bereits in verschiedenen vom BSI bis GEHEIM zugelassenen Anwendungen verwendet wird [3].

2.1. Technologie der Cloud

In der heutigen IT-Landschaft ist die Cloud einer der Basispfeiler. Etablierte Cloudanbieter wie Google, Amazon, IBM und Microsoft bieten die Dienste ihrer Clouds für Privatkunden und für Unternehmen an. Analysten prognostizieren für den so aufgespannten Markt einen Wert von 761 Milliarden Dollar in 2027 [1].

Eine Cloudlösung besteht aus einer Kombination von Software und Hardware. Ziel ist es, vorhandene Ressourcen in geeigneter Weise anzubieten. Eine Variante sind hier virtuelle Maschinen (VMs). Diese virtuellen Maschinen bilden ein standardisiertes Maschinenmodell ab und sind damit unabhängig von der tatsächlich verwendeten Hardware. Sie können dynamisch mit verschiedenen CPU- und Speicherausstattungen konfiguriert werden.

Als Hardwareausstattung sind typischerweise mehrere Servermaschinen unterschiedlicher Hardwaregenerationen mit üppigen Speicher- und CPU-Ausstattungen vertreten. Im laufenden Betrieb wird die Hardwareausstattung dabei stetig aktualisiert und erweitert.

Zum Softwarestack gehören unter anderem VM-Management, software-defined Networking, Authentifizierungs-Dienste und -Appliances sowie ein Storagelayer für persistenten Speicher.

Eine wesentliche Komponente ist der Hypervisor, welcher als Betriebssystemsoftware auf den Servermaschinen läuft und die physischen Ressourcen verwaltet und als virtuelle Maschinen bereitstellt.

Alle bisher beschriebenen Komponenten werden von einer Cloud-Managementsoftware zentral verwaltet und über deren Nutzerschnittstelle administriert.

Dem Kunden werden dabei Ressourcen wie Rechenzeit und Speicher in verschiedenen Formen verkauft. Gängige Modelle sind IAAS, PAAS und SAAS.[2] All diesen Lösungen

[2] Infrastructure as a Service, Platform as a Service, Software as a Service

liegen virtuelle Maschinen zugrunde, welche dynamisch durch den Hypervisor erzeugt, gesteuert und konfiguriert werden.

In diesem Beitrag wollen wir uns auf den Hypervisor beschränken. Der Hypervisor ist in seiner Bedeutung für eine Cloud so zentral, dass eine hochsichere Cloud ohne einen hochsicheren Hypervisor nicht möglich ist.

Im Folgenden wollen wir analysieren, welche Funktionalität ein Hypervisor mitbringen muss. Danach werden wir KVM als ein wichtiges Beispiel eines gängigen Hypervisors analysieren und hinsichtlich seiner TCB bewerten.

2.2. Der Hypervisor – zentraler Bestandteil jeder Cloud

Hauptaufgabe des Hypervisors ist es, virtuelle Maschinen auszuführen. Die dazu nötige Technologie nennt man Virtualisierung. Wir werden also zunächst Virtualisierung etwas näher erläutern und danach die Anforderungen für Hypervisoren analysieren.

2.2.1. Virtualisierung als Voraussetzung für die Cloud

Die Fähigkeit, virtuelle Maschinen auszuführen, erfordert Virtualisierung, d.h. die Bereitstellung einer virtuellen Kopie eines physischen Computers. Dazu müssen CPU, Arbeitsspeicher und Geräte virtualisiert werden. CPU und Speicher werden in modernen CPUs hardwareunterstützt virtualisiert. Mehr Designspielraum gibt es jedoch bei der Gerätevirtualisierung:

Beim **Gerätedurchgriff** wird ein physisch vorhandenes Gerät dem Gastsystem direkt zur Verfügung gestellt und von dessen Treibern allein gesteuert. Dies bringt die beste Performanz, da es keine Indirektionen gibt, ermöglicht aber im Regelfall keine gleichzeitige Nutzung eines Geräts durch mehrere VMs.

Um ein solches Sharing zu erreichen, muss der Hypervisor den VMs virtuelle Geräte bereitstellen und Zugriffe darauf auf gemeinsam genutzte physische Geräte abbilden. Virtuelle Geräte können entweder ein weit verbreitetes Gerätemodell **emulieren**, sodass Gastsysteme diese Geräte mit ihren existierenden Treibern verwenden können, oder eine speziell für den Hypervisor entworfene Geräte-API implementieren; dann spricht man von **Paravirtualisierung**. Letztere vermeidet den hohen Implementierungsaufwand, die hohe Komplexität und die eingeschränkte Performanz einer Emulation, benötigt aber spezielle Treiber im Gastsystem.

Virtio ist ein solches paravirtualisiertes Interface, das bereits die wichtigsten Gerätearten unterstützt. Der Vorteil von virtio ist, dass es eine ausgereifte Technologie ist, welche von vielen Hypervisoren wie z.B. KVM bereits unterstützt wird. Auch sind alle typischerweise benötigten virtio-Treiber bereits in Linux enthalten und für andere Systeme einfach integrierbar.

Insgesamt lässt sich feststellen, dass Virtualisierung ein komplexes Thema darstellt und eine vollständige Implementierung damit aufwendig ist und viel Entwicklungsarbeit erfordert.

2.2.2. Funktionsweise und Sicherheitseigenschaften des Hypervisors

Der Hypervisor ist eine komplexe Software. Wir werden daher die vielfältigen Aufgaben eines Hypervisors strukturiert erklären.

Konzeptuell sind gängige Hypervisoren vollständige Betriebssysteme. Sie enthalten Gerätetreiber zur Ansteuerung von Geräten wie Netzwerkkarten, Festplatten und Grafikeinheiten. Sie ermöglichen die Ausführung von Anwendungen und bringen dafür sowohl reichhaltige Programmierschnittstellen als auch Infrastruktur wie Dateisysteme und Netzwerkstacks mit.

Weiterhin implementieren Hypervisoren die Fähigkeit, virtuelle Maschinen zu konfigurieren, zu steuern und zu kontrollieren. Konfiguration bedeutet die Orchestrierung der Ausstattung der VMs mit Arbeitsspeicher, CPUs und Geräten. Sollen Netzwerkkarten eingebunden werden, so gehört zur Konfiguration zum Beispiel auch das Zuweisen einer IP-Adresse und die Verbindung mit dem Netzwerk. Zur Steuerung und Kontrolle von VMs gehört das Erzeugen und Zerstören der VMs genauso wie deren Starten und Beenden.

In den virtuellen Maschinen werden Betriebssysteme und Anwendungen der Kunden, also zum Beispiel von verschiedenen Unternehmen, ausgeführt. Der Cloudbetreiber hat typischerweise keinen Einfluss auf diese Gastsysteme. Darum ist für eine hochsichere Cloud eine verlässliche und vertrauenswürdige Trennung zwischen den virtuellen Maschinen wichtig.

Ein Hypervisor muss durch eine Cloudmanagementsoftware fernsteuerbar sein (Hypervisor-API). Typischerweise wird dies in speziellen Anwendungen implementiert, welche die Befehle der Cloudmanagementsoftware für den Hypervisor übersetzen und mit dessen Mechanismen ausführen.

Ein Beispiel für eine Hypervisor-API ist Libvirt. Dieses Open-Source-Projekt unterstützt mehrere Hypervisorlösungen, z.B. Xen, KVM und NetBSD. Das VM-Managementsystem OpenStack setzt Libvirt als eine Hypervisor-API ein, und es existieren ausgereifte Tools für Anwender und Administratoren.

2.2.2.1. Analyse von KVM als Beispiel eines Hypervisors

KVM ist ein Linux-basierter Hypervisor, der seit vielen Jahren entwickelt und in vielen Bereichen erfolgreich eingesetzt wird [4]. Da er unter einer Open-Source-Lizenz lizenziert ist, eignet er sich hervorragend für unsere Analyse.

KVM ist in den Linux-Kern integriert und hat damit Zugriff auf die Betriebssystemfunktionalitäten von Linux, wie zum Beispiel Gerätetreiber, Dateisysteme und Netzwerkstacks.

KVM selbst ist als ein Kernelmodul für den Linux-Kern implementiert. Dieses Kernelmodul implementiert dabei die CPU und Speichervirtualisierung für die VMs. Zusätzlich benutzt KVM ein Linux-Nutzerprogramm für die Steuerung und Kontrolle der VM sowie für das Bereitstellen von Geräten. Wir nennen dieses Nutzerprogramm Virtual Machine Monitor (VMM). Für KVM gibt es mit Qemu und Firecracker mittlerweile

mehrere VMMs. Qemu ist dabei der Allrounder, der sowohl viele Gerätemodelle emulieren kann als auch Virtio als Paravirtualisierungsschnittstelle unterstützt. Qemu kann damit eine Vielzahl von Gast-Betriebssystemen, wie z.B. Windows, ausführen. Alternativ gibt es Firecracker, welcher sich auf Linux als Gastsystem spezialisiert und ausschließlich Virtio-Geräte unterstützt.

KVM wird direkt von der Hypervisor-API Libvirt unterstützt. Libvirt wird dabei in einer Linux-Anwendung mit vollem Zugriff auf die POSIX-API implementiert.

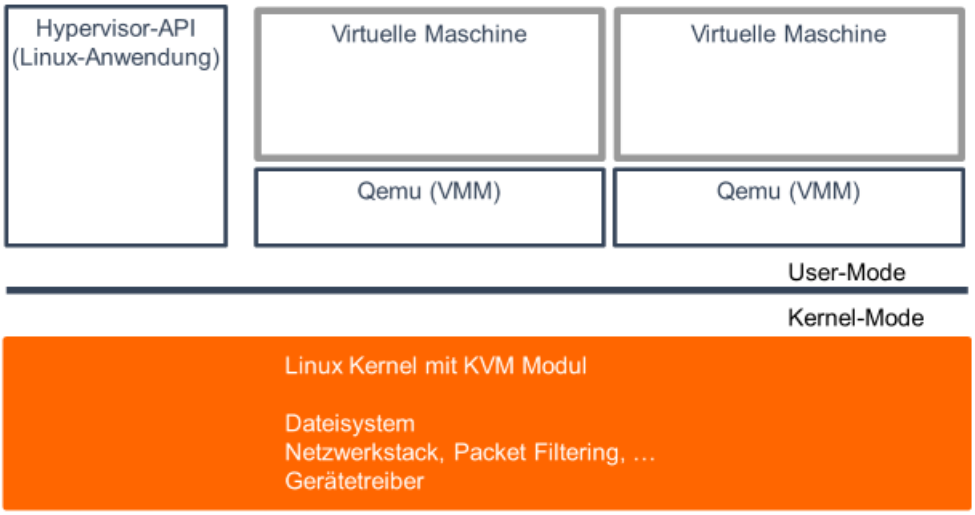


Abb. 1: Architekturdiagramm des Hypervisors KVM. Orange hinterlegte Komponenten sind kritisch für die Isolationseigenschaften.

2.2.2.2. Vertrauenswürdigkeit der Isolationseigenschaften von KVM

Code, der im höchstprivilegierten Modus der CPU – dem Kernel-Modus – ausgeführt wird, ist stets kritisch, denn es existieren in diesem CPU-Modus keine durch Hardware durchgesetzten Grenzen. Mit anderen Worten, ein einzelner Fehler kann schlimmstenfalls dazu führen, dass die Trennung untergraben wird und das System nicht mehr funktionsfähig bzw. vertrauenswürdig ist.

Linux und damit auch KVM ist ein typischer Vertreter monolithischer Betriebssysteme. So läuft bei KVM neben dem Hypervisor KVM auch der gesamte Linux-Kern im Kernel-Modus. Dazu gehören auch Funktionalitäten, die für die Virtualisierung nicht zwingend benötigt werden, wie z.B. Netzwerkstacks und eine Vielzahl von Gerätetreibern und Dateisystemen. So kommt der Hypervisor KVM auf eine TCB im Umfang von mehreren Millionen Zeilen Code.

Diese große Menge an Code müsste in einer Zulassung auf etwaige Fehler untersucht, dokumentiert und evaluiert werden. Eine solche Untersuchung ist jedoch mit heutigen Mitteln in einem vernünftigen Zeitrahmen nicht durchführbar, und damit fallen KVM und alle Hypervisorlösungen mit monolithischer Architektur als Basis für eine Cloud mit sehr hohen Sicherheitsanforderungen aus.

2.3. Mikrokernarchitektur

In einem Mikrokernsystem wird die Menge an Code, die höchstprivilegiert läuft, per Design minimal gehalten. Der Mikrokern implementiert dabei nur solche Funktionalität, welche nicht im unprivilegierten CPU-Modus als Anwendung implementiert werden kann. Im Wesentlichen sind das:

- IPC: Kommunikation zwischen Anwendungen.
- Speicherkonfiguration: Mechanismen zur sicheren Manipulation der Speicherkonfiguration und Durchsetzung von Isolation.
- Scheduling: Zuweisung von Rechenzeit an Anwendungen.

Damit kann die Menge an Code, die mit höchsten Privilegien im Kernel-Modus läuft, sehr klein gehalten werden. Der L4Re-Mikrokern hat in einer typischen Konfiguration ca. 30.000 Zeilen Code.

Alle weiteren Betriebssystemfunktionalitäten werden als Anwendung implementiert. Damit können komplexe Komponenten wie Gerätetreiber, Dateisysteme und Netzwerkstacks voneinander getrennt ausgeführt werden. Die Speichertrennung zwischen den Komponenten wird dabei in Hardware durch die Memory Management Unit (MMU) der CPU durchgesetzt. Das hat zur Folge, dass ein Fehler in einer solchen Komponente sich nicht direkt auf andere Komponenten auswirken kann.

Das bedeutet auch, dass bei einer Untersuchung der TCB einer Anwendung – neben dem Mikrokern – nur solche Komponenten betrachtet werden müssen, deren Funktionalität von dieser Anwendung direkt oder indirekt benutzt wird.

Damit erreicht man eine anpassbare und überschaubare TCB, welche mit heutigen Mitteln in überschaubarer Zeit evaluierbar ist und damit zulassungsfähig sein kann.

2.4. Mikrokern und virtuelle Maschinen

Im L4Re-System wurde die Unterstützung von virtuellen Maschinen ähnlich wie in KVM in zwei Teile aufgespalten: Mikrohypervisor und VMM.

Der Mikrokern übernimmt dabei den Part des Hypervisors und wird somit zum Mikrohypervisor. Dazu implementiert er zwei Mechanismen: zum einen einen Systemaufruf zum Wechseln vom Host- in den Gastmodus; zum anderen Funktionalität, um den Gastspeicher zu konfigurieren. Diese Funktionalitäten sind sicherheitskritisch und müssen architekturell zwangsweise im Kernel-Mode implementiert werden. Ähnlich der Behandlung von Ausnahmen (Exceptions und Faults) von Mikrokernanwendungen reflektiert der Mikrokern Virtualisierungsereignisse der VM zurück an eine Nutzeranwendung.

Die Steuerung der virtuellen Maschine wird komplett von einer Nutzeranwendung, dem VMM, übernommen. Eine Implementierung eines solchen VMM ist die L4Re-Anwendung UVMM.

Beim Design von UVMM haben wir – ähnlich wie beim Mikrokern – das Minimalitätsprinzip angewandt. Hauptziele dabei sind Performanz und minimale Komplexität. Da

die Emulation von Hardwaregeräten in jedweder Form komplex und damit rechenaufwendig und fehleranfällig ist, wollten wir soweit wie möglich darauf verzichten.

UVMM unterstützt zunächst Linux als Gastsystem. Mit dieser Einschränkung stand es uns offen, in Linux bereits vorhandene virtio-Schnittstellen zu verwenden und damit auf Emulation von Gerätemodellen weitgehend zu verzichten. Da wir das Bootprotokoll von Linux genau kennen, konnten wir auch auf eine Emulation eines BIOS und Unterstützung von Bootloadern verzichten. Eine Instanz von UVMM steuert dabei stets genau eine virtuelle Maschine, was dessen Design weiter vereinfacht. Selbstverständlich können mehrere UVMMs (für jeweils eine VM) gestartet werden.

Neben UVMM gibt es für L4Re auch andere VMMs [5] [6].

3. Forschungsgebiet Mikrohypervisor

Unserer Erfahrung nach ist es BSI-zugelassenen Produkten gemein, dass deren Systemkonfiguration statisch ist und selbst Gegenstand der Prüfung und Zulassung ist. Konkret wird beim Entwurf L4Re-basierter Produkte ein Systemsetup festgelegt, in dem zum einen die auszuführenden Programme, deren Speicherausstattungs-Zuweisung und deren Kommunikationskanäle festgeschrieben sind. Dieses Systemsetup wird in einer Konfigurationsdatei abgelegt, welche gleichzeitig das Setup beschreibt, und auch zur Laufzeit zur Systemkonfiguration herangezogen wird. Diese Konfigurationsdatei ist integraler Teil der Evaluierung und wird mittels Secure Boot gegen Veränderungen abgesichert. Dabei legen wir zum Entwurfszeitpunkt auch die Speicherausstattung und Konnektivität von virtuellen Maschinen fest, und diese Ressourcen- und Kommunikationsausstattung wird durch den Mikrokern durchgesetzt und kann sich zur Laufzeit des Produkts nicht ändern. Dabei können diese Produkte durchaus dynamische Bestandteile enthalten; so erlauben wir z.B. einen Neustart von virtuellen Maschinen. Eine Änderung der Systemkonfiguration ist jedoch nicht ohne Neuzulassung möglich.

Ein grundlegender Unterschied eines Cloud-Hypervisors zu bisherigen Mikrokernprodukten ist die dynamische Natur eines Hypervisors. Die Anzahl und Konfiguration der virtuellen Maschinen soll zur Laufzeit durch die Cloudmanagementmentsoftware festgelegt werden und ist weder zum Entwurfszeitpunkt noch bei der Zulassung bekannt, ebensowenig wie das Softwaresetup von Gastsystemen: Stattdessen wird die gesamte Softwareausstattung in einer VM vom Kunden des Cloudbetreibers bereitgestellt.

Kernaufgabe eines Cloud-Hypervisors ist es, mehrere virtuelle Maschinen gleichzeitig auf einem Virtualisierungsserver auszuführen. Dabei hat die Trennung zwischen den VMs größte Bedeutung. So muss es absolut ausgeschlossen sein, dass Informationen einer VM durch eine andere VM gelesen oder verändert werden können. Es gibt hier also ein Spannungsfeld zwischen Sharing auf der einen Seite und Trennung auf der anderen Seite.

Soll ein Cloud-Hypervisor also für die Verwendung von Gastsystemen aus höheren Geheimhaltungsstufen wie z.B. GEHEIM zugelassen werden, muss sein Durchsetzungsvermögen bezüglich Trennung evaluiert werden.

3.1. Architektur des Mikrohypervisors

Wie bereits in den vorangegangenen Abschnitten beschrieben, bringt das L4Re -Mikrokernsystem die nötigen Mechanismen mit, die es funktional zur Ausführung von virtuellen Maschinen benötigt.

Damit das L4Re-System jedoch in einer Cloud als Hypervisor zum Einsatz kommen kann, muss auch eine geeignete Hypervisor-API unterstützt werden. Die Frage ist, wie man eine solche API unterstützen kann, ohne die hervorragenden Isolationseigenschaften des Mikrokernsystems zu untergraben.

Wie bereits dargelegt, werden Hypervisor-APIs typischerweise in Nutzeranwendungen implementiert, welche dann direkt auf dem verwendeten Hypervisor ausgeführt werden und dessen Schnittstellen und reichhaltigen APIs verwenden.

Eine Möglichkeit wäre es, die nötigen APIs für unser Mikrokernsystem neu zu implementieren. Da eine solche Implementierung jedoch sehr zeit- und kostenintensiv ist, bietet sich eine Alternative auf Basis von virtuellen Maschinen an: Wir definieren eine statische VM, in welcher wir ein speziell auf die Aufgabe zugeschnittenes Linux-System mit einer bereits existierenden Hypervisor-API-Implementation ausführen. Das erlaubt uns eine Verwendung der reichhaltigen Schnittstellen von Linux und vermeidet eine aufwendige Neuimplementierung von komplexen Modulen wie Dateisystemen und Netzwerkstacks. Wir nennen diese spezielle VM die *Command and Control VM* (CCVM).

Da die CCVM nur genau eine Aufgabe hat, nämlich die Ausführung der Hypervisor-API, kann sie sehr klein gehalten werden. Benötigt werden neben der Hypervisor-API noch ein VPN-Zugang und dessen Konfiguration. Es ist möglich, ihren Massenspeicher komplett aus einem nicht-schreibbaren Bootmodul (initrd) zu beziehen. Dieses Boot-

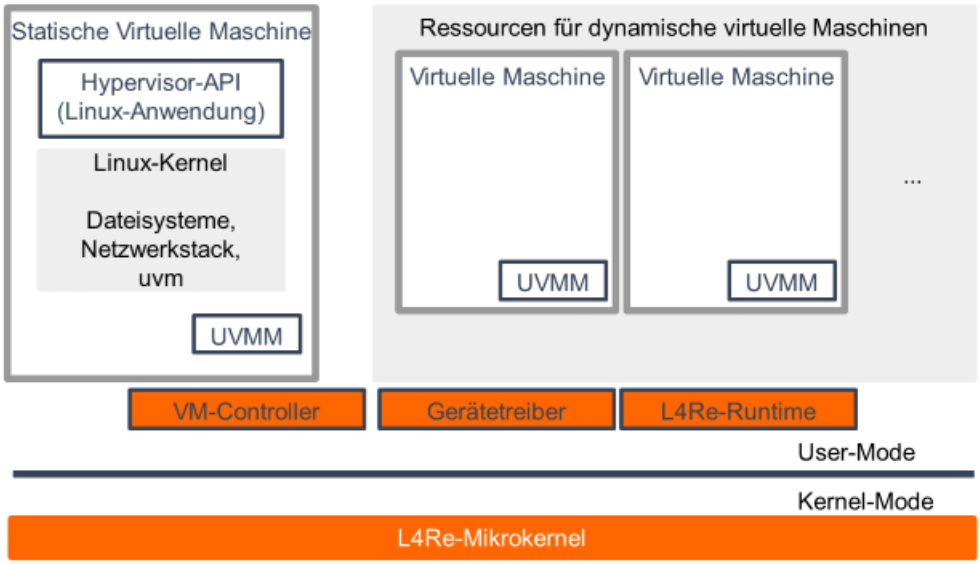


Abb. 2: Architektur des Mikrohypervisors. Orange hinterlegte Komponenten sind kritisch für die Isolationseigenschaften

modul kann zusätzlich mit Mitteln des Secure Boot abgesichert werden. In ersten Experimenten konnten wir dieses Bootmodul auf 50 MiB beschränken. Die CCVM wird weiterhin fest an eine CPU gebunden und fix mit 128 MiB Ram versehen. Eine Anbindung an physische Netzwerkhardware wird dabei statisch konfiguriert. In dieser CCVM können wir als Beispiel die Hypervisor-API Libvirt ausführen.

Im L4Re-System sind Gastsysteme in virtuellen Maschinen nicht in der Lage, Systemaufrufe des Mikrokerns aufzurufen. Es ist daher eine weitere Komponente nötig, die eine Verbindung von der Hypervisor-API zum L4Re-System herstellt. Wir schlagen dafür eine native L4Re-Anwendung mit dem Namen *VM-Controller* vor. Diese wird im Systemsetup mit allen nötigen Berechtigungen ausgestattet und damit ermächtigt, über die entsprechenden Systemressourcen zu verfügen und dynamisch VMs zu erzeugen und zu steuern.

Als Kommunikationskanal zwischen Hypervisor-API in der CCVM und dem VM-Controller kann ein generischer Kanal, wie z.B. virtio-console, verwendet werden.

3.2. Vertrauenswürdigkeit der Isolationseigenschaften des Mikrohypervisors und Vergleich mit KVM

In diesem Abschnitt wollen wir die Isolationseigenschaften des Mikrohypervisors analysieren und mit denen von KVM vergleichen. Dazu analysieren wir zunächst die TCB.

Zur TCB muss der Mikrokern zählen, denn er läuft mit höchsten Privilegien. Im Fall des L4Re-Mikrokerns sind das ca. 30.000 Zeilen Code. Dazu kommt Infrastruktur, welche zum Booten und der Speicherkonfiguration sowie dem Verwalten von Hardwaregeräten dient. Bei L4Re sind das in einer typischen Konfiguration ca. 50.000 Zeilen Code.

Der VMM UVMM kontrolliert pro Instanz genau eine VM und ist damit nicht kritisch für die Isolation zwischen VMs. Die CCVM und die darin laufende Hypervisor-API sind in eine virtuelle Maschine eingesperrt. Die CCVM dient ausschließlich der Aufgabe, die Hypervisor-API auszuführen und deren Befehle an den VM-Controller weiterzuleiten. Damit benötigt die CCVM keinen weiteren Zugriff auf die Ressourcen der Servermaschine. So sind komplexe Betriebssystemkomponenten wie die Dateisysteme und Netzwerkstacks nicht kritisch für die Isolationseigenschaften und müssen daher nicht evaluiert werden.

Der VM-Controller ist hingegen zentral für die Isolation, da er mit Zugriffsberechtigungen auf alle Ressourcen der Servermaschine ausgestattet ist. Wir schätzen, dass ein VM-Controller im L4Re-System mit wenigen tausend Zeilen Code umsetzbar ist, welche zur TCB zählen und evaluiert werden müssen.

Da der VMM UVMM stets nur genau eine virtuelle Maschine steuert, ist auch diese Komponente nicht kritisch für die Trennung der virtuellen Maschinen.

Komponenten, die von mehreren virtuellen Maschinen gemeinsam genutzt werden, welche zur TCB der VMs gehören, müssen evaluiert werden. Ein Beispiel hierfür sind Gerätetreiber.

Insgesamt ist festzustellen, dass die Komplexität des Mikrohypervisors einem vergleichbaren Hypervisor wie KVM in nichts nachsteht. Auch im Mikrohypervisor werden komplexe Komponenten wie VMMs, Dateisysteme und Netzwerkstacks verwendet. Durch den neuartigen Architekturansatz konnte diese Funktionalität jedoch mit Hilfe hardwaredurchgesetzter Grenzen gekapselt werden und ist damit – im Unterschied zu KVM – nicht kritisch für die Isolation und muss daher nicht evaluiert werden.

Insgesamt ist es mit dem Ansatz des Mikrohypervisors möglich, die zu evaluierende Codemenge auf ca. 100.000 Zeilen Code einzuschränken. Diese Menge ist mit heutigen Mitteln mit überschaubarem Aufwand evaluierbar. Das L4Re-Mikrokernsystem wird bereits heute in mehreren Projekten eingesetzt, die bis zur Geheimhaltungsstufe GEHEIM zugelassen wurden. Damit ist gezeigt, dass die Softwarebasis höchsten Anforderungen genügt. Wir sind davon überzeugt, dass auch ein Mikrohypervisor aufbauend auf dem L4Re-Mikrokernsystem bis GEHEIM zulassungsfähig ist.

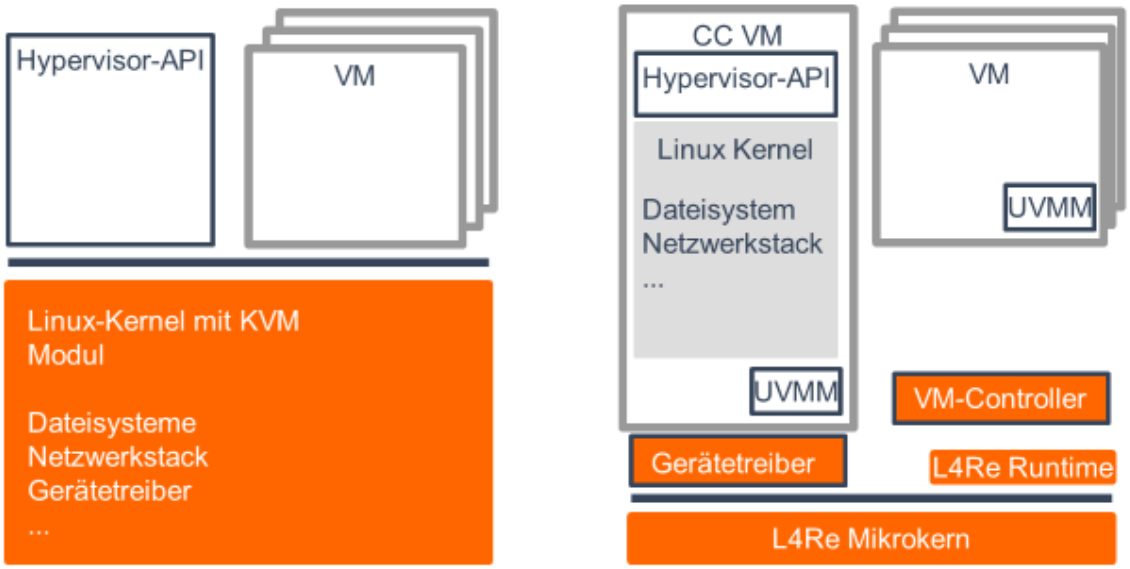


Abb. 3: Vergleich monolithischer Hypervisor KVM auf der linken Seite und Mikrohypervisor auf der rechten Seite. Orange hinterlegte Komponenten sind kritisch für die Isolation zwischen virtuellen Maschinen.

4. Zusammenfassung

In diesem Beitrag haben wir aufgezeigt, dass es mit L4Re möglich ist, einen hochsicheren und zulassungsfähigen Mikrohypervisor zu bauen. Eine solche Architektur ebnet den Weg für eine hochsichere Cloud, welche auch in höheren Geheimhaltungsstufen wie z.B. GEHEIM zum Einsatz kommen kann.

Literaturhinweise

[1] „GlobeNewswire,“ 01 10 2020. [Online]. Available: https://www.globenewswire.com/news-release/2020/10/01/2102033/0/en/Cloud-Computing-Market-Worth-760-98-Billion-at-18-6-CAGR-Tech-Giants-Such-as-IBM-and-Microsoft-to-Focus-on-Introducing-Advanced-Cloud-Solutions-Fortune-Business-Insights.html.

[2] Y. K. Malaiya und J. Denton, „Estimating defect density using test coverage,“ 1998.

[3] T. Günther, M. Hohmuth, A. Lackorzynski und M. Lange, „Mikrokern für zulassungspflichtige Hochsicherheitssysteme,“ in DACH Security, München, 2017.

[4] „Kernel Virtual Machine,“ [Online]. Available: https://www.linux-kvm.org/page/Main_Page. [Zugriff am 08 01 2021].

[5] H. Schild, M. Peter, A. Lackorzynski und A. Warg, „Virtual machines jailed: Virtualization in systems with small trusted computing bases,“ in VDTS '09: Proceedings of the 1st EuroSys Workshop on Virtualization Technology for Dependable Systems, 2009.

[6] S. Liebergeld, „Lightweight Virtualization on Microkernel-based Systems,“ Diplomarbeit Technische Universität Dresden, 2010.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Anwendungsintegration sicherer Digitaler Identitäten unter Berücksichtigung von Security by Design Prinzipien

Tobias Assmann[1], Ann-Kristin Derst[2], Dr. Detlef Hühnlein[1], Tina Hühnlein[1], Hanno Koop[2], Thorben Pohl[2], Michael Rauh[1], Tobias Wich[1]

Kurzfassung:
Der Personalausweis mit Online-Ausweisfunktion steht mittlerweile allen ausweispflichtigen Bürgerinnen und Bürgern zur Verfügung und muss „nur noch“ in entsprechende Anwendungen integriert werden, um weitverbreitet im Internet eine starke pseudonyme Authentisierung oder einen zuverlässigen elektronischen Identitätsnachweis zu ermöglichen. Der vorliegende Beitrag erläutert die praktische Umsetzung der Sicherheitsprinzipien „Security by Design“ und „Security by Default“ am Beispiel der Integration der Online-Ausweisfunktion in die populäre Nextcloud Webanwendung und zeigt, wie unkompliziert eine Anwendungsintegration sicherer digitaler Identitäten in der Praxis erfolgen kann.

Stichworte: Authentifizierung/Authentisierung, BSI TR-02102-1, BSI TR-03110, BSI TR-03116-4, BSI TR-03130, Digitale Identitäten, eID, eIDAS, eID-Template, Personalausweis, SAML, Security by Default, Security by Design, Sichere Online-Ausweisfunktion, SkIDentity

1. Einleitung, Motivation und Zielsetzung

10 Jahre nach seiner deutschlandweiten Einführung steht der Personalausweis mit Online-Ausweisfunktion (kurz: „eID“) inzwischen allen ausweispflichtigen Bundesbürgerinnen und Bundesbürgern für einen sicheren und verlässlichen Nachweis der realen Identität in der digitalen Welt zur Verfügung.

Als Zusatzfunktion bietet der Personalausweis auch eine konfigurierbare Wiedererkennung über das dienste- und kartenspezifische Kennzeichen („Pseudonym“) der Ausweiskarte zur starken Authentisierung des Ausweisinhabers an Webanwendungen und die datenschutzfreundliche Altersverifikation. Durch die sichere Ausweiskarte, die einen besitzabhängigen Authentifizierungsfaktor darstellt, die nur lokal verarbeitete PIN und die zugrundeliegende Sicherheitsinfrastruktur der Online-Ausweisfunktion ist damit ein deutlicher Sicherheitsgewinn gegenüber gängigen Nutzername/Passwort-Verfahren erzielbar.

Im Rahmen eines Projektes[3] zur sicheren und anwenderfreundlichen Integration der Online-Ausweisfunktion wurden sogenannte „eID-Templates“ für weit verbreitete Webanwendungen, wie z.B. Nextcloud (https://nextcloud.com), entwickelt. Diese „eID-Templates“ ermöglichen die starke Authentifizierung und Identifizierung mit dem Personalausweis („eID-Login“). Nach Abschluss der Qualitätssicherung werden die „eID-Templates“ als Open Source veröffentlicht. Durch die Installation über den „Nextcloud App Store“ (https://apps.nextcloud.com) und den Einsatz eines geeigneten „eID-Service“

[1] ecsec GmbH, Sudetenstraße 16, 96247 Michelau
[2] Bundesamt für Sicherheit in der Informationstechnik, Postfach 200363, 53133 Bonn
[3] Projekt 396: eID-Templates – Anwendungsintegration mobiler elektronischer Identitäten auf substanziellem Sicherheitsniveau

(https://skidentity.de), können die Administratoren und Anwender diese „eID-Templates“ auf sehr einfache und anwenderfreundliche Weise – quasi per Mausklick – in Betrieb nehmen. Im Rahmen des Projektes wird der SkIDentity-Dienst [17] für Zwecke der starken Authentisierung kostenlos bereitgestellt. Dieser mehrfach international ausgezeichnete und in den Patentschriften EP2439900 und EP2919145 näher beschriebene Dienst stellt anwendungsspezifische und aus dem Personalausweis abgeleiteten Identitäten und Pseudonyme bereit, die zur starken Authentifizierung genutzt werden können.

Bereits bei der Konzeptionierung wurden Anforderungen zur grundlegenden Absicherung der „eID-Templates“ identifiziert und für die Implementierung sowie den gesamten Softwarelebenszyklus berücksichtigt (Security by Design). Ferner werden die „eID-Templates“ in einem sicheren Zustand ausgeliefert und müssen nicht erst durch zusätzlich notwendige Konfigurationen sicher gemacht werden (Security by Default).

Der vorliegende Beitrag erläutert zunächst in Kapitel 2 die Sicherheitsanforderungen anhand der Prinzipien „Security by Design“ und „Security by Default“, bevor in Kapitel 3 die sichere Anwendungsintegration am Beispiel von Nextcloud im Detail erläutert wird. Der Beitrag schließt mit einem Ausblick auf zukünftige Entwicklungen in Kapitel 4.

2. Konzeptionelle Sicherheitsanforderungen

Durch die Online-Ausweisfunktion des Personalausweises und das beweisbar sichere „Extended Access Control“ (EAC) Protokoll [1] wird ein zuverlässiger gegenseitiger Identitätsnachweis zwischen Ausweisinhaber und Diensteanbieter sowie die feingranulare und datenschutzfreundliche Verwaltung von Berechtigungen bzw. Einwilligungen für den Zugriff auf Dokumenten- und Identitätsattribute ermöglicht.

Neben einer datenschutzfreundlichen Altersverifikation und Wohnortabfrage bietet der Personalausweis insbesondere auch ein dienste- und kartenspezifisches Kennzeichen (Pseudonym), mit dem ein sicheres Authentisierungsverfahren auf Basis der zwei Faktoren Besitz (Ausweis) und Wissen (PIN) ermöglicht wird. Damit das konzeptionell sichere eID-basierte Login-Verfahren am Ende auch in der Praxis sicher ist, müssen zahlreiche Sicherheitsaspekte beim Entwurf, der Implementierung und beim Betrieb der Lösung beachtet und sorgfältig umgesetzt werden.

Im Vergleich zu den typischen Diensten und Organisationen, wie z.B. Bundes- und Landesbehörden und große Unternehmen, die bisher die Online-Ausweisfunktion des Personalausweises integriert haben, wird mit den „eID-Templates“ eine komplett neue Zielgruppe angesprochen. Bislang musste der Anbieter und Betreiber eines Dienstes entsprechendes Wissen über die notwendigen organisatorischen Prozesse, insbesondere die Beantragung des Berechtigungszertifikates, sowie technisches Wissen für die Integration des „eID-Servers“ gemäß BSI TR-03130 [2] über den Dienst haben, um die eID-Funktionalität in seiner Anwendung nutzbar zu machen. Nun wird mit den „eID-Templates“ die eID insbesondere für kleine Organisationen und Anbieter zugänglich gemacht. Deshalb ist es erstrebenswert, auf individuelle Sicherheitsbetrachtungen und sicherheitskritische Entscheidungen im Betrieb sowie die Beantragung und die Inbetriebnahme eines eigenen Berechtigungszertifikates nach Möglichkeit zu verzichten und

trotzdem ein hohes Sicherheitsniveau, auch für im Umgang mit eID unerfahreneren Betreibern, zu gewährleisten.

Bereits beim Entwurf der „eID-Templates“ wurde der Prozess der Registrierung des Anwendungsdienstes beim „eID-Service“ (SkIDentity) optimiert und sichere Werte vorkonfiguriert, um den Betreiber bei der sicheren Inbetriebnahme zu unterstützen („Security by Default“). Bei der Konzeption und Entwicklung der Lösung wurden bewusst die im OWASP-Projekt entwickelten „Security by Design“-Prinzipien [3] berücksichtigt.

Von den zehn Prinzipien sind beim spezifischen „eID-Template“-Anwendungsfall, bei dem eine Login- und Registrierungs-Komponente in einer bereits existierenden Webanwendung ergänzt wird, die folgenden Grundsätze besonders wichtig:

- **Sichere Grundeinstellungen (Establish Secure Defaults)**
  Der Betreiber muss bei der Konfiguration des „eID-Templates“ keine Auswahl von Protokollparametern treffen, da bereits sichere Werte voreingestellt sind.
- **Zuverlässige Ausnahmebehandlung (Fail Securely)**
  Konstrukte im Quellcode, die zu einem risikobehafteten Systemzustand führen können, werden systematisch vermieden, was durch Code Reviews und Tests überprüft wird.
- **Vertraue keinem externen Dienst (Don’t Trust Services)**
  Die zuverlässige Validierung der von einem externen Dienst erhaltenen Daten erfolgt durch die Mechanismen des eingesetzten Single Sign-On-Protokolls, wie z.B. SAML [4]. Neben dem notwendigerweise vertrauenswürdigen Pluginverzeichnis und dem auch als „eID-Service“ fungierenden Identity Provider Service gibt es keine weiteren externen Dienste.
- **Keine Sicherheit durch Verschleierung (Avoid Security by Obscurity)**
  Es kommen nur sicherheitstechnisch wohluntersuchte Standardprotokolle, wie beispielsweise SAML [4], in Verbindung mit geeigneten kryptografischen Algorithmen gemäß den einschlägigen Empfehlungen des BSI (vgl. [5] und [13]) zum Einsatz.
- **Bevorzuge einfache Sicherheitslösungen (Keep Security Simple)**
  Die „eID-Templates“ bringen alle Komponenten, wie z.B. eine SAML Service Provider Implementierung mit. Es muss keine zusätzliche externe Software installiert werden.
- **Korrekte Behebung von Sicherheitsproblemen (Fix Security Issues Correctly)**
  Automatisierte Tests sind Teil des Entwicklungszyklus und werden sukzessive auch um Tests für auftretende Sicherheitslücken ergänzt.
- **Mehrstufige Sicherheitsmechanismen (Principle of Defense in Depth)**
  Mehrstufige Abwehrmechanismen, die über die SAML-basierte Anmeldung an einem Benutzerkonto hinausgehen, sind unabhängig von der mit den „eID-Templates“ realisierten starken Authentifizierung. Jedoch kommen bei bestimmten

Webanwendungen, wie z.B. Nextcloud, neben den anwendungsspezifischen Sicherheitsmechanismen („Annotations“) weitere Prüfungen zum Einsatz, um sicherzustellen, dass bestimmte Funktionen nur von einem Administrator genutzt werden können.

Als weniger relevant oder nicht anwendbar wurden folgende Security by Design-Prinzipien klassifiziert:

- **Minimale Angriffsfläche (Minimize Attack Surface Area)**
  Die Konfiguration und Administration eines Benutzerkontos ist bei einem Login-Verfahren per Definition nur für angemeldete Nutzer möglich. Eine Trennung in weitergehend geschützte und zusätzlich getrennte Backendsysteme ist bei der Realisierung als Plugin Komponente nicht anwendbar.
- **Minimale Rechte (Principle of Least Privilege)**
  Die Anwendungen, in denen die „eID-Templates“ laufen, besitzen typischerweise entsprechende Rollenkonzepte. Allerdings ist diese Autorisierung unabhängig von der mit den „eID-Templates“ realisierten starken Authentifizierung und gegebenenfalls Identifizierung.
- **Verteilte Verantwortung (Separation of Duties)**
  Rollen werden von der Hauptanwendung verwaltet und es werden keine speziellen zusätzlichen Rollen für die „eID-Templates“ benötigt.

3. Praktische Umsetzung am Beispiel Nextcloud

3.1. Das eID-Login System im Überblick

Das Gesamtsystem für das anvisierte „eID-Login“ für die Nextcloud Webanwendung ist in Abbildung 1 dargestellt.

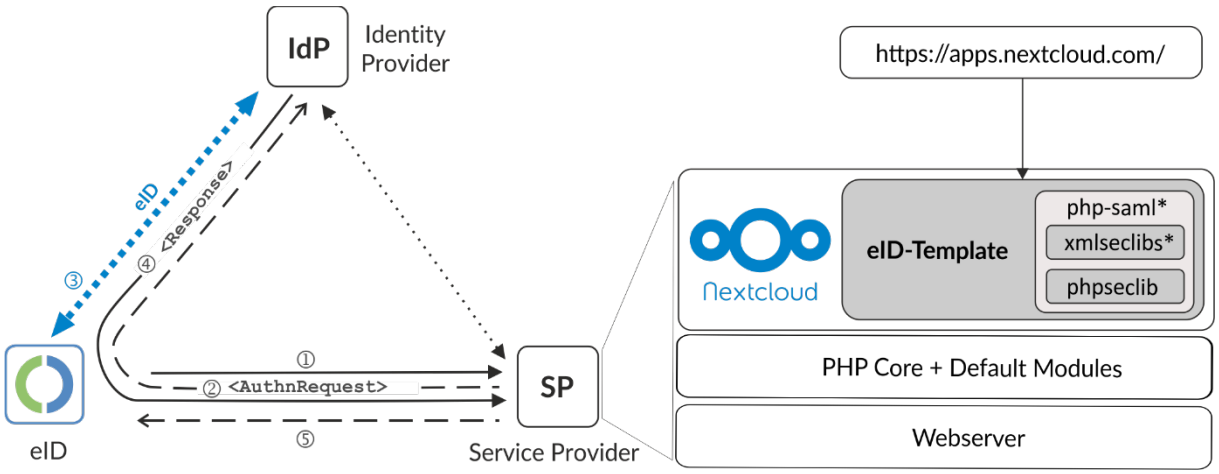


Abbildung 1:Das eID-Login-System im Überblick

Insgesamt besteht das Gesamtsystem aus dem Service Provider (SP) und dem Identity Provider (IdP) gemäß SAML [4] und der „eID-Komponente“ beim Nutzer. Diese Komponente besteht technisch aus dem Personalausweis mit Online-Ausweisfunktion gemäß [6] und [7] und einem eID-Client gemäß [8], wie z.B. der Open eCard App [9] oder der

AusweisApp2 [10]. In den Identity Provider ist ein „eID-Server“ gemäß [2] integriert, der die Abwicklung des EAC-Protokolls übernimmt, um die Authentifizierung und gegebenenfalls Identifizierung des Nutzers durchzuführen und das Ergebnis unter Beteiligung des Nutzers in einer digital signierten Bestätigung („Assertion“) über das SAML-Protokoll [4] an den Service Provider zu übertragen.

3.2. Architektur des Service Providers und des eID-Templates

Im anvisierten Anwendungsfall ist der Service Provider die PHP-basierte Webanwendung Nextcloud, die auf einem Webserver mit einer entsprechenden PHP-Installation („PHP Core + Default Modules“) läuft. Damit das gewünschte „eID-Login“ für die Anmeldung an der Nextcloud-Anwendung genutzt werden kann, wurde der vorgesehene Erweiterungsmechanismus genutzt und eine geeignete „Nextcloud-App“ („eID-Template“) entwickelt. Nach Abschluss der Qualitätssicherung wird diese App als Open Source veröffentlicht und kann dann einfach über den „Nextcloud App Store“ (https://apps.nextcloud.com) installiert werden.

Um im Einklang mit BSI TR-02102-1 [5] und BSI TR-03116-4 [13] (Abschnitte 4.3.1 und 4.7), wie dies ab dem Jahr 2021 für Projekte der Bundesregierung gefordert ist, RSA-Signaturen mit zufallsgesteuertem Signaturpadding („Probabilistic Signature Scheme“, RSASSA-PSS) gemäß RFC 8071 [14] bzw. PKCS #1 (Version 2.2) zu unterstützen, musste das eingesetzte SAML PHP Toolkit „php-saml“ [11] und die von diesem genutzte XML-Signatur-Bibliothek „xmlseclib“ [15] diesbezüglich erweitert werden, was in Abbildung 1 durch den „*“ angedeutet wird. Da RSASSA-PSS bislang auch nicht von Standard-PHP-Installationen unterstützt wird, musste zusätzlich eine spezialisierte kryptografische Bibliothek „phpseclib“ [16] für die Realisierung der „eID-Templates“ genutzt werden.

3.3. Anwendungsspezifisches Pseudonym

Da der Diensteanbieter die Authentifizierung und Identifizierung des Nutzers an einen geeigneten Identity Provider mit entsprechendem „eID-Service“ (z.B. SkIDentity [17]) überträgt, ist kein eigenes Berechtigungszertifikat zum Auslesen des Pseudonyms erforderlich. Dadurch wird die Realisierung eines eID-basierten Login-Verfahrens erheblich vereinfacht.

Hierbei wird nicht das vom Personalausweis berechnete Identity Provider-spezifische Pseudonym (siehe BSI TR-03110 [6], Teil 2, Abschnitt 3.6) an den angeschlossenen Dienst weitergeleitet, sondern nur eine daraus mit einer schlüsselbasierten kryptografischen Einwegfunktion (siehe BSI TR-02102-1 [5], Abschnitt 5.3) „abgeleitete Identität“, die letztlich mit dem Benutzerkonto des Anwenders verknüpft wird. So entsteht ein sehr sicheres und gleichzeitig datenschutzfreundliches Authentifizierungsverfahren auf Basis des dienst- und kartenspezifischen Kennzeichens des Personalausweises.

3.4. Praktische Umsetzung des Prinzips „Security by Default“

Durch die angestrebte Bereitstellung der „eID-Templates“ im für die jeweilige Webanwendung geeigneten Pluginverzeichnis und die einfache Anbindung an einen geeigne-

ten Identity Provider kann die Einrichtung des „eID-Templates“ mit wenigen Mausklicks in einer sehr einfachen Weise erfolgen. Der Administrator des „eID-Templates“ muss sich nicht um die Erzeugung von kryptografischen Schlüsseln oder die korrekte Einstellung sicherheitsrelevanter Parameter, wie z.B. die eingesetzten kryptografischen Algorithmen und Schlüssellängen im Einklang mit BSI TR-02102-1 [5] und BSI TR-03116-4 [13], kümmern. Vielmehr werden stets sichere Werte verwendet und Fehlkonfigurationen vermieden (Security by Default). Darüber hinaus steht dem erfahrenen Administrator die Möglichkeit der manuellen Konfiguration zur Verfügung.

3.5. Praktische Umsetzung des Prinzips „Security by Design“

Bei der Entwicklung der „eID-Templates“ wurden sicherheitsrelevante Aspekte von Anfang an berücksichtigt (Security by Design). So konzentriert sich das Plugin bewusst ausschließlich auf die Kernfunktionalität (Authentifizierung) und bietet Angreifern eine möglichst geringe Angriffsfläche. Zudem wird mit SAML [4] ein standardisiertes und sicherheitstechnisch sehr gut untersuchtes Protokoll verwendet und, soweit möglich[4], auf APIs und Funktionalitäten zurückgegriffen, die die zugrundeliegende Plattform bereits bietet.

Sichere Grundeinstellungen (Establish Secure Defaults)

Bei der Nutzung des SAML 2.0 Protokolls [4] in den „eID-Templates“ werden per Default geeignete sicherheitsrelevante Features genutzt. So wird z.B. die Signatur der vom Identity Provider kommenden Metadaten und Assertions erzwungen. Die vom Service Provider versendeten Requests werden genauso wie die ihn beschreibenden Metadaten nur signiert ausgeliefert. Darüber hinaus wurden bei der Implementierung der „eID-Templates“ einschlägige Sicherheitsempfehlungen [18] berücksichtigt, um bekannte Angriffsvektoren auszuschließen.

Ähnlich wie bei der Aktualisierung von Plugins (siehe „Korrekte Behebung von Sicherheitsproblemen“ unten) müssen auch kryptografische Schlüssel und Zertifikate in regelmäßigen Abständen erneuert werden. In diesem Zug können auch Algorithmen und Schlüssellängen an den aktuellen Stand der Technik angepasst werden. Daher wurde in den Anwendungen ein (semi-)automatischer Key-Rollover implementiert. Da für die bidirektionale Absicherung der SAML-Kommunikation auch die Konfiguration am Identity Provider angepasst werden muss, ist eine vollständige Automatisierung nicht ohne Weiteres möglich. Der Admin wird mit entsprechender Vorlaufzeit per E-Mail benachrichtigt und muss daraufhin lediglich die automatische Aktualisierung manuell anstoßen, sobald die Anpassungen beim Identity Provider erfolgt sind.

Zuverlässige Ausnahmebehandlung (Fail Securely)

Konstrukte im Quellcode, die zu einem risikobehafteten Systemzustand führen können, werden systematisch vermieden, was durch Code Reviews und entsprechende Tests überprüft wird. Bei der Verarbeitung der SAML-Assertion beim Service Provider können diverse Zustände im System zu unterschiedlichen Fehlern führen. Möglicherweise fehlen wichtige Daten in der SAML-Assertion, vielleicht wurde der Benutzer deaktiviert

[4] Zu den diesbezüglichen Grenzen siehe Abschnitt 3.2.

oder der Benutzer kann im System nicht gefunden werden. In all diesen Fällen wird der Login-Vorgang abgebrochen und über einen Ausnahme-Mechanismus sichergestellt, dass der Client im nicht angemeldeten Zustand eine geeignete, für einen potenziellen Angreifer nicht hilfreiche, Fehlermeldung erhält. Auf der anderen Seite werden Fehler im internen Logging des Systems gespeichert, sodass später gegebenenfalls die Nachvollziehbarkeit sichergestellt ist. Erst nachdem die Konfiguration komplett erfolgreich abgeschlossen ist, wird dem Benutzer das eID-Login mit der Online-Ausweisfunktion angeboten. Bis dahin werden etwaige Ausnahmezustände sorgfältig behandelt.

Vertraue keinem externen Dienst (Don't Trust Services)
Die Anbindung an den Identity Provider erfolgt, wie dies bei SAML [4] üblich ist, kryptografisch abgesichert über den Austausch von Zertifikaten innerhalb entsprechender Metadatenstrukturen [12], wodurch die Integrität und Authentizität der übermittelten Daten mittels digitaler Signaturen gewährleistet ist. Die zuverlässige Validierung der von einem externen Dienst erhaltenen Daten erfolgt durch die Mechanismen des SAML-Protokolls [4] unter Berücksichtigung der einschlägigen Sicherheitsempfehlungen [18]. Neben dem Pluginverzeichnis und dem notwendigerweise vertrauenswürdigen und auch als „eID-Service" fungierenden Identity Provider Service gibt es keine weiteren externen Dienste.

Keine Sicherheit durch Verschleierung (Avoid Security by Obscurity)
Der Quelltext der „eID-Templates" wird bei der Veröffentlichung auf GitHub der Öffentlichkeit als Open Source bereitgestellt. Es können also alle interessierten Parteien den Quelltext einsehen und im Detail analysieren. Auf den ohnehin fraglichen Versuch, einen sicherheitsrelevanten Vorteil durch Geheimhaltung des Quelltextes zu erzielen, wird bewusst verzichtet. Vielmehr sind alle Mitglieder der Open Source Community und alle Sicherheitsexperten herzlich eingeladen, nicht nur etwaige Verbesserungsvorschläge oder neu entdeckte Sicherheitslücken in einer geeigneten Weise zu melden, sondern auch aktiv an der Bereitstellung von weiteren „eID-Templates" für andere Webanwendungen mitzuwirken.

Bevorzuge einfache Sicherheitslösungen (Keep Security Simple)
Die „eID-Templates" bringen alle funktionalen Komponenten wie z.B. eine SAML Service Provider Implementierung mit, und es muss keine zusätzliche externe Software installiert werden. Dafür mussten im Projekt, wie in Abschnitt 3.2 näher erläutert, verschiedene sicherheitsrelevante Bibliotheken ergänzt werden, um die für XML-Signaturen formulierten algorithmischen Vorgaben aus BSI TR-03116-4 [13] zu erfüllen. Diese Ergänzung widerspricht aber etwas der empfohlenen Bevorzugung von einfachen Sicherheitslösungen. Ob der Einsatz von RSASSA-PSS gemäß RFC 8071 [14] bzw. PKCS #1 (Version 2.2) im SAML-Kontext, bei dem ein Angreifer die zu signierenden Daten nicht uneingeschränkt beeinflussen oder gar gezielt wählen kann, faktisch zu einem Sicherheitsgewinn führt und welchen Einfluss die in [13] geforderte Unterstützung von RSASSA-PSS auf die Interoperabilität hat, konnte innerhalb des Projekts noch nicht abschließend bewertet werden. Erst danach könnte eine „einfachere Sicherheitslösung für XML-Signaturen" auch in die „eID-Templates" integriert werden.

Korrekte Behebung von Sicherheitsproblemen (Fix Security Issues Correctly)
Für die Absicherung einer Webanwendung ist es von zentraler Bedeutung, dass Sicherheitslücken zeitnah behoben und kritische Sicherheitspatches eingespielt werden. Gerade bei Installationen, die nicht von einer eigenen IT-Abteilung betreut und gewartet werden, erfolgt dies häufig zu selten, was Angreifern zahlreiche Angriffsmöglichkeiten bietet. Viele Webanwendungen bieten inzwischen automatische Aktualisierungen an, was bereits einen großen Sicherheitsgewinn darstellt. Häufig vernachlässigt werden aber noch die Erweiterungen, die ebenfalls Ziel zahlreicher Attacken sind. Im Rahmen des Projekts wurde daher versucht, die Möglichkeiten der automatischen Aktualisierung zu nutzen, die die jeweilige Plattform bietet. Bei Nextcloud wird der Anwender zumindest via Notifizierung auf ein zur Verfügung stehendes Update hingewiesen. Darüber hinaus sind automatisierte Tests Teil des Entwicklungszyklus und es ist perspektivisch geplant, sukzessive auch Tests für zukünftig bekannt werdende Sicherheitslücken zu ergänzen. Sollten zukünftig Schwachstellen offenbar werden, würden hierbei nicht nur etwaige „Symptome“ behandelt, sondern es könnte direkt die eigentliche Ursache des Sicherheitsproblems behoben werden.

Mehrstufige Sicherheitsmechanismen (Principle of Defense in Depth)
In der Nextcloud-Plattform existiert ein Rollenkonzept, welches bestimmte Anwender zu Administratoren mit erweiterten Rechten erhebt. Diese besonders privilegierte Gruppe von Benutzern ist überhaupt nur in der Lage die „eID-Templates“ zu installieren und zu konfigurieren.

Um sicherzustellen, dass bestimmte Methoden nur für Administratoren zugänglich sind, können diese mit entsprechenden „Annotations“ versehen werden.

```
/**
 * Prepare a SAML certificate rollover.
 *
 * @EnforceTls
 * @AdminRequired
 *
 * @return DataResponse
 */
public function prepareRollover() : DataResponse {
```

Abbildung 2: Beispiel für Annotations bei Nextcloud

Sollte nun dieser Mechanismus durch einen Fehler in der Verarbeitung von „Annotations“ oder ähnlichem versagen, so wird von der Anwendung zusätzlich geprüft, dass die Aktion von einem angemeldeten Administrator durchgeführt wird.

3.4. Integration, Einrichtung und Nutzung des eID-Login-Verfahrens

Die „eID-Templates“ wurden gemäß den Designrichtlinien der jeweiligen Plattform (z.B. Nextcloud) umgesetzt. Die Integration erfolgt üblicherweise über den jeweiligen App- bzw. Plugin-Store. Somit wird der Code einem zusätzlichen Review-Prozess unterzogen und die Installation der „eID-Templates“ erfolgt von einer vertrauenswürdigen Quelle per Mausklick.

Die Nutzung der Online-Ausweisfunktion oder die Integration von SAML-basierten Authentisierungsverfahren ist im Allgemeinen mit erheblichen technischen Hürden verbunden. Im Rahmen dieses Projekts wurde dieser Prozess soweit möglich vereinfacht. Im einfachsten Fall wird mit einem einzigen Mausklick die betreffende Instanz als Service Provider bei einem vorkonfigurierten Identity Provider registriert, wobei nur die einmalige Authentisierung und Identifizierung des Administrators als Vertreter des Anbieters des Online-Dienstes mittels Online-Ausweisfunktion nötig ist.

Nach der korrekten Anbindung an den Identity Provider haben Benutzer des Dienstes die Möglichkeit, die Online-Ausweisfunktion des Personalausweises als zusätzliche, sichere Authentisierungsoption zu nutzen. Auch hierfür ist eine einmalige Authentisierung beim Identity Provider ausreichend. In diesem Zug wird die abgeleitete Identität dem Benutzerkonto zugeordnet, was im Anschluss eine Zwei-Faktor-Authentisierung ohne die Nutzung von Benutzernamen und Passwort ermöglicht.

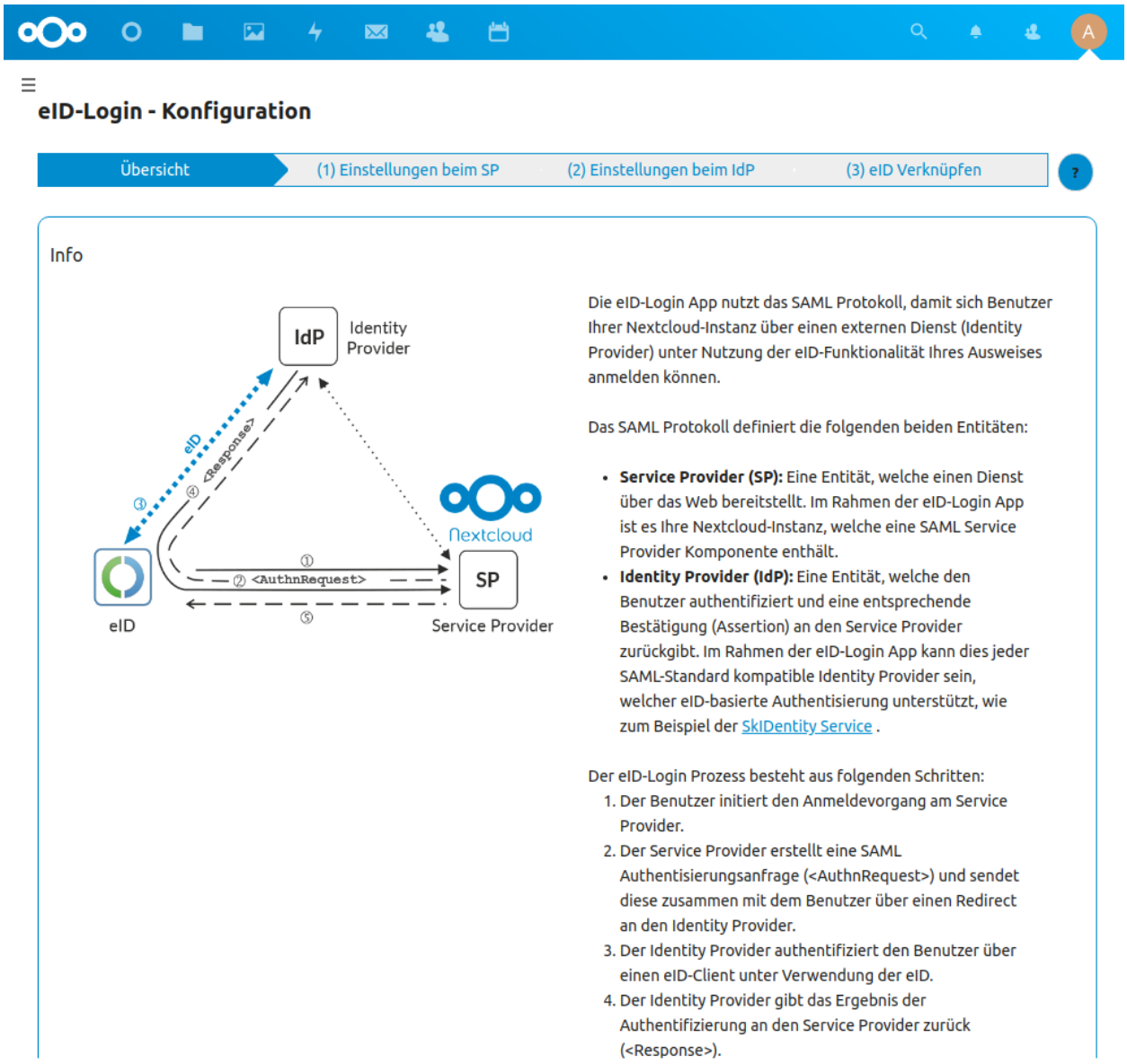


Abbildung 3: Konfiguration der eID-Login-App in Nextcloud

Insgesamt kann, wie in Abbildung 3 angedeutet, das eID-Login-Verfahren durch einfache und möglichst leicht verständliche Konfigurationsschritte sicher eingerichtet und in Betrieb genommen werden.

Wie in Abbildung 4 dargestellt, kann der Endanwender das „eID-Login“ einfach über einen entsprechenden Button auf der Anmeldeseite von Nextcloud nutzen.

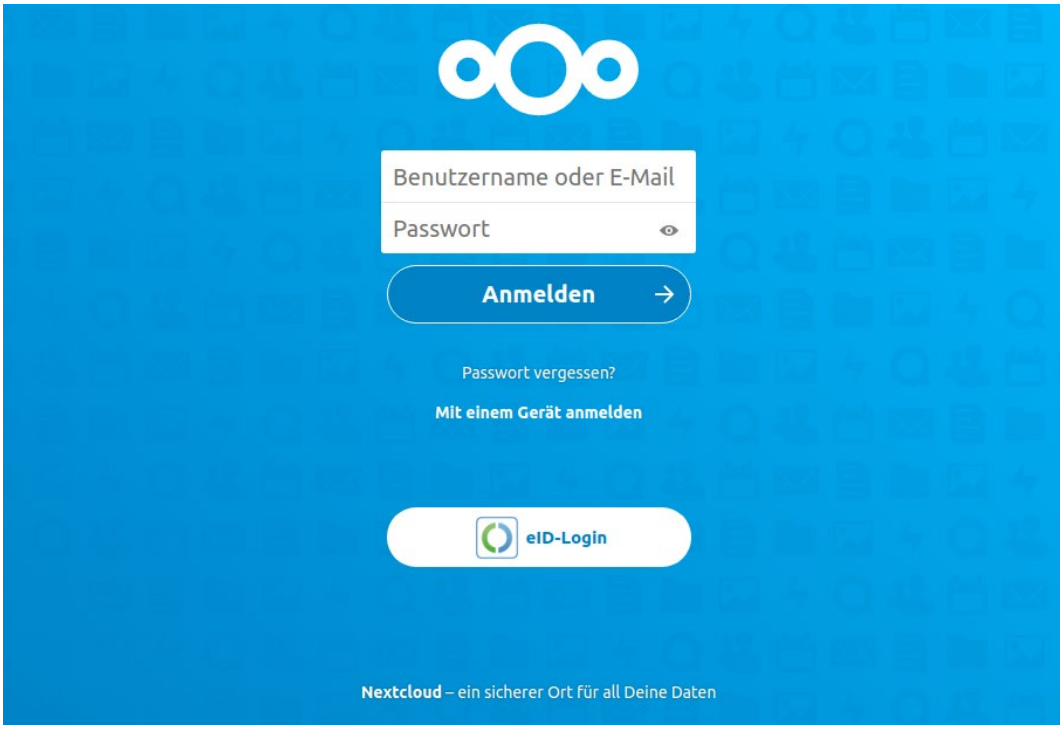


Abbildung 4: Der „eID-Login“ Button auf der Anmeldeseite von Nextcloud

4. Zusammenfassung und Ausblick

Im Rahmen des hier vorgestellten Projektes werden von der ecsec GmbH im Auftrag des BSI „eID-Templates“ für WordPress, Nextcloud und TYPO3 erstellt, die nach Abschluss der Qualitätssicherung als Open Source veröffentlicht werden. Durch den im Rahmen des Projektes kostenlos bereitgestellten „eID-Service“ (SkIDentity) kann die Online-Ausweisfunktion des Personalausweises in diesen Anwendungen sofort zur starken Authentisierung genutzt werden. Darüber hinaus kann die Online-Ausweisfunktion des Personalausweises für die datenschutzfreundliche Altersverifikation und den sicheren Identitätsnachweis genutzt werden. Gegenstand des Projektes ist zudem die Bereitstellung eines Entwicklungsleitfadens mit dem Dritte selbst eID-Templates für weitere Webanwendungen entwickeln können.

Auf Basis dieser Vorarbeiten sind unterschiedliche zukünftige Weiterentwicklungen denkbar. Eine naheliegende Erweiterung der heute verfügbaren „eID-Templates“ könnte darin bestehen, neben dem Personalausweis auch die geplante Smartphone-basierte eID-Variante zu unterstützen, was eine geringfügige Anpassung der Protokollabläufe bezüglich der eID-Aktivierung notwendig machen würde.

Darüber hinaus könnten neben dem Personalausweis auch weitere Chipkarten der e-Card-Strategie des Bundes [20], gemäß der eIDAS-Verordnung notifizierte Identifizierungsmittel oder OZG-Nutzerkonten gemäß [21] unterstützt werden, was insbesondere eine geeignete Auswahlmöglichkeit („Identity Selector“) für den Nutzer nötig machen würde.

Schließlich können die als Open Source verfügbaren Module der hier erstellten „eID-Templates“ an die spezifischen Bedürfnisse anderer Cloud- und Webanwendungen angepasst werden, damit man elektronische Ausweise perspektivisch überall im Internet für die starke Authentisierung und zuverlässige Identifizierung nutzen kann.

Literaturhinweise

[1] Ö. Dagdelen, M. Fischlin: *Security Analysis of the Extended Access Control Protocol for Machine Readable Travel Documents*, International Conference on Information Security, Springer, 2010

[2] BSI: *eID-Server*, BSI TR-03130, Part 1-4, 2017-2020

[3] OWASP: *Security by Design Principles*, https://wiki.owasp.org/index.php/Security_by_Design_Principles

[4] S. Cantor, J. Kemp, R. Philpott, E. Maler: *Assertions and Protocol for the OASIS Security Assertion Markup Language (SAML) V2.0*. OASIS Standard, 15.03.2005, http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf, 2005

[5] BSI: *Kryptographische Verfahren: Empfehlungen und Schlüssellängen*, BSI TR-02102-1, 2020

[6] BSI: *Advanced Security Mechanisms for Machine Readable Travel Documents and eIDAS Token*, BSI TR-03110, Part 1-4, 2015-2016

[7] BSI: *eID-Karten mit eID- und eSign-Anwendung basierend auf Extended Access Control – Personalausweis, elektronischer Aufenthaltstitel und eID-Karte für Unionsbürger*, BSI TR-03127, 2020

[8] BSI: *eID-Client*, BSI TR-03124, Part 1-2, 2017

[9] ecsec: *Open eCard App*, https://www.openecard.org/

[10] Governikus: *AusweisApp2*, https://www.ausweisapp.bund.de/

[11] OneLogin: *OneLogin's SAML PHP Toolkit*, https://github.com/onelogin/php-saml

[12] S. Cantor, J. Moreh, R. Philpott, E. Maler: *Metadata for the OASIS Security Assertion Markup Language (SAML) V2.0*, OASIS Standard, 15.03.2005, https://docs.oasis-open.org/security/saml/v2.0/saml-metadata-2.0-os.pdf, 2005

[13] BSI: *Kryptographische Vorgaben für Projekte der Bundesregierung, BSI TR-03116-4, Kommunikationsverfahren in Anwendungen*, 2020

[14] K. Moriarty, J. Jonsson, B. Kaliski, A. Rusch: *PKCS #1: RSA Cryptography Specifications Version 2.2*, IETF RFC 8017, 2016

[15] *xmlseclib*, https://github.com/robrichards/xmlseclibs

[16] *phpseclib*, https://github.com/phpseclib/phpseclib

[17] ecsec: *SkIDentity – Sicherer Identitätsnachweis im Netz*, https://skidentity.de

[18] OWASP: *SAML Security Cheat Sheet*, https://cheatsheetseries.owasp.org/cheatsheets/SAML_Security_Cheat_Sheet.html

[19] Kantara: *SAML V2.0 Deployment Profile for Federation Interoperability*, 2019, https://kantarainitiative.github.io/SAMLprofiles/saml2int.html

[20] B. Kowalski: *Die eCard-Strategie der Bundesregierung im Überblick*. BIOSIG 2007: Biometrics and Electronic Signatures, 108 LNI, SS. 87–96, http://subs.emis.de/LNI/Proceedings/Proceedings108/gi-proc-108-008.pdf , 2007

[21] BSI: *Servicekonten*, BSI TR-03160, Teil 1-2, 2020

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Sicherheit von Single Sign-On: Ein Überblick

Maximilian Westers[1], Prof. Dr.-Ing. Andreas Mayer[1]

Kurzfassung:
Single Sign-On (SSO), basierend auf den Standards OAuth 2.0 und OpenID Connect, ist eine vielversprechende und weit verbreitete Möglichkeit, das existierende Problem der Passwortinflation und die damit verbundenen Sicherheitsprobleme im Internet zu lösen. Jedoch ist auch SSO nicht frei von Sicherheitsmängeln. Besonders bei der Integration von SSO in Anwendungen kommt es häufig zu vermeidbaren Sicherheitslücken. Aber auch bei dem in SSO notwendigen zentralen Authentifizierungsdienst (Identity Provider), existieren vielfältige sicherheitsrelevante Anforderungen und Sicherheits-Fallstricke.

Dieser Artikel erklärt die technische Vorgehensweise der Authentifizierung und Autorisierung über SSO und erläutert anschließend bekannte Sicherheitsprobleme. Diese existieren sowohl auf der Protokollebene als auch bei der Umsetzung. Es werden für beide Bereiche Schwachstellen aufgezeigt, die daraus resultierenden Angriffsmöglichkeiten erklärt und existierende Security Best Practices erläutert.

Stichworte: Cross-Site Request Forgery (CSRF), OAuth 2.0, OpenID Connect, OIDC, PKCE, Proof Key for Code Exchange (PKCE), Security Best Practices, Single Point of Failure, Single Sign-On, SSO

1. Einleitung und Motivation

Die Passwort-basierte Authentifizierung dominiert auch heute noch das Internet. Kaum ein Unternehmen bietet seinen Nutzern nicht die Möglichkeit, sich über die altbekannte Kombination aus Nutzername und Passwort bei Diensten zu authentifizieren. Doch diese Methode hat viele bekannte Probleme und Schwächen [1], [2]. Längst nicht alle Nutzer halten sich bei der Vergabe von Passwörtern an gute Passwort-Richtlinien[2] oder sie verwenden das gleiche Passwort für mehrere Dienste. Dies führt dazu, dass schwache Passwörter per Brute Force-Angriff erratbar sind oder schlecht geschützte Zugangsdaten eines kompromittierten Anbieters den Zugriff auf weitere Dienste erlauben [3]. Selbst die Nutzung von Passwortmanagern bietet nicht immer ausreichend Schutz vor den genannten Problemen [4].

Eine vielversprechende Lösung für diese Probleme ist Single Sign-On (SSO). Dieses Konzept beschreibt das Vorgehen, den Prozess der Authentifizierung eines Nutzers für einen bestimmten Dienstleister – auch Service Provider (SP) oder Client genannt – an eine vertrauenswürdige dritte Partei auszulagern. Diese wird Identity Provider (IdP) genannt. Sobald der Nutzer bei einem IdP registriert ist, kann er sich über diesen bei beliebigen SPs, mit denen eine Vertrauensbeziehung besteht, authentifizieren. Ein Nutzer muss sich somit keine Vielzahl von unterschiedlichen Passwörtern mehr merken, son-

---

[1] Hochschule Heilbronn

[2] Siehe: https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Cyber-Sicherheitsempfehlungen/Accountschutz/Sichere-Passwoerter-erstellen/sichere-passwoerter-erstellen_node.html

dern authentifiziert sich einmalig mit einem starken Passwort am IdP. Für ein noch höheres Sicherheitsniveau kann dabei auch eine Mehrfaktorauthentifizierung eingesetzt werden.

SSO wird von vielen großen Unternehmen wie beispielsweise Amazon, Google, Microsoft und PayPal unterstützt. Konkret sind vor allem die Standards OAuth 2.0 [5] sowie OpenID Connect (OIDC) [6] im Internet weit verbreitet. Während OAuth die Autorisierung bei Multiparteien-Anwendungen erlaubt, erweitert OIDC den OAuth-Standard um die Möglichkeit der Authentifizierung. Beide Protokolle ermöglichen SSO und erhöhen somit den Benutzerkomfort deutlich. Jedoch entsteht aus Sicherheitssicht ein Single Point of Failure. Verschiedene Studien zeigen, dass SSO-Protokolle nicht Out of the Box sicher sind [7], [8]. Sowohl SP als auch IdP müssen bei der Umsetzung von SSO eine Vielzahl von Anforderungen erfüllen, damit der Nutzer vor Angriffen, wie beispielsweise Cross-Site Request Forgery (CSRF) [9] sicher ist. Bereits 2014 untersuchten Zhou und Evans automatisiert die Sicherheit von SSO-Anwendungen [10]. Dabei wurden bei 20 % der 1660 SPs, welche SSO über Facebook anboten, Schwachstellen gefunden. Obwohl vorgefertigte SDKs für die Einbindung der SSO-Technologie existieren, scheint die Integration sicherheitskritischer Komponenten schwierig zu sein.

Sicherheitsvorfälle können sowohl auf der Seite der Clientanwendung – dem SP – als auch auf der Seite des Autorisierungs- und Authentifizierungsdiensts – dem IdP – entstehen. Aus Sicht des SP gibt es verschiedene Fallstricke, auf die die Entwickler bei der Implementierung achten müssen. Dokumentierte Sicherheitslücken zeigen aber, dass dies nicht immer gelingt. So existiert beispielsweise der CVE-Eintrag CVE-2019-10315 für das Jenkins GitHub Authentication Plugin, welches auf eine fehlerhafte Integration von SSO zurückzuführen ist. Sicherheitsrelevante Parameter wurden hier nicht überprüft und führten somit zu der Möglichkeit CSRF-Angriffe durchzuführen. Gleiches gilt auch für eine Schwachstelle in MediaWiki (CVE-2019-15150).

Aber auch größere Anbieter wie z. B. Google treffen bei der Implementierung von SSO auf Schwierigkeiten. Googles eigene OAuth-Bibliothek, um per Java mit den OAuth-Diensten von Google zu kommunizieren, enthielt eine schwerwiegende Schwachstelle (CVSS 3.1 Score: 9,1), die vor allem die Anmeldung über native Anwendungen stark gefährdete (CVE-2020-7692). Schwachstellen, die in Bibliotheken der IdP-Anbieter selbst existieren, sind hierbei besonders gefährlich, da von einer weiten Verbreitung ausgegangen werden kann. Somit war sehr wahrscheinlich eine Vielzahl von Nutzern betroffen.

Eine weitere schwerwiegende Sicherheitslücke wurde im März 2020 veröffentlicht und betraf die Anmeldedienste von Apple [11]. Hierbei konnte sich ein Angreifer Zugang zu verschiedenen Diensten verschaffen, wenn er nur die bei den Diensten hinterlegte E-Mail des Opfers kannte. Für diese Schwachstelle wurden 100.000 $ im Rahmen des Bug Bounty Programms von Apple ausgezahlt. Da sich die Schwachstelle direkt bei den Authentifizierungsdiensten von Apple befand, waren potenziell alle Dienste betroffen, die den Login per Apple anboten.

Dieses Paper soll deshalb bekannte sicherheitsrelevante Bereiche der genannten Protokolle und Schwierigkeiten bei der Implementierung zusammenfassen, sowie bei der sicheren Integration von SSO unterstützen. Dafür werden bekannte Probleme aufgezeigt, Gegenmaßnahmen vermittelt und Best Practices dargestellt.

In Kapitel 2 wird dazu auf die grundlegende Funktionsweise von SSO eingegangen und es werden die Gemeinsamkeiten bzw. Unterschiede der beiden Protokolle OAuth und OIDC beleuchtet. Kapitel 3 stellt existente Schwachstellen auf der Protokollebene dar und zeigt auf, wie konkrete Angriffe auf das Protokoll stattfinden können. Kapitel 4 beschäftigt sich mit den gängigen Security Best Practices und wie diese umgesetzt werden. In Kapitel 5 wird dann der aktuelle Stand der Forschung betrachtet, bevor in Kapitel 6 ein Fazit gezogen und ein Ausblick gegeben wird.

2. Single Sign-On Protokolle

2.1. Erläuterung der Funktionsweise

Protokoll-Flow	OAuth 2	OIDC	Einsatzzweck
Code-Flow	X	X	Vertrauenswürdige Anwendungen Alle Anwendungen (mit PKCE)
Implicit-Flow	X	X	Browser-basierte Anwendungen (veraltet)
Hybrid-Flow	n/a	X	Hybride Anwendungen

Tabelle 1: Übersicht der verschiedenen SSO-Flows

In OAuth 2.0 und OIDC gibt es, je nach Anwendungsfall, verschiedene Protokoll-Flows. Da OIDC auf OAuth 2.0 basiert, sind diese recht ähnlich. In Tabelle 1 finden sich die wichtigsten Flows im Vergleich. Diese haben verschiedene Zielsetzungen und kommen daher in unterschiedlichen Szenarien zum Einsatz. Zur näheren Erklärung wird der Code-Flow herangezogen, welcher in Abbildung 1 dargestellt ist. Die Werte in Klammern stehen für in den HTTP-Nachrichten mitgesendete Parameter. Sind diese zusätzlich in eckigen Klammern aufgeführt, handelt es sich dabei um optionale Parameter. Die orangefarbenen Kommunikationspfade stellen Kommunikationen im Backchannel dar. Diese finden außerhalb des Zugriffsbereichs des Nutzers statt, da sie nicht über den Browser laufen. Der Ablauf wird im Folgenden skizziert.

1. Der Prozess wird durch die Anfrage einer geschützten Ressource am SP gestartet (Schritt 1).
2. Der SP liefert als Antwort einen Authorization Request in einem HTTP-Redirect zum IdP des entsprechenden Anbieters zurück (Schritt 1.1).
3. Der Browser leitet den Authorization Request an den IdP weiter (Schritt 2). In dem Request sind/können die folgenden Parameter enthalten sein:
   1.1 *response_type*: Die Art des Protokoll-Flows (code / implicit / hybrid)
   1.2 *client_id*: Eine ID, mit der sich der SP bei dem IdP vorab registriert hat
   1.3 *redirect_uri*: Das Weiterleitungsziel nach der Authentifizierung
   1.4 *scope*: Anzufragende Informationen über den Nutzer.

1.5 *state*: Zufälliger Wert, welcher zum Verhindern von CSRF-Angriffen dient (vgl. Abschnitt 3.1)

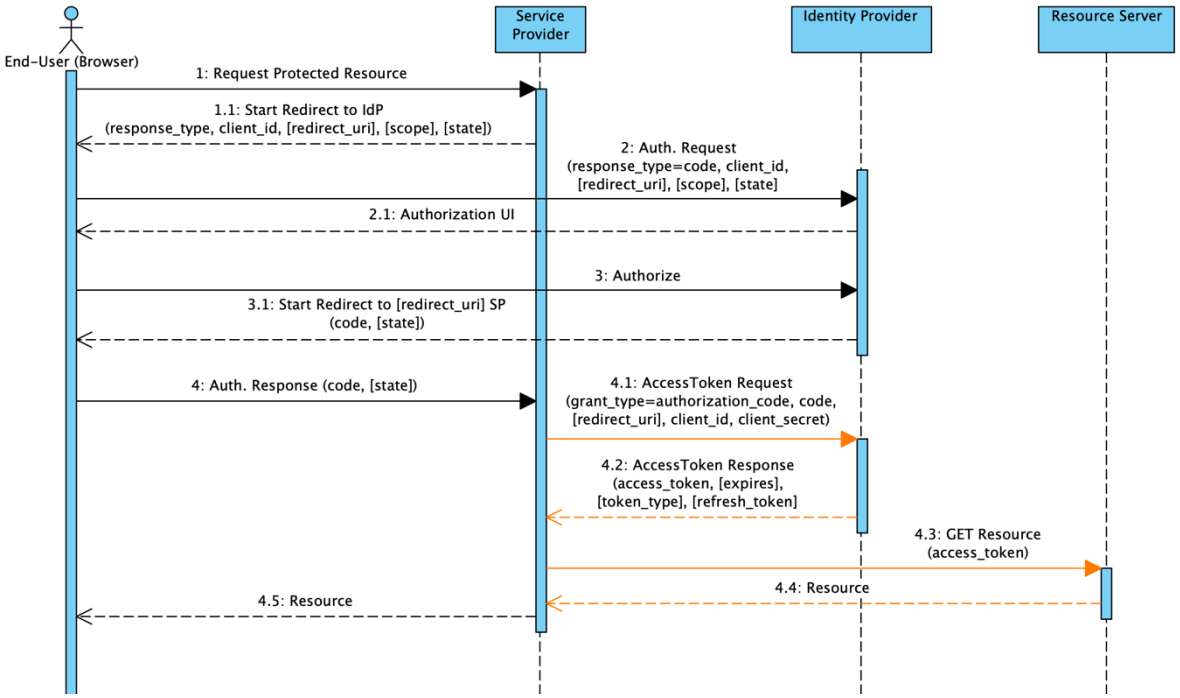


Abbildung 1: OAuth 2.0 - Code-Flow

4. Findet die Anmeldung für einen bestimmten SP erstmalig statt, muss der Nutzer den Zugriff auf seine Daten durch den SP autorisieren (Schritte 2.1 und 3). Dabei werden ihm auf einer Consent-Page die Informationen angezeigt, die für den SP freigeben werden sollen. Diese kann der IdP aus den Werten des Parameters *scope* extrahieren.
5. Nach erfolgreicher Anmeldung wird der Nutzer über den Browser vom IdP wieder an den SP geleitet (Schritt 3.1), wodurch der SP einen authorization code erhält (Schritt 4).
6. Mit dem erhaltenen Code kann der SP nun einen Access-Token Request an den IdP senden (Schritt 4.1). Mittels *client_id* und *client_secret* kann er sich zudem gegenüber dem IdP als berechtigter SP authentifizieren.
7. Der IdP validiert den Request und sendet im Erfolgsfall einen Access-Token an den SP (Schritt 4.2).
8. Der SP kann mit dem Access-Token eine Anfrage für die geschützte Ressource stellen (Schritt 4.3). Der Resource Server liefert bei erfolgreicher Validierung des Access-Token, diese an den SP zurück (Schritt 4.4).
9. Die Ressource kann der SP dann z. B. an den Browser weiterleiten oder auf deren Grundlage entscheiden, andere Daten an den Nutzer auszuliefern (Schritt 4.5).

Bei diesem Flow muss sich der SP mit Hilfe der *client_id* und des *client_secrets* beim IdP authentifizieren (Schritt 4.1). Nur wenn die im Auth. Request übermittelte *client_id*

zu dem sich authentifizierenden SP passt, erhält er einen Access-Token für den Zugriff auf die Ressourcen des Nutzers. Dies kann nur funktionieren, wenn das Client-Secret vor unberechtigtem Zugriff geschützt werden kann – beispielsweise innerhalb einer Web-Applikation, bei der ein Webserver als SP fungiert. Dieser kann das Client-Secret sicher speichern, ohne dass ein Nutzer der Web-Applikation darauf Zugriff erlangen kann. Bei einer nativen App müsste das Client-Secret allerdings auf jedem ausführenden Gerät gespeichert werden, da keine Backchannel-Komponente existiert, über die der Flow abgewickelt werden kann. Ein Angreifer könnte dann diese Zugangsdaten auslesen und sich damit beim IdP als Client ausgeben. Daher kann der Code-Flow in dieser Form nur dann sicher verwendet werden, wenn die Anwendung (SP) auf einer vertrauenswürdigen Infrastruktur ausgeführt wird und das client_secret vor unberechtigtem Zugriff geschützt werden kann.

Als Alternative zum Code-Flow wurde der Implicit-Flow in OAuth 2.0 integriert. Dieser sollte den Autorisierungsprozess vereinfachen und Anwendungen ohne Backchannel-Komponente eine einfache Möglichkeit der Autorisierung geben. Besonders für Browser-basierte Applikationen, wie Single Page JavaScript-Anwendungen, war dieser Flow gedacht. Bei der Verwendung des Implicit-Flows entfällt die Notwendigkeit, dass der SP einen Code gegen einen Access-Token einlösen muss. Stattdessen wird der Access-Token direkt im Frontchannel – in diesem Beispiel dem Browser des Endnutzers – verarbeitet. Bei Schritt 3.1 wird daher statt dem Code direkt der Access-Token zurückgeliefert, ohne dass eine Prüfung über ein client_secret erfolgt. Mit diesem Access-Token kann dann die geschützte Ressource abgerufen werden. Dadurch entfallen die Schritte 4.1 sowie 4.2, und Schritt 4.3 wird direkt von dem Browser durchgeführt. Bei diesem Vorgehen existieren einige Sicherheitsprobleme, weshalb davon abgeraten wird, diesen Flow weiter zu nutzen [12]. In Kapitel 3 wird hierauf genauer eingegangen.

Der Hybrid-Flow existiert einzig bei OIDC und vereint den Code-Flow mit dem Implicit-Flow. Dieser kann beispielsweise für hybride Apps[3] nützlich sein. Tokens, wie beispielsweise der Access-Token, werden hierbei sowohl im Back- als auch Frontchannel verarbeitet.

Festgelegt werden die verschiedenen Abläufe mit dem Wert des Parameters *response_type*, welcher im Authorization Request in Schritt 1.1 gesetzt wird. Die möglichen Werte mit der Zugehörigkeit der Abläufe sind in Tabelle 2 aufgelistet. Diese stehen in direktem Bezug zu den geforderten Rückgabewerten. Je nach Wert wird nach erfolgreicher Autorisierung ein Code, Access-Token und / oder ID-Token an den SP vom IdP in Schritt 3.1 zurückgeliefert.

- **Code**: Der Code ist ein vom IdP zurückgesendeter Wert, welcher vom SP beim IdP gegen einen Access- und/oder ID-Token eingelöst werden kann.

[3] Hybride Apps kombinieren native Anwendungen und Web-Applikationen. Sie basieren auf Web-Technologien und laufen in einem Web-View-Container. Gleichzeitig können Sie auf die Schnittstellen der Betriebssysteme zugreifen.

- **Access-Token**: Mit dem Access-Token erhält man Zugriff auf eine geschützte Ressource. Wird der Wert des Parameters *response_type* auf „token“ gesetzt, erhält der Browser des Nutzers direkt den Access-Token und kann diesen einlösen. Eine Zuordnung an einen bestimmten Client ist dann nicht mehr möglich, da der Access-Token ohne Zuordnung an einen beliebigen Nutzer ausgegeben wird.
- **ID-Token**: Der ID-Token enthält spezifische Informationen zu dem angemeldeten Benutzer, wie z. B. einen eindeutigen Identifier. Er existiert nur bei OIDC und erweitert die Autorisierung (OAuth) um eine Authentifizierung (OIDC). Beim Code-Flow wird dieser zusammen mit dem Access-Token im Backchannel ausgeliefert. Beim Implicit-Flow wird dieser direkt an den Browser ausgeliefert und beim Hybrid-Flow erhalten beide – sowohl Front- als auch Backchannel – einen ID-Token.

	OAuth 2	OIDC
Code-Flow	code	code
Implicit-Flow	token	id_token id_token token
Hybrid-Flow	n/a	code token code id_token code token id_token

Tabelle 2: Rückgabewerte der SSO-Protokollabläufe

2.2. Gemeinsamkeiten und Unterschiede

OIDC dient als Authentifizierungsschicht und baut auf OAuth 2.0 auf [6]. Daher sind die Abläufe der Protokollvorgänge größtenteils identisch. Charakteristisch für OIDC ist der ID-Token, welcher zusätzlich zum Access-Token zur Authentifizierung des Nutzers mitgesendet wird. Es handelt sich dabei um ein Base64 encodierter JSON Web Token (JWT) [13], der u. a. weiterführende Informationen zur Identität des Nutzers beinhaltet. Im Beispiel des Code-Flows (Abb. 1) wird dieser Token in Schritt 4.2 zusätzlich zum Access-Token zurückgegeben. Während OAuth 2.0 sich somit um die Autorisierung kümmert, liegt der Nutzen von OIDC in der Authentifizierung. Daher werden statt der Begriffe Authorization Request und Authorization Response bei OIDC die Begriffe Authentication Request und Authentication Response genutzt. Die Verwendung von OIDC lässt sich anhand des Parameters *scope* erkennen. Soll OIDC verwendet werden, muss dort der Wert „openid“ angegeben werden. Auf technischer Seite erweitert OIDC den OAuth 2.0 Standard um weitere Abläufe und Prozesse wie z. B. den Hybrid Flow.

3. Bekannte Sicherheitsprobleme

3.1. Protokollebene

Die vorgestellten Protokollabläufe bieten verschiedene optionale Parameter. Dadurch wird dem Entwickler eine große Freiheit bei der Implementierung der SSO-Protokolle

eingeräumt. Doch aus großer Flexibilität folgt eine große Verantwortung und Komplexität. Einige der Parameter beeinflussen direkt die Sicherheit von SSO. Ein wichtiges Beispiel hierfür ist der Parameter *state*. In der OAuth 2.0 Spezifikation wird dieser als empfohlen angegeben und wie folgt beschrieben:

> „*An opaque value used by the client to maintain state between the request and callback. […] The parameter SHOULD be used for preventing cross-site request forgery […]*" [5], Abschnitt 4.1.1.

Der SP generiert hierfür einen zufälligen und nicht vorhersehbaren Wert, welcher nur einmalig verwendet werden darf und an die User-Session gebunden ist. Diesen sendet er in der Authorization Request mit. Nach erfolgreicher Anmeldung integriert der IdP diesen in der Antwort – der Authorization Response. Der SP kann beim Erhalt der Antwort, die per HTTP-Redirect über den Browser des Nutzers gesendet wird, überprüfen, ob die Werte übereinstimmen. Ist dies der Fall und passen die User-Sessions zueinander, wurden beiden Anfragen vom gleichen User gesendet und gehören zusammen. Somit bindet der state-Parameter die Authorization Response an den Authorization Request. Da die Spezifikation diesen aber nicht verpflichtend verlangt, können SPs den Parameter bei der Authorization Request weglassen. Dadurch kann die Authorization Response nicht mehr eindeutig zugeordnet werden und CSRF-Angriffe werden möglich.

Ein weiterer Parameter, der die Sicherheit der SSO-Implementierung direkt beeinflusst ist der Parameter *nonce*. Dieser existiert nur in der OIDC-Spezifikation und kann ebenfalls bei dem Authentication Request wie der Parameter *state* als optionaler Parameter angegeben werden. Der Einsatzzweck ist die Bindung des ID-Tokens an die User-Session. Zusätzlich können damit auch Replay-Angriffe verhindert werden. Die vom SP generierte Nonce muss vom IdP in den zurückgegebenen ID-Token integriert werden. Beim Erhalt des ID-Tokens überprüft der SP, ob:

1. die User-Session, an welche der Wert der Nonce gebunden wurde, zu der User-Session passt, die den ID-Token übermittelt.
2. die Nonce von ihm generiert und noch nie genutzt wurde.

Nur wenn beide Punkte zutreffen, ist der Token valide. Aber auch hier kann die Entscheidung, den Parameter als optional zu definieren, zu genau dem Angriff führen, den der Parameter verhindern soll. Die Parameter *state* und *nonce* bringen einige Gemeinsamkeiten mit sich. Zum besseren Verständnis werden die Einsatzzwecke in Tabelle 2 dargestellt.

State	Nonce
- Schutz vor CSRF-Angriffen durch die Bindung des Autorisierungs- bzw. Authentifizierungsprozess an die User-Session - Wird im Auth. Request angegeben und in der Auth. Response vom IdP zurückgeliefert	- Schutz der Validität des ID-Tokens durch die Bindung des ID-Tokens an die User-Session - Wird in dem Auth. Request angegeben und vom IdP im erstellten ID-Token zurückgeliefert - Kann zusätzlich zum Schutz vor CSRF-Angriffen genutzt werden

Tabelle 3: Parameter state und nonce im Vergleich

Neben Parametern beeinflussen aber auch die Abläufe selbst die Sicherheit der SSO-Protokolle, wie z. B. der Implicit-Flow. Dieser Flow sendet Access- und ID-Tokens direkt in der Auth. Response mit. Da somit die sicherheitsrelevanten Tokens über den Browser des Nutzers geleitet werden, ist der Flow anfällig für Token Leakage. Eine Bindung des Access-Tokens an einen bestimmten Client – beispielsweise über die Kombination aus Client-Id und Client-Secret – wie im Code-Flow, ist ebenfalls nicht mehr möglich, da bei diesem Flow keine Authentifizierung des SP am IdP zum Erhalt eines Access-Tokens notwendig ist. Dies führt dazu, dass z. B. der Access-Token von verschiedenen Clients eingelöst werden kann, ohne dass die Identität des SPs sichergestellt ist.

Aber auch der Code-Flow kann unter gewissen Umständen nicht für die Sicherheit der Access-Tokens garantieren. Bei OAuth 2.0 gibt es zwei verschiedene Client-Anwendungs-Typen: Confidential Client und Public Client. Der Confidential Client – wie in dem Beispielablauf aus Abschnitt 2.1 skizziert – ist in der Lage sich gegenüber dem IdP auf sichere Weise zu authentifizieren. Hierzu verwendet dieser eine Client-Id und ein Client-Secret. Beides kann der SP sicher speichern und vor unberechtigtem Zugriff schützen. Die Ausführung eines Authorization Requests auf Code-Flow Basis ist daher sicher möglich. Dagegen sind Public Clients wie z. B. native Applikationen oder SPAs nicht in der Lage, ein Client-Secret sicher vor unbefugtem Zugriff zu schützen, da sie in der nicht vertrauenswürdigen Umgebung des Endnutzers laufen. Somit werden diese anfällig für Authorization Code Interception-Angriffe. Der zurückgelieferte Code aus der Authorization Response kann dann unberechtigt von einem Angreifer beim IdP gegen einen Access-Token eingelöst werden. Um dies zu erreichen, muss der Angreifer:

1. Zugriff auf Client-Id und Client-Secret des SPs besitzen
2. den Code der Authorization Response abfangen
3. in der Lage sein, eine eigene gültige OAuth 2.0-Applikation bei dem Client zu installieren, die sich als originale Applikation ausgibt, um den Code einzulösen.

Der Prozess ist in Abbildung 2 dargestellt. Da dies nur bei Public Clients der Fall ist und Confidential Clients bereits durch den Schutz des Client-Secrets geschützt sind, kann die Attacke nur bei Public Clients durchgeführt werden.

Neben den aufgelisteten Flows aus Abschnitt 2.1 existieren noch weitere Flows. Da die Darstellung aller Protokollabläufe in diesem Paper zu weit führen würde, soll an dieser Stelle nur auf einen weiteren Flow eingegangen werden, welcher aufgrund von Sicherheitsproblemen nicht weiter genutzt werden soll – der „Resource Owner Password Credentials Grant“. Dieser kommt dann zum Einsatz, wenn der Resource Owner dem SP vollständig vertraut – beispielsweise, wenn IdP und SP von demselben Betreiber gehostet werden. Beim „Resource Owner Password Credentials Grant“ sind dem SP die Zugangsdaten eines Nutzers für den IdP bekannt. Mit diesen kann sich der SP dann am IdP anmelden und erhält einen Access-Token für den Nutzer. Da hier Nutzername und Passwort vom SP direkt übermittelt werden müssen, existieren mehr potenzielle Angriffsmöglichkeiten, und die Wahrscheinlichkeit wird erhöht, dass die Zugangsdaten mitge

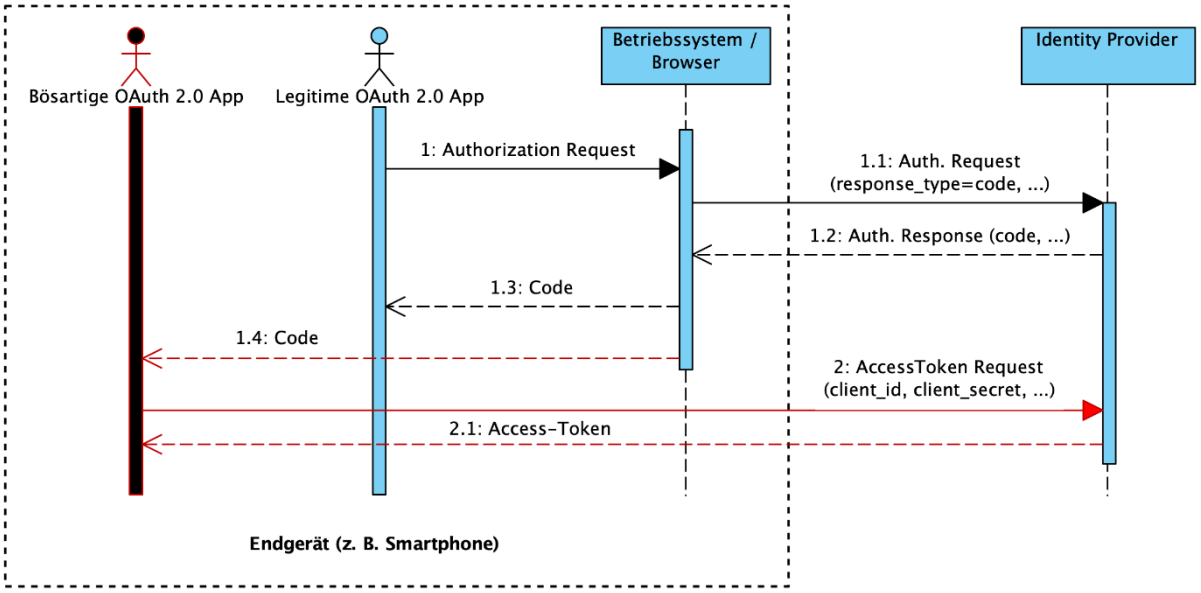


Abbildung 2: Code Injection Angriff bei Public Clients

lesen werden können. Sollte dieser Flow verwendet werden und der SP nicht vertrauenswürdig sein, besteht zudem die Gefahr, dass die Zugangsdaten missbraucht werden, um an weitere Informationen des Resource Owners zu gelangen.

Zu guter Letzt stellen die Weiterleitungen selbst eine Gefahr bei den SSO-Protokollen dar. Sobald ein SP einen Auth. Request sendet, kann er eine *redirect_uri* definieren. Ist dieser Parameter unzureichend geschützt, kann ein Angreifer dafür sorgen, dass der ausgestellte Code, Access-Token oder ID-Token nicht an den vorgesehenen SP weitergeleitet wird, sondern an eine vom Angreifer kontrollierte Domain. Eine Änderung des Auth. Requests ist u. a. dadurch möglich, dass dieser per HTTP-Redirect über den Endnutzer an den IdP gesendet wird. Diese Kommunikation kann mitgelesen, abgefangen und geändert werden. Zudem kann der Angreifer selbst einen Protokollablauf starten.

Ist die *redirect_uri* gut geschützt, kann auch eine Weiterleitung auf Ebene des SPs die Sicherheit aushebeln. Wenn dieser z. B. einen Parameter definiert, um den Nutzer nach erfolgreicher Authentifizierung auf die anfänglich besuchte Seite weiterzuleiten, kann es sich um einen sogenannten Open Redirect[4] handeln. Sobald ein Angreifer diesen kontrollieren kann, können Parameter wie z. B. der Code oder Access-Token bei dem anschließenden Redirect auf das vom Angreifer spezifizierte Weiterleitungsziel aus dem HTTP Referer-Header geleaked werden.

3.2. Angriffe auf die Protokolle

Die aufgezeigten Schwachstellen aus Abschnitt 3.1 können bei Nichtbeachtung oder falscher Umsetzung zu verschiedenen Sicherheitsproblemen führen. Zuerst soll der Parameter *state* näher betrachtet werden. Sollte ein SP den SSO-Protokollablauf auch ohne diesen unterstützen, kann ein Angreifer einen CSRF-Angriff durchführen. Gleiches gilt,

[4] siehe auch OWASP Top Ten 2013 – „A10 – Unvalidated Redirects and Forwards" (https://owasp.org/www-pdf-archive/OWASP_Top_10_2013_DE_Version_1_0.pdf)

wenn der Wert des state-Parameters vorhersehbar ist (z. B. ein statischer Wert). Um dies auszunutzen, startet ein Angreifer den Autorisierungsprozess an einem SP, fängt allerdings die Weiterleitung des SPs an den IdP ab (Authorization Request). Statt diesen selbst an den IdP zu senden, bringt er das Opfer dazu, den abgefangenen Request auszuführen (z. B. durch einen manipulierten Link). Da die manuelle Autorisierung (Schritt 2.1 / 3.) entfällt, wenn der Nutzer beim IdP für den entsprechenden SP bereits seine Zugriffserlaubnis erteilt hat, kann der Angreifer so das Opfer unbemerkt in den Account des Opfers anmelden. Des Weiteren besteht die Möglichkeit, dass der Angreifer ein Opfer unbemerkt in den Account des Angreifers anmeldet. Hierfür führt der Angreifer den Autorisierungsprozess bis zum Erhalt einer validen Auth. Response durch (Schritt 3.1). Diese kann nun erneut dem Opfer untergeschoben werden und bei fehlender oder fehlerhafter Validierung des State-Parameters ist das Opfer im Account des Angreifers eingeloggt. Besonders unter Berücksichtigung von weiteren Schwachstellen, können sich hieraus verschiedenste Szenarien ergeben, in denen das Opfer Aktionen in seinem Account oder dem des Angreifers ausführt, welche die Schutzziele kompromittieren können.

Der nonce-Parameter schützt den SP bei der Durchführung von OIDC-Abläufen vor Replay-Attacken. Er bindet dafür den ID-Token an eine User-Session. Sollte ein SP diesen Parameter allerdings nicht oder fehlerhaft validieren, ist eine Anmeldung des Angreifers im Namen des Opfers an dem jeweiligen SP möglich. Hierzu muss der Angreifer zuerst in den Besitz eines ID-Tokens des Opfers kommen. Dieser kann dann während des Authentifizierungsprozesses ausgetauscht werden. Am besten funktioniert dies bei Anwendungen, die den Implicit-Flow unterstützen. Um den Angriff durchzuführen, fängt der Angreifer die Authentication Response des IdPs an den SP ab, welche über seinen Browser weitergeleitet wird. In dieser tauscht er nun den eignen ID-Token gegen den gestohlenen ID-Token aus. Der SP arbeitet durch die fehlerhafte Prüfung der nonce nun mit den im gestohlenen ID-Token enthaltenen Informationen weiter und gewährt dem Angreifer z. B. Zugriff auf den Account des Opfers. Wäre der nonce-Parameter richtig validiert worden, wäre beim Austausch bereits aufgefallen, dass der ID-Token nicht zu der aktuellen Session passt bzw. dieser bereits beim SP eingelöst wurde.

Als nächstes werden die verschiedenen Protokollabläufe von SSO betrachtet. Der Implicit-Flow wurde bereits als veraltet definiert. Daher soll dieser, beziehungsweise jeder Flow, der den Access-Token direkt in der Authorization Response integriert, nicht verwendet werden (vgl. Abschnitt 2.1). Ausnahmen gelten nur dann, wenn der Client aktiv verhindert, dass gestohlene Access-Tokens ausgetauscht werden können und alle Möglichkeiten für Access-Token Leakage beseitigt werden. Von einer Verwendung dieses Flows wird aber generell abgeraten, da die existierenden Bedrohungen zu vielfältig sind. Vor allem die direkte Auslieferung des Access- und ID-Tokens an den Nutzer ist problematisch, da verschiedenen Methoden existieren, um die Tokens zu stehlen und zu missbrauchen. Hierfür eignen sich beispielsweise Cross-Site Scripting-Schwachstellen, über die ein Access-Token aus dem Browser des Opfers extrahiert werden kann oder auch Redirect-Angriffe, bei denen der Token über den HTTP Referer-Header geleaked wird. Da bei diesem Flow keine Möglichkeit besteht, den Access-Token an die Client-Applikation zu binden, kann der gestohlene Access-Token dann von einer beliebigen

Anwendung eingelöst werden. Wird OIDC verwendet, kann der Angreifer zusätzlich Zugriff auf den ID-Token erlangen. Somit ist es ein leichtes, an Informationen und Daten des Opfers zu kommen und mit geklauten Access-Tokens auf Ressourcen des Users zuzugreifen.

Als weiterer Flow sind die Gefahren des „Resource Owner Password Credentials Grant" offensichtlich. Sollte dessen Nutzung möglich sein, kann ein Angreifer, der Zugriff auf den SP besitzt, nicht nur Informationen und Daten über ein Opfer in Erfahrung bringen, sondern im schlimmsten Fall den kompletten Account des Opfers übernehmen. Den Zugriff kann sich der Angreifer über weitere Schwachstellen verschaffen oder auch ganz legitim erhalten haben, z. B., weil er die Anwendung administriert. Ist dies der Fall, können Nutzernamen und Passwort im Klartext beim Versand an den IdP oder der Eingabe am SP mitgelesen werden.

Einen weiteren Angriffspunkt auf Protokollebene stellt der Parameter *redirect_uri* dar. Sollte die Weiterleitung durch korrekte Validierung nicht richtig geschützt werden, kann ein Angreifer das Ziel, an das ausgestellte Codes, Access- und/oder ID-Tokens zurückgeliefert werden, beliebig abändern. Da diese Überprüfung von dem IdP durchgeführt wird, ist ein Angreifer dabei auf eine fehlerhafte Verifizierung des IdPs angewiesen. Das würde bedeuten, dass Firmen wie Google oder Facebook den Parameter *redirect_uri* fehlerhaft gegen vorher festgelegte URLs der registrierten Client-Anwendungen prüfen müssten. Wahrscheinlicher ist es dagegen eine Redirect-Schwachstelle beim SP (Open Redirect) zu finden. Diese tritt z. B. dann auf, wenn ein SP den Nutzer nach erfolgreicher Anmeldung wieder zu der Seite weiterleiten möchte, auf die ursprünglich zugegriffen wurde (Schritt 1, Abbildung. 1). Ist diese Information in einem Parameter enthalten, der auch an den IdP gesendet wird, kann der Angreifer einen Redirect-Angriff durchführen. Ein verwundbarer HTTP Authorization Request dafür könnte wie folgt aussehen:

> https://accounts.google.com/o/oauth2/v2/auth?client_id=exampleclient&
> response_type=token&redirect_uri=auth.example.com&
> state=randomString%2B**url%3Dattacker.com**

Hierbei wurde die Seite, von der der Nutzer die Anmeldung gestartet hat, an den Wert des Parameters *state* angehängt. Durch dessen Manipulation wird der Nutzer nach erfolgreicher Autorisierung nun an attacker.com weitergeleitet. Wird clientseitig keine oder nur eine fehlerhafte Validierung solcher Weiterleitungen durchgeführt, kann das Opfer an eine vom Angreifer kontrollierte Domain weitergeleitet werden. Access-, ID-Tokens und/oder Codes können dann auf dem Angreiferserver aus dem HTTP Referer-Header ausgelesen werden.

4. Security Best Practices

4.1. Zusammenfassung der Security Best Practices

Für OAuth 2.0 wurden aktuelle Security Best Practices bereits in einem Dokument zusammengefasst [12]. In diesem werden unter anderem auch die in Abschnitt 3.1 aufgelisteten Sicherheitsprobleme adressiert. Wichtige Punkte sind dabei der Schutz des Redirect-basierten Ablaufs, die richtige Nutzung der verschiedenen Flows, Maßnahmen

gegen Token Replay, die Einschränkung der Berechtigungen von Zugangstokens, sowie die Client Authentication. Da OIDC auf OAuth aufbaut, können diese Security Best Practices auch hierfür herangezogen werden. Zusätzlich werden in Abschnitt 16 der OIDC Spezifikation noch weiter Hinweise bezüglich Sicherheit von OIDC gegeben [6].

Für die Redirect-basierten Abläufe bei SSO gilt, dass der Schutz vor Manipulation entsprechender Redirect-URIs gegeben sein muss. Hier stehen auch die IdPs in der Pflicht. Bei der Registrierung von neuen Client-Anwendungen, müssen dazu valide URLs angegeben werden, an die eine Authorization Response weitergeleitet werden darf. Sollte diese nicht mit der URL des Parameters *redirect_uri* übereinstimmen, welche im Auth. Request mitgeliefert wurde, darf keine gültige Auth. Response gesendet werden. Für die SP gilt: Es dürfen keine offenen Redirects innerhalb der Parameter solcher Anfragen existieren. Daraus folgt, dass ein Angreifer keine Möglichkeit haben darf, den Nutzer nach einer erfolgreichen Autorisierung auf eine beliebige, von ihm kontrollierte (Sub-) Domain weiterzuleiten. Ebenfalls müssen die SP sicherstellen, dass Redirect-Antworten nur dann durchgeführt werden, wenn diese von genau dem IdP stammen, an den der entsprechende Request gesendet wurde und diese von der gleichen User-Session stammen. Damit können Angriffe verhindert werden, bei denen der Angreifer einen IdP kontrolliert (sogenannte Mix-Up Attacks).

Um den Schutz vor CSRF-Angriffen zu gewährleisten, muss der Parameter *state* korrekt genutzt werden. D. h. er muss vorhanden sein, und der Wert wird für jeden Protokoll-Durchlauf frisch und zufällig generiert. Eine Wiederverwendung wird ausgeschlossen. Ist state sicher implementiert, kann dadurch überprüft werden, ob die Authorization Response zu dem gesendeten Authorization Request gehört. In OIDC kann dafür auch der Parameter *nonce* genutzt werden.

Einer der wichtigsten Punkte der Security Best Practices ist die Entscheidung, den Implicit-Flow nicht (weiter) zu nutzen. Gleiches gilt für den „Resource Owner Password Credentials Grant“. Stattdessen sollen Flows verwendet werden, die den Access-Token im Backchannel bei der Token Response zurückliefern – z. B. der Code-Flow. Somit kann die Zugehörigkeit des Access-Tokens zu einem SP bzw. Client zugesichert werden. Dies geschieht, indem sich der SP am IdP authentifiziert. SPs sollen hierbei nach Möglichkeit asynchrone Verfahren zur Authentifizierung, wie z. B. „Mutual-TLS Client Authentication and Certificate-Bound Access Tokens“ [14] verwenden, da so der IdP keine sensiblen Informationen, wie das Client-Secret speichern muss.

Aber auch der Code-Flow hat für Public Clients, wie native oder Browser-basierte Applikationen spezielle Sicherheitsanforderungen. Solange die Client-Applikation nicht in der Lage ist ein Client-Secret sicher zu speichern, können gestohlene Codes weiterhin von einem unberechtigten Angreifer eingelöst werden. Daher wurde das „Proof Key for Code Exchange“ (PKCE) Verfahren entwickelt [15]. Dies soll es auch Public Clients möglich machen, auf sichere Art einen Code zu erhalten und diesen nur für eine bestimmte Anwendung einlösen zu können. Hierfür wird der Code-Flow um einen Parameter *code_verifier* ergänzt. Aus diesem wird eine *code_challenge* berechnet, welche in dem Auth. Request an den IdP mitgesendet wird. Die Berechnung wird mittels einer

code_challenge_method durchgeführt. Aktuell unterstützt die Spezifikation zwei Arten für die *code_challenge_method*:

- S256 (Base64-kodierter SHA256 Hash des *code _verifiers*)
- plain (Plain-Text des *code_verifiers*)

Die Methode plain darf nur verwendet werden, wenn der Client aufgrund von technischen Limitierungen z. B. nicht in der Lage ist, einen SHA256 Hash zu generieren. Nachdem der Client den Code in der Authorization Response erhalten hat, sendet er diesen in dem Access-Token Request zusammen mit dem originalen *code_verifier* an den IdP. Dieser kann anhand des generierten Werts überprüfen, ob die Anfrage des übermittelten Codes zu dem Client gehört, der aktuell dafür einen Access-Token einlösen möchte. Sollte dies nicht der Fall sein, wird die Anfrage abgebrochen. Das Vorgehen ist in Abbildung 3 skizziert. Mittels PKCE können zum einen Code Interception-Angriffe ausgehebelt und zum anderen nativen und Browser-basierten Applikationen eine Möglichkeit der Autorisierung und Authentifizierung mit erhöhtem Sicherheitsniveau gegeben werden. In der Spezifikation wird PKCE für alle Clients empfohlen, auch wenn das Verfahren initial für Public Clients vorgesehen war.

Der Access-Token kann bei OAuth als die wichtigste Information angesehen werden. Mit diesem kann ein Zugriff auf die Ressourcen des Resource-Owners ausgeführt werden. Dementsprechend sollten Access-Tokens ein besonders hohes Maß an Sicherheit erfüllen. Dies beginnt bereits bei der Vergabe der Berechtigungen, die an einen Token gebunden sind. Je höher die Rechte eines Tokens sind, umso schlimmer die Auswirkungen bei einem Missbrauch. Somit sollte hier immer das Sicherheits-Prinzip “Least Privilege” angewandt werden, um dem Access-Token nur die Rechte zuzuordnen, die für die Erfüllung der Aufgabe minimal notwendig sind. Um die Möglichkeit des Missbrauchs weiter zu verringern, sollte die Gültigkeitsdauer eines Tokens möglichst kurz gehalten werden. Ebenfalls sollten Access-Token nur für spezifische Resource-Server gültig sein. Bestenfalls nur für einen. Dies muss sowohl bei der Ausstellung des Tokens definiert als auch bei der Nutzung des Tokens durch den IdP verifiziert werden.

Zu guter Letzt sollte natürlich die komplette Kommunikation zwischen den beteiligten Parteien über TLS abgesichert werden.

4.2. Kosten-Nutzen Auswirkung bei der Umsetzung der Gegenmaßnahmen

Wie bei allen Anwendungen, müssen die Gefahren und der Nutzen der Maßnahmen gegeneinander abgewogen werden. Viele der aufgeführten Security Best Practices aus Abschnitt 4.1 sind relativ einfach umzusetzen und helfen damit, das Sicherheitsniveau deutlich zu erhöhen. Daher sollten diese Security Best Practices als Minimum der umzusetzenden Maßnahmen angesehen werden. Diese lassen sich nach Verantwortlichkeitsbereichen aufteilen. Es gibt Maßnahmen, die vom IdP umgesetzt werden müssen, aber noch mehr, die von den SPs adressiert werden sollten.

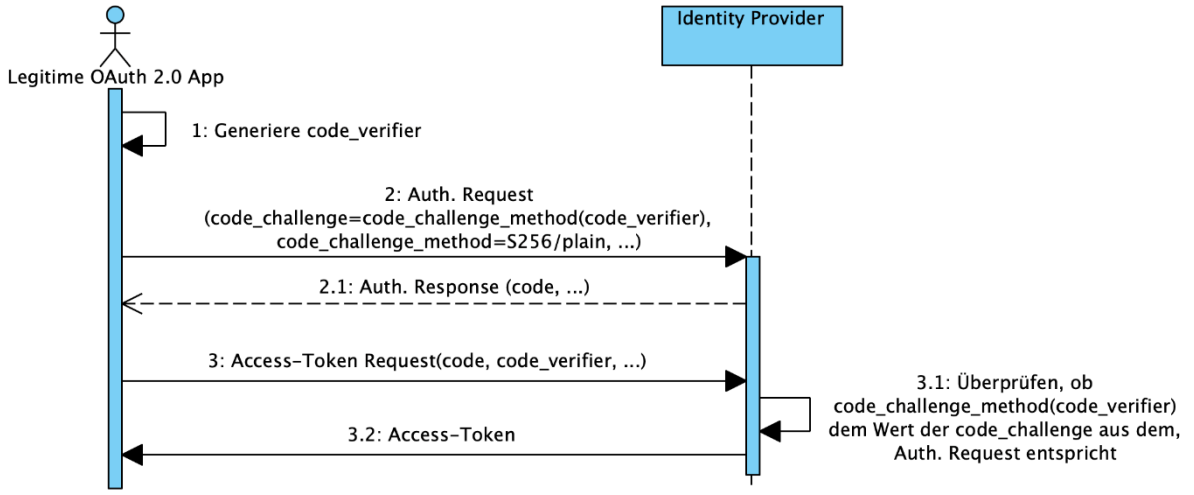


Abbildung 3: Ablauf des Code-Flows mit PKCE

Aus Sicht des IdP steht vor allem die saubere Implementierung des Protokolls im Vordergrund. Solange die Anfragen richtig verarbeitet werden, können die SPs einen sicheren Autorisierungs- und Authentifizierungsprozess aufsetzen. Sollten die IdPs allerdings Implementierungsfehler enthalten, kann auch der Client die Sicherheit nicht mehr gewährleisten. Hierzu zählen z. B. fehlende oder falsche Validierung der mitgesendeten Redirect-URIs aber auch die richtige und sichere Speicherung der Berechtigungen von Access-Tokens. Da hiervon viele SPs abhängen, sollten alle IdPs auf die sichere und korrekte Implementierung besonderen Wert legen. Solange dies gegeben ist, liegen die Risiken vorwiegend bei der Umsetzung auf SP-Ebene.

Zu den einfach umzusetzenden Maßnahmen auf Ebene der SPs gehört vor allem der Einsatz und die richtige Validierung von bereits vorgesehenen Security-Parametern. Durch die Beachtung der Parameter *state* und *nonce*, können CSRF- und Replay-Angriffe effektiv verhindert werden. Hierfür ist der Aufwand der Implementierung gering, da es sich hauptsächlich um das Generieren und das Prüfen von zufälligen Werten handelt.

Open Redirects stellen eine weitere gut adressierbare Gefahr dar. Das Hauptproblem besteht hierbei vor allem im Aufspüren solcher Weiterleitungen. Ein guter technischer Gesamtüberblick der SP-Anwendung kann dabei helfen. Wenn dies erst einmal geschehen ist, können verwundbare Prozesse umgestellt werden und mit mäßigem Aufwand, wie z. B. dem Einführen von Whitelists für Weiterleitungen, abgesichert werden.

Der Umstieg von nativen oder Browser-basierten Anwendungen kann unter Umständen mit einem deutlich erhöhten Aufwand einhergehen. Sollte eine Anwendung noch den Implicit-Flow nutzen, bedeutet dies eine größere Anpassung. Der Prozess des Code-Flows mit Verwendung von PKCE ist komplexer und muss zunächst verstanden und dann richtig eingebunden werden. Befindet sich eine Anwendung bereits auf Basis des Code-Flows, begrenzt sich der Aufwand auf die Integration von PKCE und ist somit einfacher zu handhaben. Ähnliches gilt für den „Resource Owner Password Credentials

Grant“. Hier ist eine komplette Umstellung des Autorisierungs- und Authentifizierungsprozesses allerdings unumgänglich.

Festzuhalten ist, dass alle Maßnahmen einen deutlichen Gewinn an Sicherheit für die eigene Anwendung mit sich bringen. Maßnahmen, die einen hohen Aufwand erfordern, sind bereits durch ein großes Sicherheitsrisiko geprägt und sollten daher trotz allem umgesetzt werden. Die restlichen Maßnahmen sind einfacher umzusetzen, bringen aber trotzdem einen großen Gewinn an Sicherheit. Daher können die Security Best Practices durchaus als gutes Gesamtpaket betrachtet werden, welches in jeder OAuth 2.0- / OIDC-Applikation Anwendung finden sollte. Auch wenn SSO durch eine Bibliothek des IdP Betreibers realisiert wird, bedeutet das nicht, dass die Autorisierungs- und Authentifizierungsprozesse sicher sind. Diese Bibliotheken müssen in die vorhandene Codebasis eingepflegt und der Prozess konfiguriert werden. Beachtet man hierbei nicht die Sicherheit, kann die eigene Anwendung durch eine Fehlkonfiguration leicht trotzdem verwundbar sein. Beispielsweise können veraltete Prozesse wie der Implicit-Flow sowohl von Bibliotheken als auch IdPs immer noch unterstützt werden.

5. Stand der Forschung

Seit dem offiziellen Erscheinen von OAuth 2.0 im Oktober 2012 und OIDC im November 2014 wurden die Protokolle hinsichtlich ihrer Sicherheit mehrfach untersucht. 2014 stellten Yang und Manoharan bezüglich der Sicherheit von OAuth 2.0 fest, dass durch den Vereinfachungsschritt von OAuth 1.0 zu OAuth 2.0 einige Sicherheitskonzepte nicht mehr existierten [16]. Dadurch waren vermehrt Angriffe auf das Protokoll möglich. Auch Sun und Beznosov hatten bereits 2012 fehlende Sicherheitsmaßnahmen auf diese Vereinfachungen zurückführen können [17]. Li und Mitchell untersuchten 2014 die Sicherheit bei 60 chinesischen OAuth 2.0 Implementierung und fanden alleine bei 21 davon CSRF-Schwachstellen [18]. Die Autoren der Studie schätzen, dass viele Millionen Nutzer durch diese Schwachstelle verwundbar waren. Noch einschneidendere Ergebnisse lieferte die Studie derselben Autoren 2014, als sie die Sicherheit beim Einsatz der Google OIDC-Implementierung bei verschiedenen Clients untersuchten [19]. Bei den insgesamt 103 untersuchten Nutzungen, entdeckten sie weit verbreitet Schwachstellen, welche schwerwiegende sicherheitskritische Folgen haben können. Bei vielen konnte sich ein Angreifer als Opfer anmelden.

Die formale Analyse aus 2016 von Fett et al. zeigt, dass OAuth 2.0 durchaus sicher eingesetzt werden kann [20]. Allerdings funktioniert dies nicht out-of-the-box. Es müssen verschiedene Gegebenheiten erfüllt werden. Unter anderem die Einhaltung der Security Best Practices. Das gleiche gilt für OIDC, welches 2017 ebenfalls von Fett et al. einer formalen Analyse unterzogen wurde [21]. Solange Security Best Practices und Guidelines eingehalten werden, kann nach aktuellem Stand davon ausgegangen werden, dass ein sicherer Einsatz gewährleistet ist.

Dass die Integration von verbreiteten SSO-Bibliotheken nicht ohne weitere Sicherheitsmaßnahmen ausreicht, zeigten Mainka et al. mit ihrer Sicherheitsanalyse von OIDC [7]. Unter anderem untersuchten sie acht auf der offiziellen OIDC-Website empfohlene Referenzimplementierungen bezüglich verschiedener Angriffsmöglichkeiten. Davon

konnten nur zwei alle Verifizierungsschritte vollständig und sicher durchführen. Durch komplexe Angriffe, die von der Spezifikation nicht ausreichend abgedeckt werden, konnte keine der offiziellen Bibliotheken eine bedenkenlose sichere Umsetzung gewährleisten. Im Zuge dessen, wurde auch das Open Source Pentesting-Tool PrOfESSOS[5] entwickelt. Mit diesem können bekannte Angriffe auf SSO automatisiert getestet werden.

6. Fazit und Ausblick

Wie in Kapitel 5 dargestellt, existieren Untersuchungen, die bestätigen, dass der sichere Einsatz von OAuth 2.0 und OIDC möglich ist. Dem gegenüber stehen Studien, die zeigen, wie weit verbreitet Schwachstellen in Implementierungen sind. Dies kann durchaus auf das fehlende Sicherheitsverständnis der Protokolle bei Entwicklern zurückgeführt werden. Erschwerend kommt hinzu, dass unsichere Abläufe und Konfigurationen – wie der Implicit-Flow – standardkonform sind. Entwickler könnten deshalb davon ausgehen, dass deren Einsatz sicher und vor allem vertretbar ist, da diese so im Standard definiert wurden. Durch die Aufteilung des Standards in viele Dokumente (Definition, Security Best Practices, PKCE, etc.) wird dies zusätzlich erschwert. Nicht jeder Entwickler, der SSO in eine Anwendung integriert, ist zudem Experte auf dem Themengebiet der IT-Sicherheit. Die richtige Information zur sicheren Umsetzung zu finden kann aufgrund der Fülle der Dokumente problematisch sein. Wenn dann noch nicht alle Entwickler die komplette Spezifikation sowie die Security Best Practices lesen, bevor eine Integration von SSO-Protokollen erfolgt, resultieren hieraus schnell schwere Sicherheitslücken.

Erschwerend kommt hinzu, dass viele falsche Code-Beispiele im Internet existieren und diese vor dem Einsatz nicht immer genügend auf die Sicherheit überprüft werden. Stack Overflow – eins der wichtigsten Portale, um auf programmatische Fragen Antworten zu erhalten – besitzt z. B. einen sehr hohen Anteil an unsicheren Antworten [22]. Diese werden durch den Fragesteller häufig trotz allem akzeptiert oder sogar von den Usern am höchsten bewertet. Es ist durchaus möglich, dass daraus fehlerhafte Implementierungen und Konfigurationen im Produktiveinsatz entstehen.

Dieses Paper soll dabei helfen, das Bewusstsein für die Sicherheit bei SSO im Allgemeinen zu erhöhen und SSO – im speziellen OAuth 2.0 und OpenID Connect – sicher zu integrieren. Die gezeigten Schwachstellen sind zum Großteil auch auf andere populäre Standards, wie z. B. SAML übertragbar.

Durch die reine Integration von Bibliotheken können schnell schwerwiegende Sicherheitslücken in den eigenen Anwendungen entstehen. Unter Beachtung der Security Best Practices und Verständnis der zugrundeliegenden Protokollabläufe, kann durch den Einsatz von SSO ein deutlich höheres Sicherheitsniveau erreicht werden. Dies geht einher mit einer zusätzlich verbesserten Usability.

[5] https://github.com/RUB-NDS/PrOfESSOS

Literaturhinweise

[1] R. Morris and K. Thompson, “Password Security: A Case History,” *Commun. ACM*, vol. 22, no. 11, pp. 594–597, 1979, doi: 10.1145/359168.359172.

[2] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano, “The quest to replace passwords: A framework for comparative evaluation of web authentication schemes,” 2012, doi: 10.1109/SP.2012.44.

[3] J. Bonneau, “The science of guessing: Analyzing an anonymized corpus of 70 million passwords,” 2012, doi: 10.1109/SP.2012.49.

[4] Z. Li, W. He, D. Akhawe, and D. Song, “The Emperor’s new password manager: Security analysis of web-based password managers,” in *Proceedings of the 23rd USENIX Security Symposium*, 2014, pp. 465–479.

[5] D. Hardt, “RFC6749: The OAuth 2.0 Authorization Framework,” *IETF Standard*. 2012, [Online]. Available: https://tools.ietf.org/html/rfc6749 .

[6] N. Sakimura *et al.*, “OpenID Connect Core 1.0 incorporating errata set 1,” *OpenID Foundation*. 2014, [Online]. Available: http://openid.net/specs/openid-connect-core-1_0.html .

[7] C. Mainka, V. Mladenov, J. Schwenk, and T. Wich, “SoK: Single Sign-On Security - An Evaluation of OpenID Connect,” in *Proceedings - 2nd IEEE European Symposium on Security and Privacy, EuroS and P 2017*, Jun. 2017, pp. 251–266, doi: 10.1109/EuroSP.2017.32.

[8] A. Armando, R. Carbone, L. Compagna, J. Cuéllar, G. Pellegrino, and A. Sorniotti, “An authentication flaw in browser-based Single Sign-On protocols: Impact and remediations,” *Comput. Secur.*, vol. 33, pp. 41–58, 2013, doi: 10.1016/j.cose.2012.08.007.

[9] W. Zeller and E. W. Felten, “Cross-Site Request Forgeries : Exploitation and Prevention,” *New York Times*, 2008.

[10] Y. Zhou and D. Evans, “SSOScan: Automated testing of web applications for single sign-on vulnerabilities,” 2014.

[11] B. Jain, “Zero-day in Sign in with Apple,” *https://bhavukjain.com/blog/2020/05/30/zero-day-signin-with-apple/*

[12] T. Lodderstedt; J. Bradley; A. Labunets; D. Fett, “OAuth 2.0 Security Best Current Practice draft-ietf-oauth-security-topics-16,” *IETF Standard*, 2020. https://tools.ietf.org/html/draft-ietf-oauth-security-topics-16.

[13] M. Jones, J. Bradley, and N. Sakimura, “RFC7519: JSON Web Token (JWT),” 2015. [Online]. Available: https://tools.ietf.org/html/rfc7519.

[14] T. L. Sakimura, N, B. Campbell, J. Bradley, “RFC 8705: OAuth 2.0 Mutual-TLS Client Authentication and Certificate-Bound Access Tokens,” 2020. [Online]. Available: https://tools.ietf.org/html/rfc8705.

[15] N. Sakimura (Ed.), J. Bradley, and N. Agarwal, “RFC7636: Proof Key for Code Exchange by OAuth Public Clients,” 2015. [Online]. Available: https://tools.ietf.org/html/rfc7636.

[16] F. Yang and S. Manoharan, “A security analysis of the OAuth protocol,” in *IEEE Pacific RIM Conference on Communications, Computers, and Signal Processing - Proceedings*, 2013, pp. 271–276, doi: 10.1109/PACRIM.2013.6625487.

[17] S. T. Sun and K. Beznosov, “The devil is in the (implementation) details: An empirical analysis of OAuth SSO systems,” 2012, doi: 10.1145/2382196.2382238.

[18] W. Li and C. J. Mitchell, “Security issues in OAUTH 2.0 SSO implementations,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8783, pp. 529–541, 2014, doi: 10.1007/978-3-319-13257-0_34.

[19] W. Li and C. J. Mitchell, “Analysing the security of google’s implementation of openID connect,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9721, pp. 357–376, doi: 10.1007/978-3-319-40667-1_18.

[20] D. Fett, R. Küsters, and G. Schmitz, “A comprehensive formal security analysis of OAuth 2.0,” in *Proceedings of the ACM Conference on Computer and Communications Security*, 2016, vol. 24-28-October-2016, doi: 10.1145/2976749.2978385.

[21] D. Fett, R. Kusters, and G. Schmitz, “The Web SSO Standard OpenID Connect: In-depth Formal Security Analysis and Security Guidelines,” in *Proceedings - IEEE Computer Security Foundations Symposium*, 2017, pp. 189–202, doi: 10.1109/CSF.2017.20.

[22] M. Chen, F. Fischer, N. Meng, X. Wang, and J. Grossklags, “How Reliable is the Crowdsourced Knowledge of Security Implementation,” in *Proceedings - International Conference on Software Engineering*, 2019, vol. 2019-May, pp. 536–547, doi: 10.1109/ICSE.2019.00065.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

OPTIMOS – Trusted-Service-Management-System – eine Infrastruktur für sichere mobile Dienste

Jörg Breuer[1], Christian Stengel[1], Dr. Friedrich Tönsing[1], Robert Zastrau[1]

Kurzfassung:
In diesem Beitrag soll das Trusted-Service-Management-System (TSMS) als Infrastruktur für sichere mobile Dienste vorgestellt und beschrieben werden. Das TSMS wird u.a. Karten auf die Smartphones bringen, z.B. den Online-Ausweis, und somit den 9-Punkte-Plan für ein digitales Deutschland des CIO des Bundes Dr. Markus Richter aus Juli 2020 unterstützen.
Zu diesem Zweck wird zuerst die Notwendigkeit des Systems dargestellt und anschließend ein allgemeiner Überblick über das System gegeben. Die Ausführungen werden dabei in den Kontext der Aktivitäten gestellt, die zu dieser Infrastruktur geführt und beigetragen haben.
Die TSM Infrastruktur wurde in den Förderprojekten OPTIMOS1.0 und OPTIMOS2.0 rudimentär spezifiziert und pilotiert und fand Einsatz in den Projekten MobiDO des BSI sowie dem SDI-Projekt PeopleID.

Stichworte: Digitale Identitäten, eIDAS, Mobile Dienste, OPTIMOS, SE, Secure-Elements, Trusted-Service-Management-System, TSMS

1. Motivation

Digitalisierung – mit Sicherheit!

Im Zuge der Digitalisierung werden Dienstleistungen, Produkte, Verwaltungsleistungen und weitere Geschäftsabläufe vermehrt mittels Webseiten, in Apps, durch Sprachassistenten und Chatbots abgebildet. Neben verstärkter digitaler Abbildung verschiebt sich auch der Zugang zu diesen Angeboten zunehmend auf das mobile Endgerät der Konsumenten. Nutzerfreundlichkeit, schnelles Feedback und eine schnelllebige, zielgerichtete Kundenerfahrung stehen im Mittelpunkt.

Nebenbei speichern Anbieter alle Daten, die sie legal verarbeiten dürfen, und Nutzer ziehen Bequemlichkeit dem Schutz ihrer Daten vor. Wie ernst Anbieter den Schutz der Kundendaten nehmen, wie sicher der vielseits genutzte Facebook-Login eigentlich ist oder weitere Fragen zu Datenschutz und Datensicherheit, stellen sich nicht. Ganz im Gegenteil rücken diese stetig weiter an den Rand des Kundenbewusstseins, da sowohl die Anbieter als auch die Kunden die Sicherheit von Daten und digitalen Identitäten quasi als gegeben voraussetzen.

Zur Sicherung von Logins, digitalen Identitäten und anderen digitalen Assets stellt sich stets die gleiche Frage: Wo beginnt die Wurzel des Vertrauens?

Das eSE – die sichere Ablage für Schlüsselmaterial auf mobilen Endgeräten

Moderne Betriebssysteme mobiler Endgeräte bieten standardmäßig bereits verschiedene Möglichkeiten zur Verwaltung von Schlüsselmaterial. Meist werden diese über gut gepflegte Codebibliotheken in Form einer Key-Chain oder eines Key-Stores angeboten.

[1] Deutsche Telekom Security GmbH, Bonner Talweg 100, 53838 Bonn

Diese nativen Möglichkeiten der Schlüsselablage haben ihre Wurzeln des Vertrauens meist in einem sicheren Bereich des Prozessors, dem so genannten Trusted-Execution-Environment (TEE). Für die generische Ablage von Schlüsselmaterial im Rahmen üblicher Anwendungen ist diese Form der Sicherung oft ausreichend. Erhöht sich jedoch die Kritikalität der Anwendung oder der Schutzbedarf einer bestimmten digitalen Identität, ist die Sicherheit des TEE nicht mehr adäquat, da TEEs gegen eine Vielzahl von Angriffen nur unzureichend gehärtet sind.

Eine höherwertige Alternative zum TEE bieten die so genannten Secure-Elements (SE). Diese sind spezielle Sicherheitschips, die besonders gegen physische und logische Angriffe gehärtet und mit hohem Aufwand geprüft und zertifiziert wurden. Die Sicherheit eines solchen Chips ist folglich signifikant höher als die des TEE. Secure-Elements können außerdem in verschiedenen Formfaktoren vorliegen. So stellen beispielsweise Universal-Integrated-Circuit-Cards (UICCs) in den verschiedenen Ausprägungen als UICC, embedded UICC (eUICC) oder integrated UICC (iUCC) eine spezielle Subgruppe der Secure-Elements dar. Zudem gibt es Secure-Elements in verschiedenen Formfaktoren, z.B. als embedded Secure-Elements (eSE), die direkt im mobilen Endgerät verlötet wurden.

Herausforderungen zur Nutzung von eSEs

Trotz der theoretischen Alternative zum TEE werden Secure-Elements heute nur in den wenigsten Anwendungen genutzt. Die Gründe für diesen Sachverhalt sind vielfältig. So erfordert die Nutzung der sicheren Chips sehr spezifisches Know-how. Dies stellt für viele Anbieter mobiler Services eine Hürde dar, da das benötigte Wissen nur schwer am Markt zu finden ist. Weiterhin ist der Markt der verfügbaren Secure-Elements in mobilen Endgeräten stark fragmentiert. Dies bedingt einen unverhältnismäßig hohen Aufwand für einen potenziellen Service-Provider, wenn dieser eine angemessen große Reichweite an Endgeräten adressieren möchte, um sein Angebot wirtschaftlich erfolgreich bei Endkunden zu platzieren. Eine weitere Konsequenz der starken Fragmentierung ist die Notwendigkeit eines individuellen Vertragsverhältnisses mit jedem Partner, der dem Service-Provider ein Secure-Element zur Verfügung stellt. Für viele Service-Provider ist allein der organisatorische Aufwand zur Etablierung dieser Vertragsstrukturen zu hoch.

Herausforderung angenommen: OPTIMOS!

Im Rahmen der OPTIMOS-Förderprojekte wurde an der Verbesserung und Lösung der genannten Herausforderungen gearbeitet. Dabei wurde eine konkrete Lösung in Form einer diskriminierungsfreien, generischen Trusted-Service-Management-Infrastruktur gefunden. Mit der Einführung der Rolle eines zentralen Trusted-Service-Management-Systems (TSMS) benötigen die Service Provider kein spezielles SE-Know-how. Die vorherrschende Fragmentierung des Marktes wird sowohl technisch als auch wirtschaftlich im TSMS gebündelt. Service-Provider haben auf Basis dieses Lösungsansatzes nur noch einen Ansprechpartner und müssen zur Nutzung der sicheren Hardware in mobilen Endgeräten lediglich das TSMS über ein bereitgestelltes Software-Development-Kit (SDK) in ihre Anwendungen integrieren. Dadurch wird es ihnen ermöglicht, Secure-

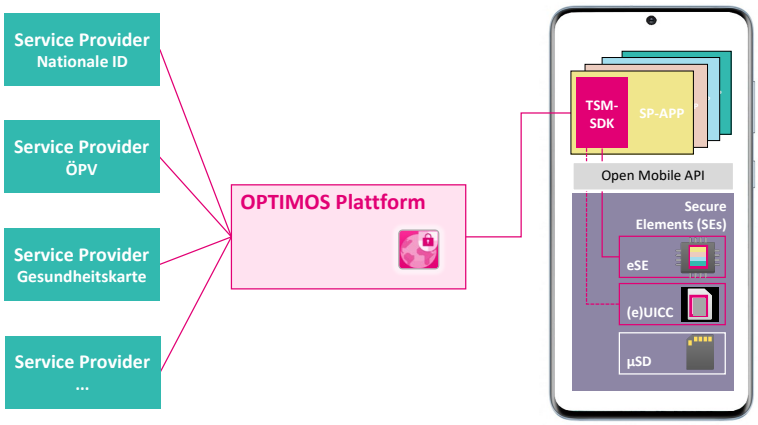


Abbildung 1: Das TSMS aus dem Projekt OPTIMOS2.0

Elements als hochwertige Wurzel des Vertrauens zur stärkeren Absicherung ihrer Anwendungen, ihrer Services und zugehöriger Kundendaten zu nutzen.

Die Aktivitäten in OPTIMOS2.0 werden in Kapitel 3 näher erläutert.

Exkurs: Wo lassen sich eIDs nutzen?

Politik und Wirtschaft haben die Bedeutung der diskriminierungsfreien, sicheren und datenschutzkonformen Nutzung von elektronischen Identitäten und der dazu erforderlichen Infrastrukturen erkannt. Beispiele hierfür sind die eIDAS-Verordnung, das Onlinezugangsverbesserungsgesetz (OZG), der Koalitionsvertrag der Bundesregierung sowie der 9-Punkte-Plan für ein digitales Deutschland des CIO des Bundes Dr. Markus Richter aus dem Juli 2020.

Die eIDAS-Verordnung der Europäischen Union regelt die Nutzung von elektronischen Identitäten in der EU und enthält technische Vorgaben zur Implementierung von skalierenden Vertrauensniveaus. Sie befindet sich derzeit in Überarbeitung und soll in Zukunft neben hoheitlichen auch private IDs umfassen, um eIDs in Europa und die angestrebte EUid auf eine breitere Grundlage zu stellen und „vertrauenswürdige Identitäten für alle“ zu schaffen. Dabei soll die Richtlinie nicht nur individuelle, sondern in Zukunft auch legale Identitäten wie IoT-Geräte umfassen. Zudem soll es jedem Bürger durch die neue Richtlinie möglich sein, sich mit seiner Identität nach einem Single-Sign-On-Ansatz auch pseudonymisiert zu identifizieren. Die Verwendung der Digitalen Identität soll dabei freiwillig sein.

Das deutsche OZG soll die Nutzung von Online-Diensten im eGovernment stärken. Es verpflichtet Bund und Länder dazu, ihre Verwaltungsleistungen bis 2022 auch elektronisch anzubieten. Außerdem müssen die Verwaltungsportale von Bund und Ländern zu einem Verbund miteinander verknüpft werden.

Der Koalitionsvertrag der aktuellen Bundesregierung hebt in besonderer Weise die Bedeutung von digitalen Infrastrukturen hervor und enthält Forderungen nach der proaktiven Umsetzung von IT-Sicherheit und Datenschutz gemäß den Prinzipien „security-by-design“, „privacy-by-design“ und „privacy-by-default“.

2. Warum Hardwaresicherheit

Im Rahmen von OPTIMOS wurde ein Trusted-Service-Management System zur Provisionierung von Daten auf hardwarebasierte Sicherheitselemente (SE) entworfen. Hierbei stellt sich die generelle Frage, warum hardwarebasierte Elemente softwarebasierten Lösungen vorzuziehen sind.

2.1 Einhaltung gesetzlicher Vorschriften

Nur die Verwendung hardwarebasierter SEs unterstützt eine einfache Lösung für mobile eID-Anwendungen, die allen Anforderungen und Empfehlungen der eIDAS-Richtlinie genügt. Die Konformität mit [eIDAS] kann durch bestehende Zertifizierungssysteme nachgewiesen werden, die von der Peer-Group der EU akzeptiert werden. Implementierer von nicht auf SE basierenden Konzepten müssten neue Konzepte entwickeln, die die Konformität mit [eIDAS] belegen, und die Peer-Group davon überzeugen, dass ein potenziell verbleibendes Risiko für den Betrieb ermittelt wird und vom anmeldenden Mitgliedstaat abgedeckt wird.

Darüber hinaus ist [eIDAS] nicht die einzige relevante Regelung. Die Zahlungsdienstleistungsrichtlinie (Payment Services Directive, PSD2) verpflichtet Zahlungsdienste, eine starke Authentifizierung und Autorisierung von Transaktionen zu unterstützen. Das SE-basierte Konzept erfüllt auch diese Anforderung in vollem Umfang.

2.2 Usability

Usability, d.h. die Verwendbarkeit ist ein wesentlicher Faktor für die Kundenakzeptanz und den Markterfolg mobiler Anwendungen. Authentifizierungsmechanismen, die in einem mobilen Gerät ohne SE-Unterstützung (z. B. Fingerabdruck) implementiert sind, sind bequem, entsprechen aber möglicherweise nicht dem von [eIDAS] erwarteten Sicherheitsniveau. Dies könnte die Sicherheit beeinträchtigen und die Datenschutzoptionen einschränken. Externe Token können alle Sicherheitsanforderungen erfüllen, verringern jedoch den Komfort. Das SE- und zertifizierungsbasierte Konzept ist der einzige verfügbare, zertifizierbare und skalierbare Ansatz, der die Integration aller Funktionen, die die Benutzerfreundlichkeit unterstützen, in das mobile Gerät ermöglicht.

2.3 Security and privacy by design

Das SE-basierte Architektur- und Zertifizierungskonzept macht das mobile Gerät zu einem vertrauenswürdigen Benutzermedium, das skaliert werden kann, um alle von [eIDAS] definierten Sicherheitsstufen zu erfüllen und dieselbe Sicherheitsstufe zu bieten wie sichere Chipkarten. Im Gegensatz zu anderen Konzepten besteht keine Notwendigkeit, verbleibende Risiken durch gezielte Maßnahmen auf der Ebene des Anwendungssystems zu mindern und damit z.B. die Datenschutzoptionen des Benutzers einzuschränken. Nur der SE-basierte Ansatz bietet die Möglichkeit, sehr sensible Daten unter der physischen Kontrolle des Benutzers zu verwalten, und eine Option für die Offline-Nutzung des Dienstes. Außerdem unterstützen SEs die anonyme Verwendung von hochwertigen Zahlungsmitteln oder Tickets.

Ein offenes Ökosystem, das auf maximale Flexibilität bei der Unterstützung mobiler Anwendungen und die vollständige Unterstützung von Datenschutz- und Datenschutzoptionen abzielt, erfordert eine vertrauenswürdige Architektur in einem mobilen Gerät. In dieser Hinsicht ist das SE-basierte Konzept eindeutig die bessere Wahl.

2.4 Diskriminierungsfreier Zugang für Service-Provider

Die etablierten Service-Provider haben gelernt, mit unsicheren Benutzermedien wie zum Beispiel Magnetstreifenkarten umzugehen. Sie verfügen über ausgeklügelte Systeme, die bei der Aufdeckung von Betrug die Transaktionen und das Verhalten der Nutzer beurteilen, und sie haben die Erfahrung, Risiken in die Preisgestaltung einzubeziehen. Neulinge haben diese Fähigkeiten kaum. Unsichere Benutzermedien würden für sie einen erheblichen zusätzlichen Aufwand auf der Anwendungsseite, ein undefiniertes geschäftliches Risiko und ein Hindernis für den Markteintritt bedeuten; im Vergleich zu etablierten Service-Providern würden Neulinge hierdurch eindeutig diskriminiert. Dies verstößt gegen ein wesentliches Prinzip eines offenen, diskriminierungsfreien Ökosystems.

2.5 Wiederverwendbarkeit in anderen Marktsektoren

Die eIDAS-Richtlinie gilt formal nur für den europäischen eGovernment-Markt. Die mobilen eID-Dienste, die in Übereinstimmung mit [eIDAS] implementiert werden, werden jedoch auf jeden Fall auch in anderen Märkten eingesetzt. Einige dieser Märkte haben eigene Anforderungen an ein höheres Sicherheitsniveau und andere Merkmale, die am besten durch einen auf SE basierenden Ansatz erfüllt werden können. Beispiele hierfür sind Zahlungen, Gesundheitswesen, Mobilität, Industrie 4.0. Es ist sehr wahrscheinlich, dass ein offenes, skalierbares Konzept, das den höheren Sicherheitsstufen von [eIDAS] entspricht, für die Verwendung in anderen Sektoren exportiert wird. Es ist daher auch in dieser Hinsicht die beste Wahl für ein offenes Ökosystem für mobile Dienste.

Service Provider und insbesondere Benutzer können von einem zusätzlichen Nutzen profitieren, wenn die in das mobile Gerät integrierte SE mit einer NFC-Schnittstelle kombiniert wird:

2.6 Intuitive Anwendung

Die NFC-Technologie kann die Interaktion und den Datenaustausch zwischen dem Benutzer und dem Dienst erheblich vereinfachen und beschleunigen.

3. Zielsetzung des Projektes OPTIMOS2.0

3.1 Hintergrund

Wie eingangs erwähnt ist die zunehmende Nutzung von Mobilgeräten für Dienste verschiedenster Art einer der weltweit bestimmenden Trends. Elektronische Identitäten, sogenannte eIDs, sind die Schlüssel zu mobilen Diensten. Der diskriminierungsfreie Zugang zu geeigneten eIDs ist die Voraussetzung für eine umfassende Teilhabe der Nutzer und Anbieter an den Möglichkeiten der digitalen Servicewelt.

OPTIMOS2.0 sollte eine Infrastruktur für mobile Dienste schaffen, die die Herausforderungen an ein offenes und diskriminierungsfrei zugängliches Ökosystems in vollem Umfang erfüllt.

3.2 Wesentliche Herausforderungen in OPTIMOS2.0

Die wesentlichen Herausforderungen an SE-basierte Architekturen für mobile Dienste bestanden im Allgemeinen darin,

- die hohe Komplexität bestehender Systeme zu vereinfachen,
- Marktreichweite zu erzielen und
- allen interessierten Service-Providern (SPs) diskriminierungsfrei einen Platz auf einem Sicherheitselement (SE) im Mobilgerät des Kunden verfügbar zu machen.

Herausforderung 1: Reduktion der Komplexität

Nach dem vorhergehenden Stand der Technik mussten in einer Secure-Element Infrastruktur alle Service-Provider Verträge mit jedem Anbieter von Sicherheitselementen bzw. sicheren Plattformen schließen (vgl. Abbildung 2 1). In Deutschland sind dies alleine drei Anbieter, europaweit ca. 80 und weltweit über 300 – ein nahezu aussichtsloses Unterfangen.

Abbildung 2: Secure Element Infrastruktur aus heutiger Sicht

Herausforderung 2: Marktreichweite

Das TSMS musste eine große Marktreichweite erreichen und mittelfristig möglichst alle Smartphone-Nutzer bedienen können. Dies erforderte, dass das TSMS sowohl auf die eUICCs der Mobilfunknetzbetreiber als auch die eSEs der Mobilgerätehersteller zugreifen können sollte.

Herausforderung 3: Diskriminierungsfreier Zugang

Das TSMS musste Service-Providern einen diskriminierungsfreien Zugang zu den Sicherheitselementen ermöglichen.

3.3 Lösungsansätze in OPTIMOS2.0

Um der ersten Herausforderung zu begegnen, wurde in OPTIMOS2.0 ein Trusted-Service-Management-System definiert. Diese Komponente diente als zentrales und neutrales Bindeglied zwischen den Service-Providern z.B. für eID-Anwendungen, öffentlichen Personenverkehr oder Car Sharing auf der einen und den Anbietern, d.h. den Herstellern, Betreibern oder Service-Providern der sicheren Plattformen (eSE, eUICC) im Mobiltelefon auf der anderen Seite.

Das System wurde in zwei Teilsysteme aufgeteilt.

Auf der einen Seite sollte die Integration der Sicherheitselemente der MNOs und OEMs durch das erste Teilsystem des **Trusted-Service-Managers für Secure Elements** unterstützt werden. Der Betreiber dieses Teilsystems ist für die Verbindung der Service-Provider zu den SEs und für die Verwaltung der SEs im Smartphone verantwortlich.

Auf der anderen Seite wurde durch den **Application-Manager** ein diskriminierungsfrei zugängliches Teilsystem definiert, das auf die Besonderheiten einzelner Anwendungen und Märkte eingehen und Zugangshemmnisse minimieren sollte. Ziel dieses Teilsystems ist es, regionale oder nationale Service-Provider einzubinden und die kommerzielle und organisatorische Schnittstelle im Sinne von „One-stop-shopping“ zum Service-Provider zu bilden.

Komponenten und Schnittstellen des TSMS wurden in die internationale Normung eingebracht. Ferner wird derzeit eine Technische Richtlinie BSI TR zur Ausgestaltung des TSMS verfasst.

3.4 Das TSMS in OPTIMOS2.0

Im TSMS-Ansatz des OPTIMOS2.0-Projekts wurde erstmals ein „generischer“ Trusted-Service-Managers für Secure Elements geschaffen, der den Zugang zu allen Anbietern von sicheren SE-Plattformen schafft. Dies war erforderlich, da nicht nur die MNOs, sondern auch die eSE der verschiedensten Mobilgerätehersteller integriert werden, um die nötige Kundenreichweite zu erreichen.

Die folgende Abbildung zeigt die grundlegende Funktion des TSMS. Wie zu erkennen ist, entfallen die in oben dargestellten, bilateralen Abstimmungen zwischen den Service-Providern auf der einen und den Anbietern der Sicherheitselemente auf der anderen Seite.

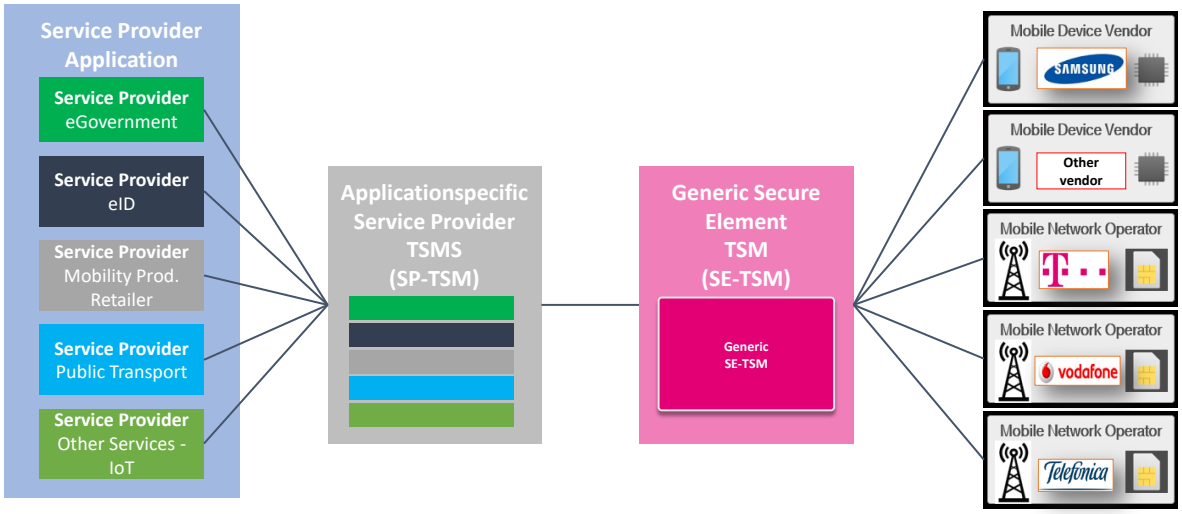


Abbildung 3: Grundlegende Funktionen des TSMS

3.4.1 Application-Manager für Service-Provider

Der Application-Manager dient als Schnittstelle für den Service-Provider, um seine Services im Gesamtsystem zu administrieren und die eigenen Daten zu verwalten. Zu diesem Zweck bietet der Application-Manager dem Service-Provider eine Benutzeroberfläche in Form eines internetbasierten, leicht bedienbaren, grafischen Webinterfaces - der Management-UI - an.

Über das Management-UI können

- Service-Provider diskriminierungsfrei angebunden werden (Onboarding),
- die eigenen Unternehmensdaten verwaltet,
- die Services und alle zugehörigen Applets verwaltet sowie
- die Nutzer des eigenen Unternehmens administriert werden.

Das Management-UI unterstützt für die genannten Prozesse unter anderem die folgenden Funktionalitäten:

- Bereitstellung der Spezifikationen der Application-Interfaces (APIs), Guidelines, Dokumentationen und rechtlichen Dokumenten
- Bereitstellung von Entwicklungsressourcen wie einem SDK, notwendigen Apps und weiteren Dateien
- Bereitstellung von APIs zur Anbindung an Backend-Systeme des Service-Providers;
- Bereitstellung von APIs für das Order- und Status-Management zur Unterstützung von Customer-Care-Prozessen
- Verwaltung von Services, zugehörigen Applets und sonstigen zugehörigen Informationen, z.B. Installationsskripte
- Verwaltung von Meta-Informationen des Service-Providers sowie Management der Nutzer des Service-Providers.

3.4.2 Trusted-Service-Manager für Secure-Elements

Neben den rein administrativen Tätigkeiten, die der Service-Provider im Application-Manager durchführt, müssen die Services des Service-Providers in Form von Applets[2] auch auf die Secure-Elements aufgebracht werden. Dieser Aspekt der TSMS-Funktionalität wird Applet-Lifecycle-Management oder kurz ALM genannt und bildet den zentralen Mehrwert des Trusted-Service-Manager für Secure-Elements.

Unter ALM wird die Menge an Prozessen zur Überprüfung der Verwendbarkeit des Applets auf dem Hardwaresicherheitselement (Eligibility-Check), zur Aufbringung auf dem jeweiligen Chip sowie zur Aktivierung/Deaktivierung und zum Löschen des Applets verstanden. Für die Realisierung des ALM ist der Trusted-Service-Manager zuständig.

Um eine einfache Anbindung an den Trusted-Service-Manager zu ermöglichen, wird Service-Providern, wie in Abbildung 4 dargestellt, ein TSMS-SDK zur Integration in die eigene Applikation auf dem Endgerät bereitgestellt. Damit wird es auf einfache Weise ermöglicht, die Lebenszyklen der eigenen Applets auf den Hardwaresicherheitselementen des jeweiligen Endgeräts zu verwalten.

Für das ALM stellt der Trusted-Service-Manager Service-Providern folgende Funktionalitäten und Leistungen bereit:

- Bereitstellung eines SDKs für die Integration der TSMS Funktionalitäten in die Service-Provider App
- Bereitstellung der APIs zum Verwalten des Lifecycles von Applets auf den Secure-Elements mobiler Endgeräte
- Bereitstellung der APIs zum Verwalten des Lifecycles von Services auf den Secure-Elements mobiler Endgeräte
- Abruf eines eindeutigen Identifiers zur Referenzierung des Endgerätes innerhalb des OPTIMOS TSMS.

Durch diese Funktionalitäten wird der Service-Provider – sortiert nach Wichtigkeit – bei der Durchführung folgender Prozesse unterstützt:

- **Provisionierung von Applets und gegebenenfalls Personalisierungsdaten auf ein Secure-Element**
  Prozess, um alle hinterlegten Applets eines Service auf ein verfügbares SE im Gerät des Endkunden aufzuspielen.
- **Löschung von Applets, gegebenenfalls inkl. Personalisierungsdaten**
  Komplette Löschung aller mit dem Service assoziierten Applets und Daten auf dem jeweiligen Secure-Element des Gerätes.
- **Eligibility-Check**
  Überprüfung, ob die Applets des Service auf einem der verfügbaren Secure-Elements des Gerätes aufgespielt werden können.

[2] Applikationen zur Ausführung auf Secure-Elements, die auf Grund ihrer geringen Größe Applet statt App genannt werden.

- **Personalisierung**
  Prozess, um mit Hilfe des TSMS über einen gesicherten Kanal Daten in die Applets des Service-Providers einzubringen.

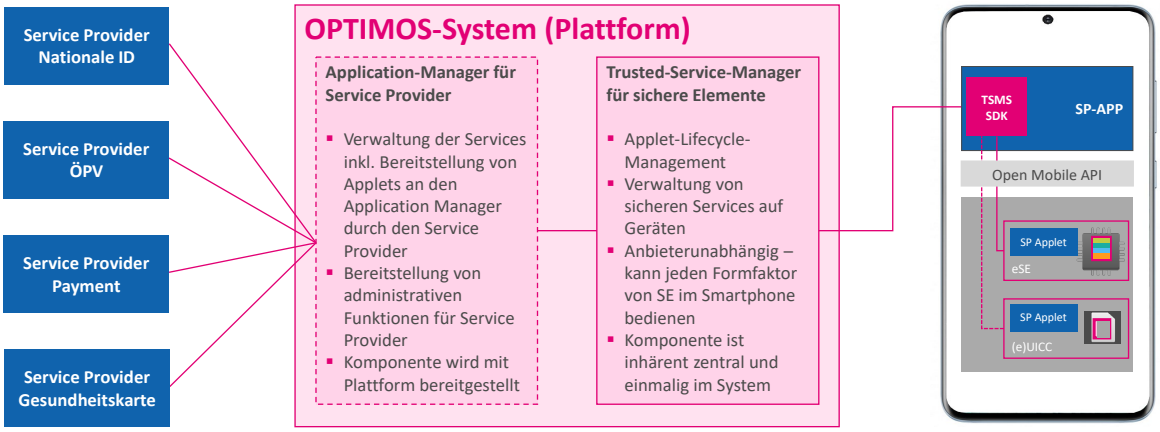


Abbildung 4: Übersichtsbild des TSMS, bestehend aus Application-Manager und Trusted-Service-Manager

4. Quo Vadis OPTIMOS2.0?

OPTIMOS bzw. die TSMS-Infrastruktur lassen sich in diversen Sektoren sowohl aus dem hoheitlichen Bereich als auch aus der Privatwirtschaft einsetzen.

Im Rahmen des Projektes OPTIMOS2.0 wurden mittels des TSMS verschiedene Use Cases exemplarisch umgesetzt. Hierzu gehörten die Provisionierung des eID-Applets als Grundlage des mobilen Personalausweises oder die Provisionierung von ÖPV-Tickets.

Das TSMS erlaubt darüber hinaus jedoch auch den Einsatz in einer Reihe weiterer Anwendungsfelder. Hierzu gehören neben der Provisionierung von Schlüsseln für das Car Sharing auch Boarding Pässe für Flugzeuge oder die Provisionierung von Daten im Internet der Dinge.

Die Anwendungsfelder von OPTIMOS2.0 werden unter anderem in der Schaufensterinitiative „Sichere Digitale Identitäten“ erprobt und weiterentwickelt.

5. Ergänzungen zum Förderprojekt OPTIMOS 2.0

Das diesem Text zugrundeliegende Forschungs- und Entwicklungsprojekt[2] wurde mit Mitteln des Bundesministeriums für Wirtschaft und Energie (BMWi) innerhalb des Technologieprogramms „Smart Service Welt II" in den Jahren 2018 bis 2020 gefördert[3] und vom Projektträger Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Köln, betreut. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

[2] https://www.bundesdruckerei.de/de/Unternehmen/Innovation/Optimos

[3] https://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/SmartServiceWeltProjekte/Wohnen_Leben/SSWII_Projekt_OPTIMOS_20.html

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Medical Device Security: Results from Project ManiMed

Julian Suleder, M.Sc.[1], Dr. Dina C. Truxius[2]

Abstract:
Medical device IT security is considered in relevant laws and regulations regarding safety and performance. The general IT security posture of medical devices is not as mature as possible, due to a strong focus on safety often contributed by a long product life cycle. Consequently, medical devices should be examined in-depth for IT security vulnerabilities throughout the product lifecycle, especially before being placed on the market. The maturity will only significantly improve if approval processes imply defined IT security requirements for medical devices. Further, a prompt and effective reaction of the medical device manufacturer after discovering vulnerabilities is only possible within a framework of established and well-defined processes. This article elucidates these introductory statements by exemplifying vulnerabilities in three medical devices, which were identified in the BSI project ManiMed.

Keywords: Bluetooth Low Energy, Coordinated Vulnerability Disclosure, CVE, HL7, Insulin Pump, Networked Medical Device, Patient Monitor, Patient Safety, Security Assessment, Security Research, Vulnerabilities

1. Introduction

Digital networking is already widespread in many areas of life. In the health sector, more and more medical devices are becoming networked. The security of these devices already plays a significant role, which is based on the steadily increasing number of networked medical devices as the Federal Office for Information Security (BSI) in Germany shows in its report on the Current State of IT Security in Germany in 2018 and 2019 [1]. According to BSI, there is a steadily increasing number of networked medical devices. This is corroborated by an increase in attacks on such devices, concomitant with a potential impact on patient safety [1].

A balance between medical functionality and IT security must be found to ensure continued availability and reliable use of the complex medical device systems in everyday life [1]. IT security is not only a continuous process, but has to be a core issue when developing and manufacturing medical devices, thereby considering trust relationships between communicating parties and exchanged data.

In its role as the federal IT security authority, the BSI aims to sensitize for security risks in networked medical devices and initiated the project Manipulation of Medical Devices (ManiMed). A security analysis of selected products was performed through security assessments by the security service provider ERNW Research GmbH on behalf of the BSI to gain insights into such devices' IT security posture on the German market.

[1] ERNW Research GmbH, Heidelberg.
[2] Federal Office for Information Security (BSI), Bonn.

2. Vulnerabilities in Medical Devices

As with any other IT system, vulnerabilities can exist in medical devices. So far, there was no systematic study that determines and evaluates the prevalence and criticality of vulnerabilities in networked medical devices as access to a large number of different devices is exceptional, and testing in production environments is risky. Therefore, only little data is available on the effectiveness of regulatory measures aiming to prevent such vulnerabilities, for instance, when a medical device gets approved for the German market. The BSI project ManiMed [2] is the first project of this kind, which evaluates IT security for networked medical devices via security assessments. The project report was published on December 31, 2020, and can be accessed on the BSI website [3].

In the past, there have been several publications about IT security of medical devices such as for example, the talk Understanding and Exploiting Implanted Medical Devices reported critical vulnerabilities in implanted medical devices [4] and the ERNW White Paper 66, which summarizes the status of medical devices' security based on publicly available information [5]. Furthermore, tests carried out by ERNW on select patient monitors, syringe pumps, electroencephalography systems, home monitoring devices, and magnetic resonance tomography devices have shown that the devices analyzed were affected by several, sometimes critical, vulnerabilities [6,7]. Different groups of people and interests such as patients, security researchers and universities have been actively researching this area for years [8,9,10]. The need to address the problems identified has never been more urgent.

3. Medical Device Security Framework

Legal requirements regulate the safety and functionality of the medical devices used in Germany. The framework for this is specified in the Medical Devices Act (MPG) in Germany and the European directives on active implants (90/385 / EEC), on medical devices (93/42 / EEC), and on in vitro diagnostics (98/79 / EC). The development of the medical device is accompanied by complex processes that stipulate risk analyses and safety requirements. The risk analysis must weigh up the benefits of the product’s intended use with the foreseeable misuse, including all risks and identified hazards. For medical devices, proof of safety and therapeutic or diagnostic performance for a medical device is attested by conformity assessments. IT security is mainly assessed in the context of potential impact on patient safety according to the assessment procedures defined in the MPG [11] or Regulation (EU) 2017/745 on medical devices (MDR) [12]. Any security risk arising from vulnerabilities that do not impact patient safety is not explicitly mentioned in the aforementioned legal frameworks.

4. Case Studies

The following sections present three medical devices with exemplifying vulnerabilities identified in project ManiMed [2], [[3].

4.1. Case Study 1: DANA Diabecare RS Insulin Pump

In the security assessments, severe vulnerabilities in the DANA Diabecare RS insulin pump by the Korean manufacturer SOOIL Developments Ltd. were identified [3], [13]. The manufacturer fixed the vulnerabilities within the framework of project ManiMed.

Patients use an insulin pump to administer insulin to treat diabetes mellitus (type 1 diabetes). In general, insulin pumps are a less invasive and significantly more flexible alternative than several manual insulin administrations per day. In addition to the physical input options, the DANA Diabecare RS insulin pump can be controlled with mobile apps via a proprietary communication protocol that uses Bluetooth Low Energy (BLE).

4.1.1. Identified Vulnerabilities

All vulnerabilities affected the proprietary communication protocol between the insulin pump and the associated mobile applications. During the security assessment, it was identified that some of the security functions were only implemented in mobile applications. Still, the corresponding checks were not carried out by the insulin pump [3], [13]. For example, users needed to select a personal identification number (PIN) to lock the pump’s control panel. This PIN was also needed for unlocking a virtual control panel in mobile applications. However, the PIN was only checked by mobile applications and not by the insulin pump. An attacker without knowledge of the PIN who interacts directly with the insulin pump could omit the verification. It also follows that the mobile application had to obtain the correct PIN for performing a check. This was implemented by exchanging the PIN every time communication is established between the insulin pump and the mobile application. Consequently, unauthenticated attackers could sniff the PIN via Bluetooth Low Energy (BLE) packets [3], [13].

Although the communication between the mobile applications and the insulin pump was encrypted and all cryptographic keys used in the proprietary communication protocol were found to be generated deterministically, these keys depended on the insulin pump's hardware clock, the serial number, and the already addressed PIN.

Further, the mobile applications' authentication solely depended on the possession of one of these keys. This so-called pairing key was exchanged as soon as the insulin pump and mobile application are paired. None of the cryptographic keys, including the pairing key, were adequately protected and communicated between the insulin pump and mobile applications [13].

4.1.2. Impact on Patients

Combining the identified vulnerabilities enabled attackers to take over the DANA Diabecare RS insulin pump via Bluetooth Low Energy (BLE) and control it remotely. Carrying out this attack required sniffing a single handshake between an insulin pump and a paired mobile application using simple Bluetooth Low Energy (BLE) hardware within an appropriate range. The attacker could then control all pump functionalities related to its Bluetooth Low Energy (BLE) interface. The attack could be performed automatically [3], [13], which is demonstrated in a published video [14].

According to § 5 MPG, the manufacturer published a Field Safety Notice on May 08, 2020, via the Federal Institute for Drugs and Medical Devices (BfArM) [15]. The manufacturer provided a firmware update for the insulin pump, eliminating all identified vulnerabilities by adapting the communication protocol. Besides, the mobile applications had to be updated on Android and iOS, thereby ensuring compatibility with the altered communication protocol. To temporarily reduce the risk of patient safety impairment, the manufacturer recommended deactivating BLE by switching the pump to airplane mode [15]. In this state, the insulin pump's therapeutic purpose could be retained as control via mobile applications is optional. Additionally, the device has implemented safeguards, such as a maximum daily dose. According to the manufacturer, there is no known exploitation of the vulnerabilities besides the demonstration in the test environment [15].

4.1.3. Summary

In total, the analysis and transmission of the vulnerabilities, evaluation, implementation of the measures, evaluation of the effectiveness of the implemented security measures, coordination of the roll-out of updated firmware and corresponding regulatory processes took nine months. The whole process was characterized by reliable and trustful cooperation with the manufacturer [13].

4.2. Case Study 2: Philips Patient Monitoring Devices

Another case study conducted as part of project ManiMed focuses on Philips patient monitoring devices, which are networked devices connected to a Philips central monitoring station. The assessed system contained a Patient Information Center iX surveillance station and two IntelliVue MX850 patient monitors and network infrastructure components, representing a nearly fully-featured clinical environment similar to productive setups. The manufacturer has an established relationship with the US-based CISA CERT to handle these kinds of advisories and is regularly communicating via CISA advisories as part of a CVD process. Hence, Philips issued a security advisory [16] as part of the CVD process. Further, multiple updates, including additional security measures to fix the vulnerabilities and strengthen the monitoring system's security maturity, are planned to be released in the future [16].

Patient monitoring solutions are used to measure patients’ vital signs during transport or hospitalization contiuously. Features may include data storage, alarming functionality, and connectivity to central medical monitoring systems.

4.2.1. Identified Vulnerabilities

During the security assessment, an application crash of the server application via a single specially crafted packet could be identified. The packet led to a reboot of the application and its server host 3].

Further, the inspected patient monitors improperly checked certificate revocations, enabling attackers with access to a trusted certificate to obtain a Man-in-the-Middle (MitM) position between patient monitors and the server applications. During testing, this position could crash the patient monitor since the device did not validate inputs.

This caused the monitor to restart after crashing. During restart, vital signs could neither be measured nor transmitted to server applications [3].

4.2.2. Impact on Patients

The vulnerabilities mentioned above resulted in unauthorized access, interrupted monitoring, collection of access information, and patient data. An attacker with expert level skills would need to gain access to the medical device network [16].

4.3. Case Study 3: HL7v2 Injections in Patient Monitors

In project ManiMed, a vulnerability in a patient monitor that sends HL7v2 messages was identified (CVE-2020-27260). A blog post about the vulnerability was published in April 2020 [17], a respective security advisory on January 07, 2021 [18].

The HL7v2 standard is a common, text-based interoperability standard for health and medical transactions between medical systems and devices to support hospital workflows. The standard enables the interoperability of heterogeneous medical information systems and active medical devices such as patient monitors. Agreements on the message structure and content representations are required to parse and process the messages handled in HL7v2 with an encoding syntax based on segments (lines) and one-character delimiters. Each segment starts with a string that identifies the segment type. Every message's first segment represents the message header segment, which determines the message type and expected message segments [17].

4.3.1. Identified Vulnerabilities

The target of evaluation, a patient monitor, used a message type that is in place to transmit observations and results from production to ordering systems. An attacker with physical access to the device could inject valid, but non-authentic HL7v2 segments into the HL7v2 messages using a connected barcode reader as encoding characters in untrustworthy inputs were processed. Hence, attackers could tamper with data transmitted to further network-connected systems. The malicious barcode was needed because it circumvents special character restrictions in the patient monitor's input fields. The intended use of such barcode readers is to scan a patient's personally identifiable information (PII) to speed up clinical processes [17].

4.3.2. Impact on Patients

A malicious barcode could contain HL7v2 special characters and a segment representing, for example, manipulated observation results instead of a primary diagnosis. As these results are not necessarily limited to the patient under observation, multiple manipulated records can be sent in a single message, which could cause misdiagnosis or medical errors. The vulnerability could be fixed by escaping the HL7v2 special characters and symbols present in potentially untrusted user input. An attacker with expert level skills would need to gain physical access to the medical device [17], [18].

5. Conclusions

For the assessments, as mentioned earlier, of three medical devices and the exemplifying vulnerabilities, various generalizations and conclusions can be drawn.

All case studies show that medical devices and communication systems often assume that trust relationships between parties of a communication system apply to the processed data. When exchanging data between different systems, it must be assumed that compromised or malicious systems might also be involved in the communication. Medical decisions that are based on manipulated data may pose a high risk to patient safety. As a result, strict validation of trust relationships and processed data is crucial.

The Philips patient monitoring system's assessment results demonstrate that the chaining of individual, per-product vulnerabilities into a complex attach scenario can often turn a successful attack on one product into an attack that also affects the other product in the system. This is related to the concept of lateral movement. It is a vital pre-condition that the manufacturer cooperates and provides a full-fledged system for testing to perform this advanced security assessment, which widens the focus beyond the individual product.

Assumptions about the environment and trust relationships at interfaces between components can cause security vulnerabilities when the assumed and actual behavior differ. System-level security testing can help to avoid discrepancies. The evaluation of complex medical systems should be preferred over assessments of single components to examine assumptions about their environment and trust relationships.

The security assessments and subsequent processes demonstrated that the IT-security posture and handling of disclosed vulnerabilities vary signigicantly between manufacturers and highly depend on their maturity.

Changes to medical devices that affect IT security are often made in the context of remedying vulnerabilities identified internally or by external parties, as reports by the ICS-CERT of CISA in the USA show [19]. Often, not all identified vulnerabilities are eliminated at once. Residual risk analysis or effort assessment is almost always carried out to prioritize the vulnerabilities. A subset of the vulnerabilities may be accepted since their criticality is too low or the effort required to remedy is too high. Additionally, the downtime incurred with applying a software patch to a medical device that is in regular use has to be taken into account.

Even in a supposedly secure system, vulnerabilities can never be completely ruled out. Therefore, the professional handling of vulnerabilities is an integral part of the manufacturer's product life cycle activities. Usually, when external parties report vulnerabilities, Coordinated Vulnerability Disclosures (CVD) are carried out. A manufacturer can only reliably carry out a CVD if a well-defined response process is established, e. g., by providing public contact information for reporting vulnerabilities, thereby enabling CVD processes to be as efficient as possible. Regulatory duties of fixing medical device vulnerabilities in Germany arise only if the vulnerabilities impact patient safety.

It is not uncommon for security vulnerabilities to arise due to discrepancies between a socio-technical system's specified and real behavior. This discrepancy can hardly be identified in documentation audits and specification reviews. Well-proven means are external penetration tests, in which a complex, active medical device is examined in-depth before productive use. Dependent on the device functionality, external medical and non-medical, wireless and wired network interfaces, Bluetooth and USB interfaces, update, maintenance, and configuration mechanisms are examined for vulnerabilities, and existing security measures are checked for effectiveness.

Manufacturers have to be motivated to proactively consider IT-security related topics such as implementing and enforcing secure development life cycles for their medical devices in combination with a timely and effective response to disclosed vulnerabilities to achieve and maintain IT security at a high level. Further, legal frameworks and standards have to cover IT security and related processes. Different parties, such as regulators and vendors, have to exchange knowledge, thereby increasing the devices' security levels. The large number of vulnerabilities identified within project ManiMed [3] corroborates previous research efforts and admits of improvement concerning IT security in the health sector.

Literaturhinweise

[1] Federal Office for Information Security (BSI): Report on the State of IT Security in Germany. Online (new URL since February 01, 2021): https://www.bsi.bund.de/EN/SecuritySituation

[2] Bundesamt für Sicherheit in der Informationstechnik (BSI). Projekte des BSI im Bereich der Medizintechnik. Online (new URL since February 01, 2021): https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/E-Health/Medizintechnik/Projekte/projekte_node.html

[3] Bundesamt für Sicherheit in der Informationstechnik (BSI). Veröffentlichungen. Online (new URL since February 01, 2021): https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/E-Health/Medizintechnik/Veroeffentlichungen/veroeffentlichungen_node.html

[4] Billy Rios, Jonathan Butts. Understanding and Exploiting Implanted Medical Devices. Black Hat USA 2018. Online (accessed January 08, 2021): https://www.blackhat.com/us-18/briefings/schedule/index.html#understanding-and-exploiting-implanted-medical-devices-11733

[5] Julian Suleder, Andreas Dewald, Florian Grunow. Medical Device Security: A Survey of the Current State. ERNW White Paper 66. 2018. Online: (accessed January 08, 2021): https://ernw-research.de/en/whitepapers/issue-66.html

[6] Süddeutsche Zeitung (SZ). Nächtliches Desaster. 2015. Online (accessed January 08, 2021): https://www.sueddeutsche.de/wirtschaft/medizintechnik-naechtliches-desaster-1.2534424

[7] Florian Grunow. The patient's last words: I am not a target!. 2015. Online (accessed January 08, 2021): https://insinuator.net/2015/07/the-patients-last-words-i-am-not-a-target/

[8] Johannes Sametinger, Jerzy Rozenblit, Roman Lysecky, Peter Ott. Security Challenges for Medical Devices. Online (accessed January 08, 2021): https://www.se.jku.at/wp-content/uploads/2015/03/TR-SE-15.03.pdf

[9] Mandeep Khera. Think Like a Hacker: Insights on the Latest Attack Vectors (and Security Controls) for Medical Device Applications. ournal of Diabetes Science and Technology. 2017;11(2):207-212. doi:10.1177/1932296816677576

[10] Jay Radcliffe. Hacking medical devices for fun and insulin: breaking the human SCADA system. Online (accessed January 08, 2021): https://media.blackhat.com/bh-us-11/Radcliffe/BH_US_11_Radcliffe_Hacking_Medical_Devices_WP.pdf

[11] Gesetz über Medizinprodukte (Medizinproduktegesetz - MPG)

[12] Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, zur Änderung der Richtlinie 2001/83/EG, der Verordnung (EG) Nr. 178/2002 und der Verordnung (EG) Nr. 1223/2009 und zur Aufhebung der Richtlinien 90/385/EWG und 93/42/EWG des Rates

[13] Julian Suleder. Safety Impact of Vulnerabilities in Insulin Pumps. ERNW White Paper 69. 2020. Online (accessed January 08, 2021): https://ernw-research.de/en/whitepapers/issue-69.html

[14] ERNW Research GmbH. Demo: Hijacking the DANA Diabecare RS Insulin Pump. Online (accessed January 08, 2021): https://www.youtube.com/watch?v=0GMe2poiYtE

[15] Federal Institute for Drugs and Medical Devices (BfArM): Urgent Field Safety Notice for Insulinpumpe DANA Diabecare RS;mobilen Anwendung AnyDANA by SOOIL Development Co. Ltd. Online (accessed January 08, 2021): https://www.bfarm.de/SharedDocs/Kundeninfos/EN/07/2020/17203-19_kundeninfo_en.html

[16] Philips via Cybersecurity & Infrastructure Security Agency (CISA). ICS Medical Advisory (ICSMA-20-254-01) - Philips Patient Monitoring Devices. Online (accessed January 08, 2021): https://us-cert.cisa.gov/ics/advisories/icsma-20-254-01

[17] Julian Suleder. Medical Device Security: HL7v2 Injections in Patient Monitors. April 23, 2020. Insinuator Blog. Online (accessed January 08, 2021): https://insinuator.net/2020/04/hl7v2-injections-in-patient-monitors/

[18] Cybersecurity & Infrastructure Security Agency (CISA). ICS Medical Advisory (ICSMA-21-007-01) - Innokas Yhtymä Oy Vital Signs Monitor. Online (accessed January 08, 2021): https://us-cert.cisa.gov/ics/advisories/icsma-21-007-01

[19] Cybersecurity & Infrastructure Security Agency (CISA) – ICS-CERT Advisories. Online (accessed January 08, 2021): https://www.us-cert.gov/ics/advisories

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Ganzheitliches Notfallangebot für KMU

Dipl. Math. Angelika Jaschob[1]

Kurzfassung:
Während Kritische Infrastrukturen und Konzerne nach einem IT-Sicherheitsvorfall auf interne Forensik-Teams zurückgreifen können, stehen hingegen kleine und mittelständische Unternehmen (KMU) und Bürger meist alleine da. Ohne die notwendige Expertise und Erfahrung fällt es ihnen schwer einen IT-Sicherheitsvorfall zu bewerten und die richtigen Schritte zu ergreifen, um den Vorfall bewältigen und den Schaden eindämmen zu können. Dies bestätigt auch die aktuelle Umfrage in der Wirtschaft von DIHK, BDI und Bitkom[2], in der sich Unternehmen ein ganzheitliches Notfallangebot der Sicherheitsbehörden mit folgenden Bestandteilen wünschen:

- eine zentrale Notfallnummer für IT-Sicherheit,
- schnelle und ausführliche Informationen zur Lage,
- Bereitstellung von Informationspaketen inkl. einer Liste kompetenter Ansprechpartner im Notfall,
- schnelle und individuelle Hilfe nach einem IT-Sicherheitsvorfall.

Hier setzt das Cyber-Sicherheitsnetzwerk (CSN) an. Mit dem CSN wird eine flächendeckende dezentrale Struktur aufgebaut, die effizient und kostengünstig KMU und Bürgern bei IT-Vorfällen Unterstützung anbietet und die bestehenden präventiven Beratungsdienstleistungen des Bundes und der Länder gut ergänzt.

Stichworte: Cyber-Sicherheitsnetzwerk, CSN, Digitale Ersthelfer, Digitale Rettungskette, IT-Dienstleister, IT-Sicherheitsvorfall, IT-Störung, KMU, Qualifizierung, Schulung, Vorfall-Behandlung, Vor-Ort-Unterstützung, Vorfall-Experten, Zertifizierung

1. Das Cyber-Sicherheitsnetzwerk als Lösungsansatz für das ganzheitliche Notfallangebot

Die Zielsetzung dieses Projektes besteht in der Schaffung eines deutschlandweiten Netzwerkes zur Hilfe und Unterstützung von Bürgern sowie kleineren und mittelständischen Unternehmen bei der Lösung und Behebung von IT-Sicherheitsvorfällen. Es soll eine Landkarte für regionale Ansprechpartner auf Augenhöhe entstehen, die nach einem IT-Sicherheitsvorfall qualifiziert unterstützen können.

Die reaktive Unterstützung bei der Bearbeitung von IT-Sicherheitsvorfälle erfolgt im Cyber-Sicherheitsnetzwerkes in verschiedenen Eskalationsstufen, angefangen bei einer zentralen Telefon-Hotline für die erste Kontaktaufnahme über tiefergehende Analyse-Gespräche bis zum Einsatz eines Teams von Experten vor Ort.

Das Cyber-Sicherheitsnetzwerk ist ein freiwilliger Zusammenschluss von qualifizierten Experten für die Vorfallbearbeitung, die sich bereit erklären, ihre individuelle Expertise und ihr individuelles Knowhow zur Behebung von IT-Sicherheitsvorfällen zur Verfügung zu stellen, und die mithelfen die IT-Sicherheitslage in Deutschland zu verbessern.

---

[1] Bundesamt für Sicherheit in der Informationstechnik, Bonn

[2] Fortschreibung der Nationalen Cyber-Sicherheitsstrategie – Ergebnis einer Umfrage in der Wirtschaft https://www.dihk.de/resource/blob/30462/30debd93d9f3ce89ffc6c7a16fdc2f1d/umfrage-cybersicherheit-2020-data.pdf

Sie helfen durch die Übernahme reaktiver Tätigkeiten, IT-Sicherheitsvorfälle zu erkennen und zu analysieren, das Schadensausmaß zu begrenzen sowie weitere Schäden abzuwenden. Dabei kann die Unterstützung je nach Vorfall- und Zielgruppe unterschiedlich ausfallen (siehe Abbildung 1).

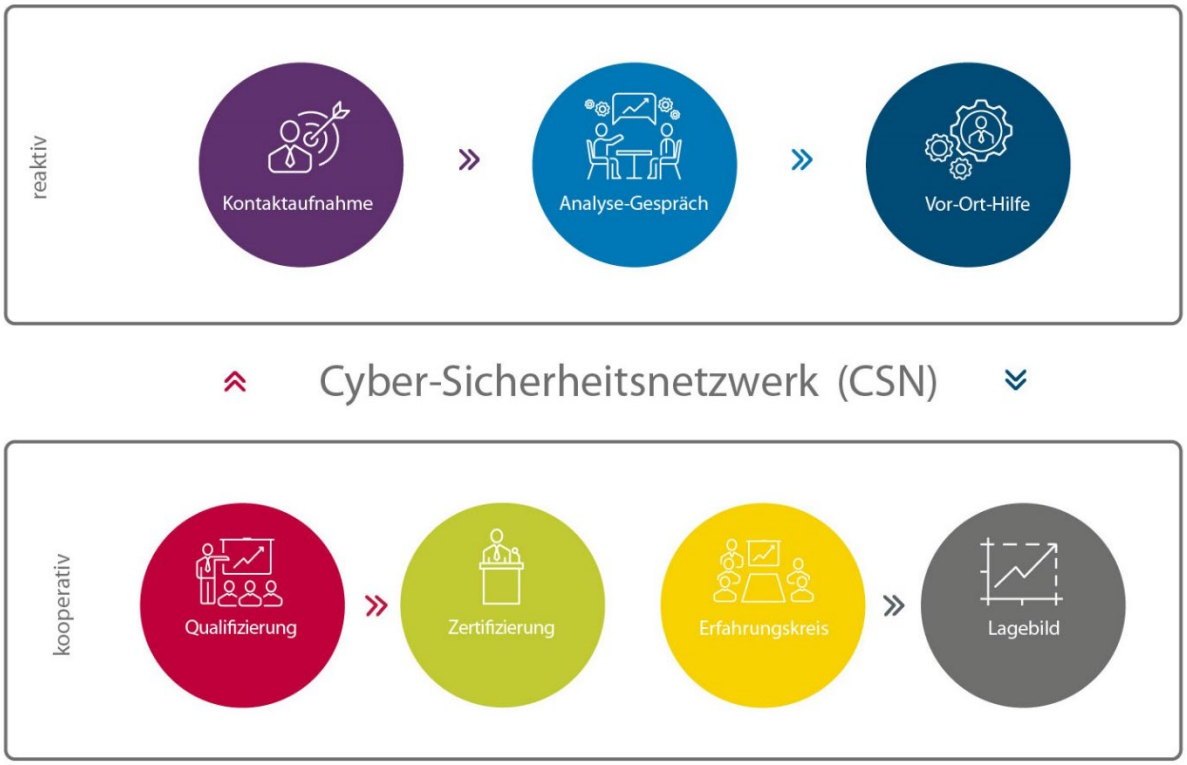


Abbildung 1: Cyber-Sicherheitsnetzwerk

Das Qualifizierungsprogramm stellt die einheitliche Qualität der Vorfallbearbeitung durch Experten (Digitale Ersthelfer und Vorfall-Experten) sicher. Zusätzlich wird die Qualifizierung der Digitalen Ersthelfer bzw. Vorfall-Experten mittels eines Testats bzw. einer Personenzertifizierung des BSI bescheinigt.

Durch den Austausch von Erfahrungen bei der Vorfallbearbeitung werden der Zusammenhalt des Experten-Netzwerks und der Aufbau einer einheitlichen Wissensbasis gefördert. Das Qualifizierungsprogramm soll auf der Grundlage aktueller Erkenntnisse aus den Vorfällen kontinuierlich erweitert werden. Zusätzlich sollen die Erkenntnisse in ein Lagebild einfließen. So können neue Empfehlungen und zusätzliche präventive Maßnahmen zielgerichteter erstellt und die Unterstützungsdienstleistungen durch das Cyber-Sicherheitsnetzwerk kontinuierlich erweitert und optimiert werden.

2. Die Aufbaustruktur des Cyber-Sicherheitsnetzwerks

Das Cyber-Sicherheitsnetzwerk bildet die zentrale erste Anlaufstelle sowohl für Betroffene als auch für Experten. Die Geschäftsstelle des CSN, welche beim Bundesamt für Sicherheit in der Informationstechnik liegt, nimmt die Registrierungen vor und beantwortet alle prozessualen und organisatorischen Fragen. Die strategische Ausrichtung

sowie die Rahmenbedingungen für das Cyber-Sicherheitsnetzwerk übernimmt ebenfalls eine Koordinierungsstelle im BSI, die von einem „Round-Table“, bestehend aus Vorfall-Experten sowie Vertretern von Behörden, Bildungsinstitutionen und unterschiedlichen Interessensgruppen, unterstützt wird. Einen Überblick über alle unterschiedlichen Rollen des Cyber-Sicherheitsnetzwerks gibt die folgende Abbildung.

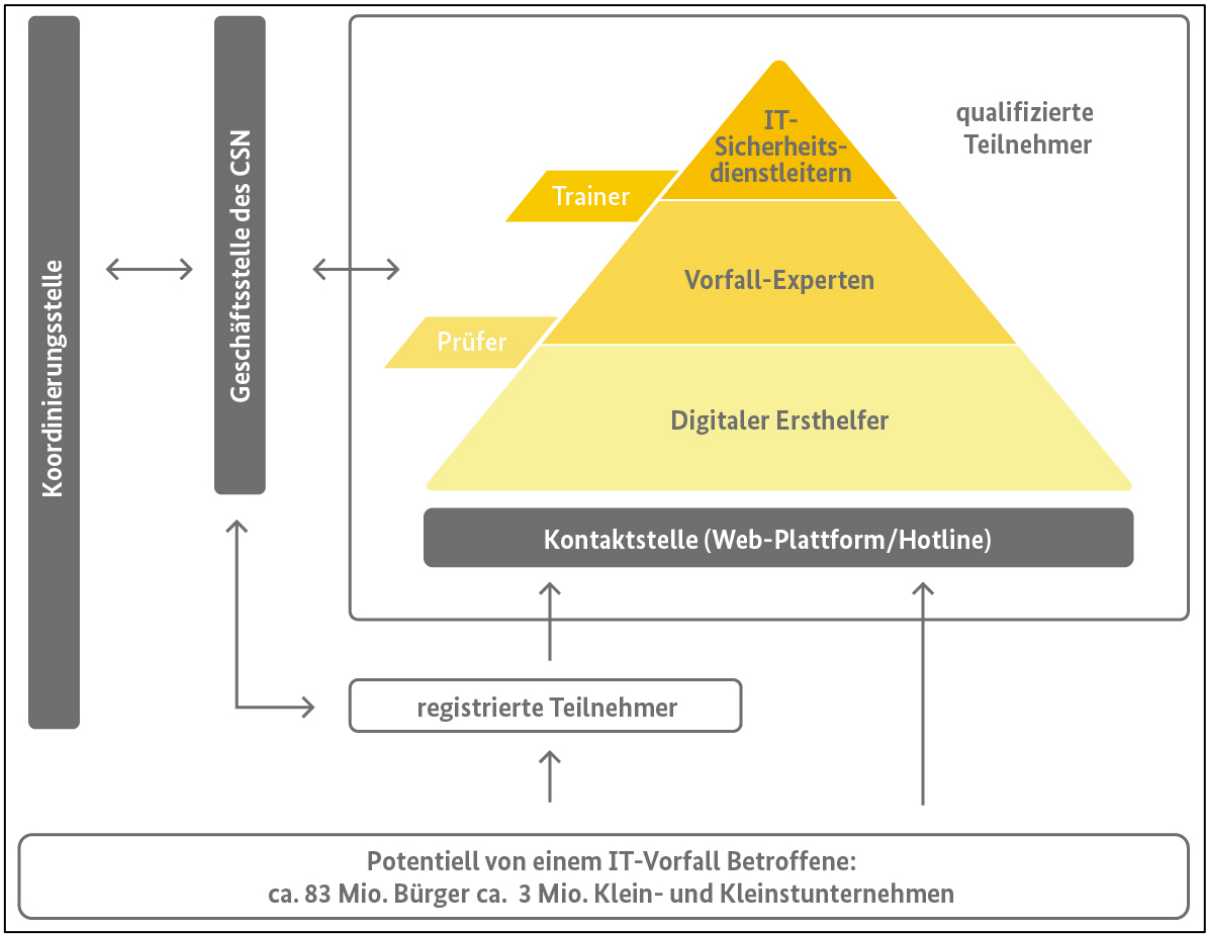


Abbildung 1: Rollen im Cyber-Sicherheitsnetzwerk

Über die Kontaktstelle kann der Betroffene nach einem IT-Sicherheitsvorfall Unterstützung durch qualifizierte Teilnehmer des Cyber-Sicherheitsnetzwerks (Digitale Ersthelfer, Vorfall-Experten oder IT-Sicherheitsdienstleister) erhalten.

Trainer und Prüfer unterstützen das Qualifizierungskonzept und sichern die Qualität des unterstützenden Dienstleistungsangebotes durch Schulung bzw. die Abnahme einer Prüfung.

3. Eine Digitale Rettungskette als Unterstützungsangebot im Notfall

Mit dem Konzept einer Digitalen Rettungskette arbeiten Digitale Ersthelfer, Vorfall-Experten und IT-Sicherheitsdienstleister aufeinander abgestimmt. Sie bilden ein übergreifendes Komplettsystem, welches beginnend mit der Identifizierung, über Hilfestellung bis hin zur umfassenden Lösungsbetreuung und Vorfallklärung eine Kette unterschiedlicher reaktiver Hilfsangebote etabliert.

Bei der Digitalen Rettungskette (siehe Abbildung 3) stehen die folgenden Eskalationsschritte eines IT-Sicherheitsvorfalls in einer aufeinander aufbauenden Reihenfolge:

1. Hilfe zur Selbsthilfe über die Web-Seiten des CSN
2. Notruf an die Kontaktstelle des CSN
3. telefonische Ersthilfe durch Digitalen Ersthelfer
4. Analyse durch den Vorfall-Experten
5. Vor-Ort-Unterstützung durch einen IT-Dienstleister

Jeder dieser Schritte stellt hierbei eine Eskalation des IT-Sicherheitsvorfalls in eine höhere Stufe dar. Es können je nach Eskalation auch einzelne Glieder der Kette in jeder Richtung übersprungen werden.

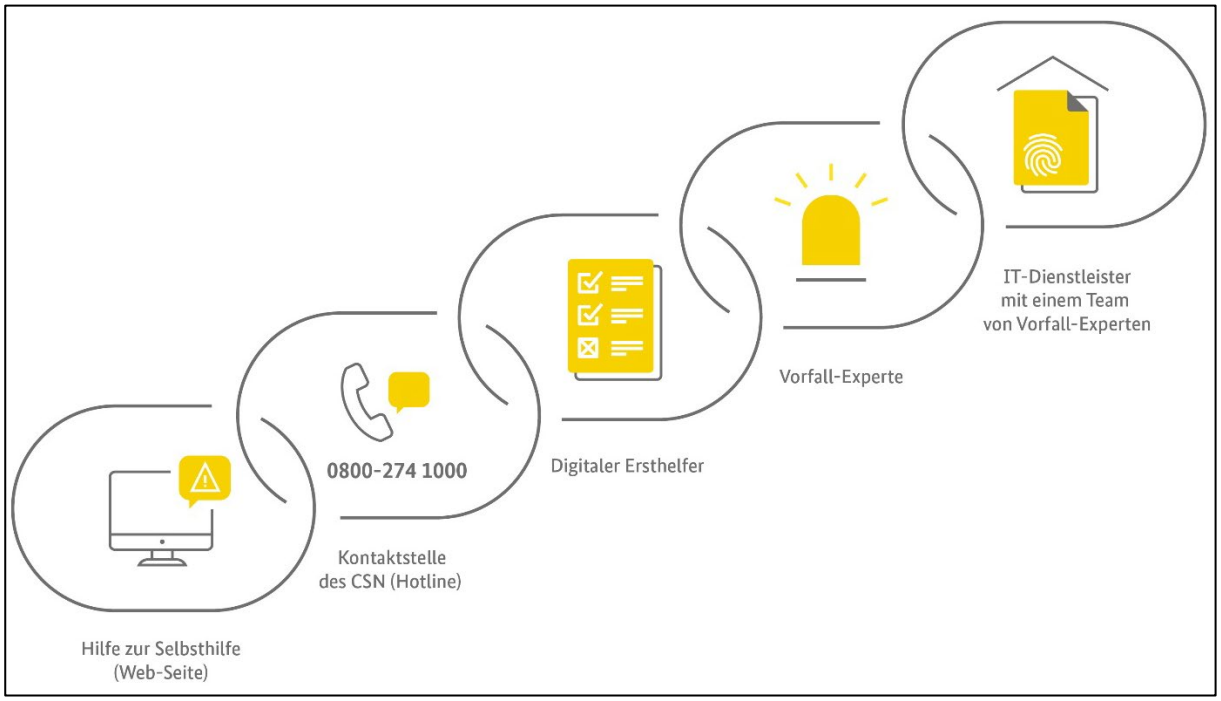


Abbildung 2: Digitale Rettungskette

Erste Anlaufstelle des Cyber-Sicherheitsnetzwerks ist die Kontaktstelle (zentrale Hotline), die dem Betroffenen hilft, den IT-Sicherheitsvorfall einzuschätzen und auf die Liste der Experten verweist.

Die Digitalen Ersthelfer bauen auf der ersten Einschätzung auf und können dann schnelle Ersthilfe leisten. Der Vorfall-Experte übernimmt die telefonische Vorfallbearbeitung vom Digitalen Ersthelfer, wenn dieser mit seiner Ersthilfe das Problem nicht lösen konnte. Ist eine abschließende Bearbeitung eines IT-Sicherheitsvorfall durch die ersten beiden Eskalationsstufen nicht möglich und besteht weiterhin der Bedarf an Unterstützung, so kann auf die Hilfe eines IT-Sicherheitsdienstleisters mit einem Team von Vorfall-Experten zurückgegriffen werden. Zwischen den einzelnen Ebenen besteht eine enge Verzahnung, welche durch eine darüber liegende Kommunikationsstruktur aufgegriffen wird.

4. Konzept für eine qualifizierte einheitliche Unterstützungsleistung

Das Schulungskonzept bildet den Rahmen für die Qualifizierung der Digitalen Ersthelfer und Vorfall-Experten, sowie die Qualitätsanforderungen an die Prüfer und Schulungsanbieter.

In einer Basisausbildung soll ein Erste-Hilfe-Programm vermittelt werden, das für eine grundlegende Unterstützung der Teilnehmer sorgt. Das Erste-Hilfe-Programm kann der Digitale Ersthelfer in einem Online-Kurs und durch Selbststudium des „Leitfadens zur Reaktion auf IT-Vorfälle für Digitale Ersthelfer“ erlernen[3].

Qualifizierte Prüfer kontrollieren die Kompetenz der Digitalen Ersthelfer in einem Prüfungs-Workshop und erteilen nach bestanderer Prüfung ein Testat. Um aktiv das Cyber-Sicherheitsnetzwerk zu unterstützen muss sich der Digitale Ersthelfer registrieren lassen. Nach dieser Registrierung erhält der Digitale Ersthelfer ein Erste-Hilfe-Paket zur Unterstützung seiner Tätigkeit im Cyber-Sicherheitsnetzwerk. Inhalte des Erste-Hilfe-Paketes ist der „Leitfaden zur Reaktion auf IT-Sicherheitsvorfälle - Digitale Ersthelfer“, der die Leitplanken seiner Tätigkeit festlegt, sowie Flyer und andere Hilfsmittel.

Abbildung 3*:* Qualifizierungsprogramm des Cyber-Sicherheitsnetzwerks

In einer Aufbauschulung für Vorfall-Experten vermitteln qualifizierte Trainer zusätzliche Inhalte für die Vorfallanalyse und -behandlung.

Aufgabe des Schulungsanbieters ist es, auf der Grundlage des im Curriculum festgelegten Lehrinhaltes eigenständig eine Aufbauschulung für Vorfall-Experten zu konzipieren.

Die Aufbauschulung umfasst mindestens drei Tage mit mindestens zwanzig Unterrichtseinheiten.

[3] Der Online-Kurs für Digitale Ersthelfer steht voraussichtlich ab März 2021 kostenlos zu Verfügung.

Folgende Themenfelder werden in der Aufbauschulung bearbeitet:

- Einführung und Rahmenbedingungen für den Vorfall-Experten
- Ablauf des Standardvorgehens mit Übungen
- Remote-Unterstützung
- Angriffsszenarien und Sofort- bzw. Gegenmaßnahmen mit Übungen
- Vor-Ort-Unterstützung: „Überblick verschaffen“ und Übungen
- Vor-Ort-Unterstützung: „Analyse“ und Übungen
- „Nach einem Vorfall ist vor einem Vorfall"- präventive Maßnahmen

Durch ein standardisiertes Schulungsprogramm ist eine breite Basis geschaffen, um ein Qualifizierungsprogramm für Vorfall-Experten anzubieten. Schulungsanbieter können IT/Informationssicherheits-Schulungsanbieter, Verbände oder Universitäten sein, die Digitale Ersthelfer, Mitglieder, Studenten aber auch Mitarbeiter von Unternehmen schulen, um bei der Vorfallbearbeitung im Rahmen des Cyber-Sicherheitsnetzwerkes in geeigneter Art und Weise zu unterstützen. Qualifizierte Schulungsanbieter konzipieren eine Aufbauschulung für Vorfall-Experten, um handlungsfähige Vorfall-Experten aus- bzw. weiterbilden zu können[4].

Ziel ist es, möglichst viele Experten (Digitale Ersthelfer und Vorfall-Experten) zu qualifizieren und für die aktive Mitarbeit im Cyber-Sicherheitsnetzwerk zu gewinnen.

Erfahrene Vorfall-Experten können nach dem Prinzip: Aus der Praxis für die Praxis als Trainer und Prüfer im Cyber-Sicherheitsnetzwerk tätig werden.

So kann innerhalb kürzester Zeit ein flächendeckendes Netzwerk von Experten geschaffen werden.

Diese Experten können dann zielgruppengerecht eine reaktive Unterstützung nach IT-Sicherheitsvorfällen leisten, um KMU und Bürger in geeigneter Art und Weise zu unterstützen.

5. Das Zertifizierungsverfahren im Cyber-Sicherheitsnetzwerk

Zur Durchführung von Vorfallbehandlungen im Rahmen der Digitalen Rettungskette des Cyber-Sicherheitsnetzwerkes werden qualifizierte Personen benötigt. Ziel des Verfahrens ist es, kompetente Personen bereitzustellen sowie die Qualität und Vergleichbarkeit der Vorfallbehandlung sicherzustellen.

Es gibt unterschiedliche Formen der Qualifikation für Personen bzw. IT-Dienstleiter die im Cyber-Sicherheitsnetzwerk aktiv sind:

- Bronze-Level: Testat zum Digitalen Ersthelfer auf der Grundlage eines Online-Kurses sowie einer bestandenen halbtägigen Prüfung[5]
- Silber-Level: Personenzertifikat als Vorfall-Experte auf der Grundlage einer Aufbauschulung sowie einer Prüfung durch das BSI

[4] Eine Liste aller beim Cyber-Sicherheitsnetzwerk registrierten Anbieter für die Aufbauschulung zum Vorfall-Experten befindet sich derzeit noch im Aufbau.

[5] Die halbtägigen Prüfungen werden von qualifizierten Prüfern z.B. zertifizierten Vorfall-Experten durchgeführt.

- Gold-Level: Qualifizierung als IT-Dienstleister mit einem Team von zertifizierten Vorfall-Experten

5.1 Anforderungen an die Qualifikation des Digitalen Ersthelfers

Die Digitale Rettungskette ist insgesamt nur so stark wie ihr schwächstes Glied. Daher gilt es, den Digitalen-Ersthelfer ausreichend zu schulen und zu unterstützen, so dass er eine flächendeckende Erstversorgung übernehmen kann. Damit die Qualität dieser Ersthilfe gewährleistet ist, haben die Digitalen Ersthelfer einen Online-Kurs zu absolvieren und eine theoretische und praktische Prüfung zu bestehen. Die Prüfung bzw. Schulung der Digitalen-Ersthelfer bzw. Vorfall-Experten übernehmen erfahrene Trainer des CSN. Damit ein Digitaler-Ersthelfer seine Aufgabe im CSN verlässlich und kompetent durchführen kann, gibt das CSN einen Rahmen für die Qualifikation eines Digitalen Ersthelfers vor.

Folgende Voraussetzungen sollte der Digitaler Ersthelfer mitbringen:

- grundlegendes Verständnis von IT-Systemen und Netzwerken
- Überblick über die Gefährdungslage
- Unterschiede Server / Clients / dedizierte Hardware
- Unterschiede in Betriebssystemen
- Grundlagen zu Internet, Mail, Webseiten
- Kenntnisse über Bedrohungen, Schwachstellen und Gefährdungen in der IT

Im Bereich soziale bzw. persönliche Kompetenzen werden folgende Voraussetzungen angenommen

- Einsatzbereitschaft
- Zeitmanagement
- Sozialverhalten z.B. ruhiges und beruhigendes Auftreten am Telefon
- Verantwortungsbewusstsein

Digitale Ersthelfer müssen vor Aufnahme ihrer Tätigkeit zunächst nachweisen, dass sie über ausreichende technische bzw. fachliche sowie persönliche und soziale Kompetenzen verfügen. Sie müssen sich außerdem im CSN registrieren.

5.2 Anforderungen an die Qualifikation des Vorfall-Experten

Vorfall-Experten müssen neben den persönlichen Eigenschaften auch Erfahrung bei der Vorfall-Behandlung mitbringen. Daher müssen sie folgende Nachweise (durch Vorlage von entsprechenden Zeugnissen oder Projektbescheinigungen) gegenüber der Zertifizierungsstelle erbringen. Dies sind unter anderem:

- Berufsabschluss
- Berufserfahrung
  - im Bereich IT-/Informationssicherheit in den letzten 5 Jahren und
  - mindestens drei Jahre Erfahrung bei der IT-Vorfall-Behandlung
- Praxiserfahrung aus Projekten im Gesamtumfang von mindestens 40 Personentagen mit dem Schwerpunkt
  - Behandlung von IT-Vorfällen oder

- Erstellung von forensischen Werkzeugen bzw. Analyse-Tools oder
- Durchführung von Penetrationstests (Detektion) oder Beratung (Prävention)

5.3 Anforderungen an den IT-Sicherheitsdienstleiter

IT-Sicherheitsdienstleister des Cyber-Sicherheitsnetzwerks sind größere IT-Dienstleister, die überregional agieren und eine größere Anzahl von zertifizierten Vorfall-Experten bereitstellen können. So können sie bei komplexeren und größeren IT-Sicherheitsvorfällen ein Team aus Vorfall-Experten mit speziellen Kenntnissen und Fähigkeiten für die Vorfallbehandlung anbieten.

IT-Sicherheitsdienstleister mit einem Team für das CSN können:

- zertifizierte IT-Sicherheitsdienstleister für den Geltungsbereich „Vorfall-Experten“ oder
- qualifizierte Dienstleister für ein Spezialgebiet z. B. DDOS oder APT sein.[6]

5.3.1 Zertifizierung IT-Sicherheitsdienstleister

IT-Sicherheitsdienstleister verfügen über ein Team von mindestens drei zertifizierten Vorfall-Experten. Für die Zertifizierung als IT-Sicherheitsdienstleister beim BSI ist ein Nachweis der Erfüllung der Anforderungen der Norm DIN EN ISO/IEC 17025:2018 (Referenzkatalog) erforderlich.

Zusätzlich muss der IT-Dienstleister nachweisen, dass er mindestens für den Geltungsbereich ein Informationsmanagementsystem (ISMS) mit einem Sicherheitskonzept auf der Basis des IT-Grundschutzes [1] hat.

Damit zeigt der IT-Sicherheitsdienstleister, dass

1. ein funktionierendes IS-Management vorhanden ist,
2. ein definiertes Sicherheitsniveau erreicht wird.

Die Einhaltung der Bestimmungen zum Umgang mit Verschlusssachen gemäß „Verfahrensbeschreibung zur Anerkennung von Prüfstellen und Zertifizierung von IT-Sicherheitsdienstleistern“ [2] sowie die Bereitschaft zur Aufnahme in die Geheimschutzbetreuung des Bundes, sind weitere Voraussetzung für die Zertifizierung des IT-Sicherheitsdienstleisters.

5.3.2 Qualifizierte Sicherheitsdienstleister

Qualifizierte Sicherheitsdienstleister haben sich anhand eines vom BSI festgelegten einheitlichen Auswahlverfahrens qualifiziert und mussten dafür gegenüber dem BSI nachweisen, dass sie in ihrem jeweiligen Tätigkeitsgebiet ein hohes Spezialwissen erlangt haben. Zum Qualifikationsnachweis müssen gegenüber dem BSI festgelegte Kriterien erfüllt und bestimmte Kompetenzen in Form eines Prüfungs-Interviews nachgewiesen werden.

Eine Aufstellung einzelner Leistungsmerkmale der qualifizierten DDoS-Mitigation-Dienstleister [3] oder der APT-Response-Dienstleister [4] sind veröffentlicht. Damit

[6] Derzeit gibt es, im Sinne § 3 BSIG bereits qualifiziert IT-Dienstleister für die Bereiche DDOS-Mitigation und APT.

wird für den Betroffenen die Auswahl eines geeigneten Dienstleisters erleichtert. Kontinuierlich wird das Angebot um weitere Anforderungskataloge für qualifizierte Sicherheitsdienstleister erweitert.

Das BSI veröffentlicht regelmäßig die qualifizierten Dienstleister sowie die Auswahlkriterien. [3][4]

5.4 Sonderqualifikation: Der Vorfall-Experte im Unternehmen

Auch für Unternehmen macht es Sinn, ihre Mitarbeiter auf den Notfall vorzubereiten. Eine entsprechende Qualifikation für mindesten einen IT-affinen Mitarbeiter des Unternehmens durch den Besuch der Aufbauschulung wird daher empfohlen. So verfügt das Unternehmen über Kenntnisse wie die Digitale Rettungskette aufgebaut ist und lernen das Vorgehen bei der Vorfall-Behandlung kennen.

Um nach einen IT-Sicherheitsvorfall möglichst schnell und effektiv reagieren zu können agiert dieser Vorfall-Experte des Unternehmens als kompetente Schnittstelle zu den Vorfall-Experten des CSN. Außerdem ist der Vorfall-Experte erster Ansprechpartner im Unternehmen für die Experten des CSN bei einem größeren Schadensfall.

Zusätzlich können Kontakte zu anderen Vorfall-Experten des CSN oder anderer Unternehmen aufgebaut werden, um ein eigenes Netzwerk zu schaffen und sich vertrauensvoll zu unterstützten.

Nach erfolgreichem Besuch der Schulung ist eine Registrierung beim CSN als „Unternehmen-Vorfall-Experte“ möglich. Erfüllt der Vorfall-Experte die Anforderungen der Personenzertifizierung kann er sich zusätzlich auch zertifizieren lassen und dann aktiv im CSN als Vorfall-Experte andere unterstützen.

Das Unternehmen kann gegenüber Externen so transparent machen, dass das Thema Notfallvorsorge für das Unternehmen ein wichtiges Thema ist.

Zusätzlich werden die Vorfall-Experten zu entsprechenden Erfahrungsaustausch-Foren eingeladen, um ihr Wissen kontinuierlich aufrecht zu erhalten und sich über die neusten Angriffsformen auszutauschen. Durch die Einladung zu regionalen Foren besteht die Möglichkeit, in einer gesicherten Umgebung die Behandlung realer IT-Vorfälle zu üben.

6. Ausblick

In einem ersten Schritt haben bereits aktive Experten zusammen mit Vertretern von Behörden, Bildungsinstitutionen und unterschiedlichen Interessensgruppen das Bundesamt für Sicherheit in der Informationstechnik unterstützt, die strategische Ausrichtung sowie die Rahmenbedingungen für das Cyber-Sicherheitsnetzwerk zu schaffen.

Kernbestandteil des Cyber-Sicherheitsnetzwerkes ist das Netzwerk aus Experten, die je nach Eskalationsstufe der Digitalen Rettungskette tätig werden. Die Außenwahrnehmung steht und fällt mit den eingesetzten Experten. Deshalb wird im zweiten Schritt der Fokus auf die Qualifikation und Zertifizierung der Experten gelegt. Als Ergebnis müssen hierbei ausreichend viele Digitale Ersthelfer und Vorfall-Experten sowie IT-Sicherheitsdienstleister mit einem Team von Vorfall-Experten im Netzwerk geschult und registriert sein sowie aktiv werden.

In einem dritten Schritt wird das Konzept der Digitalen Rettungskette in einem sechsmonatigen Pilotbetrieb erprobt. Einige Herausforderungen werden sich erst in der eigentlichen Pilotphase zeigen und dann adäquat gelöst werden. Die Pilotphase schließt mit einer Evaluierung ab, in der Empfehlungen für die nächsten Schritte zusammengefasst werden und die Planung für eine deutschlandweite Einführung der Unterstützungsdienstleistung durch das Cyber-Sicherheitsnetzwerk erfolgt.

Literaturhinweise

[1] BSI-Standard 200-2 IT-Grundschutz-Vorgehensweise, BSI, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/BSI_Standards/standard_200_2

[2] Verfahrensbeschreibung zur Anerkennung von Prüfstellen und Zertifizierung von IT-Sicherheitsdienstleistern, BSI, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Zertifizierung/VB-Stellen.html?nn=10023142

[3] Kriterien für qualifizierte Dienstleister: DDoS-Mitigation-Dienstleister, BSI, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/Dienstleister-DDos-Mitigation.html

[4] Kriterien für qualifizierte Dienstleister: APT-Response-Dienstleister, BSI, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/Auswahlkriterien_APT-Response_Dienstleister.html

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Multifunktionale und sichere Edge-Architektur für digital transformierte Industrie und Kritische Infrastrukturen

Jan Tiedemann[1], Peter Rost[1], Annegrit Seyerlein-Klug[1], Jens Kulikowski[1]

Kurzfassung:
Im Rahmen der Digitalisierung bringt die Vernetzung industrieller cyber-physikalischer Systeme mit zentralen IT-Anwendungen große wirtschaftliche Chancen, aber auch vielfältige Herausforderungen an die Cybersicherheit mit sich. Wir stellen eine industrietaugliche, multifunktionale, sichere und vertrauenswürdige Edge-Architektur für Anwendungsfälle im Industrial Internet of Things (IIoT) vor. Erfahrungen aus der praktischen Anwendung der Architektur und ein Ausblick beschließen den Beitrag.

Stichworte: Cyber-Physikalische Systeme, Digitale Transformation, Digitalisierung, Edge Computing, Edge Gateway, Industrial Internet of Things (IIoT), Industrie 4.0, Kritische Infrastrukturen

1. Digitale Transformation in Kritischen Infrastrukturen und Industrie

Die Digitalisierung bzw. digitale Transformation von Wirtschaft und Infrastruktur ist in vollem Gange. Digitalisierung bedeutet hier erstens die „Digitalmachung“ (engl. *digitization*) von Anlagen, Maschinen, Aktoren und Sensoren – kurz: Dingen, die bisher weder nach Stand der Technik[2] digitale Daten erzeugen noch diese verarbeiten konnten. Diese Digitalmachung ist die Fortsetzung der dritten Industriellen Revolution oder „Digitalen Revolution“, die bereits Mitte des 20. Jahrhunderts begann[3]. Zweitens umfasst die digitale Transformation die Vernetzung bisher nicht oder nur manuell vernetzter Dinge. Drittens bedeutet Digitalisierung auch die Optimierung digitaler und vernetzter Systeme durch immer komplexere Anwendungen bis hin zur Autonomie durch künstliche Intelligenz oder deren Vorstufen. Vision und Ziel der Aktivitäten rund um die digitale Transformation in der Wirtschaft ist es, die Industrie 3.0 in eine Industrie 4.0 (bzw. Infrastrukturen 4.0) zu transformieren. Hier spricht man auch von der Vierten Industriellen Revolution (vgl. Abb. 1).

Im Jahr 2020 hat die digitale Transformation noch einmal besonderen Schwung aufgenommen. Die Pandemie-bedingten Kontaktbeschränkungen haben deutlich gemacht, wie sehr digitalisierte Produktions-, Betriebs- und Geschäftsprozesse nicht nur Kosten senken und Innovations- sowie Produktivitätsvorteile bieten, sondern auch die Krisenfestigkeit von Gesellschaft, Infrastruktur und Wirtschaft steigern.

2. Chancen und Risiken

Die digitale Transformation ist jedoch eine Entwicklung mit zwei Gesichtern. Auf der einen Seite zeigen sich die Chancen. Innovationsfähigkeit durch die Erschließung neuer

---

[1] secunet Security Networks AG, Essen. Kontakt über https://www.secunet.com

[2] https://www.bsi.bund.de/DE/Themen/KRITIS-und-regulierte-Unternehmen/Kritische-Infrastrukturen/Allgemeine-Infos-zu-KRITIS/Stand-der-Technik-umsetzen/stand-der-technik-umsetzen_node.html

[3] Siehe auch: https://stfc.ukri.org/files/digital-revolution-infographic/

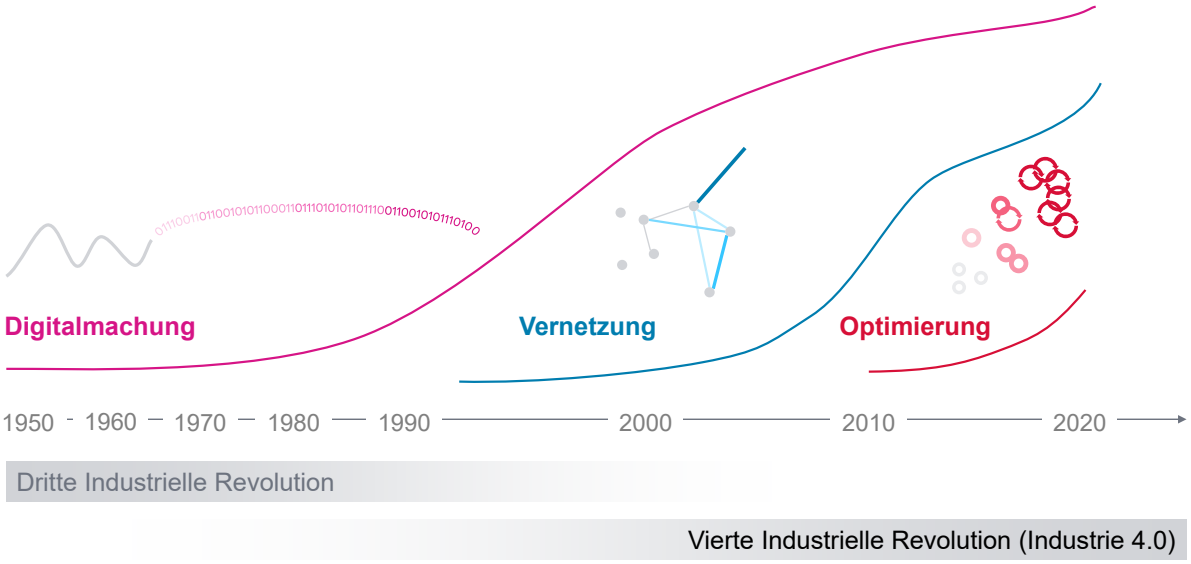


Abbildung 1: Digitalisierung – von der Dritten zur Vierten Industriellen Revolution

Möglichkeiten ist ein wichtiger Treiber.[4] Erst durch digitale Prozesse werden viele innovative Entwicklungen möglich.[5] Noch wichtiger, insbesondere für deutsche Unternehmen[6], sind durch digitalisierte Prozesse gesteigerte Effizienz und reduzierte Kosten. Mit digitaler Technologie ist es möglich, die industrielle Produktion effizient und kostenoptimiert aus der Ferne kontrollier- und steuerbar zu machen. Erforderliche Wartungsmaßnahmen können mittels datenbasierter Verschleißprognosen kurz vor dem erwarteten Ausfalltermin erkannt und automatisiert-präventiv eingeplant werden. Abläufe werden somit reibungsärmer, und der Fachkräftemangel wird entschärft.[7]

Auch die Flexibilität der Produktionsprozesse lässt sich steigern. So können Veränderungen nicht erst mit der nächsten großen Iteration, sondern bereits mit der nächsten Schicht digital eingespielt werden. Damit werden optimierte Produkte und Dienstleistungen möglich, wenn beispielsweise ein Produzent innerhalb kürzester Zeit auf die Anforderungen seines Kunden reagieren kann. In der Konsequenz kann jedes Werkstück ein Unikat werden (Schlagwort: „Losgröße 1“[8]). Durch die Digitalisierung werden zudem verbesserte Monetarisierungsmöglichkeiten über eine datenschutzkonforme Nutzung der anfallenden Daten geschaffen. Zu guter Letzt steigt durch die digitale Transformation idealerweise auch die Customer Experience (Vgl. Abb. 2).

Den Chancen stehen jedoch auch Gefahren und Risiken gegenüber. Auf alles was vernetzt ist, kann zugegriffen werden. Erfahrung und Praxisdaten[9] zeigen, dass neben not-

4 https://www.ptc.com/de/industry-insights/digital-transformation

5 Umgekehrt ergibt sich aber auch ein Risiko, denn wer bei der Digitalisierung den Anschluss verpasst, den wird die digitale Disruption über den Markt strafen.

6 https://www.idc.com/getdoc.jsp?containerId=prEUR145546419

7 https://www.bayika.de/de/aktuelles/meldungen/2020-11-26_Internationale-Studie-Digitalisierung-2020.php

8 https://www.efpf.org/post/lot-size-one-is-bigger-than-what-you-might-think

9 https://www.bsi.bund.de//Lageberichte/

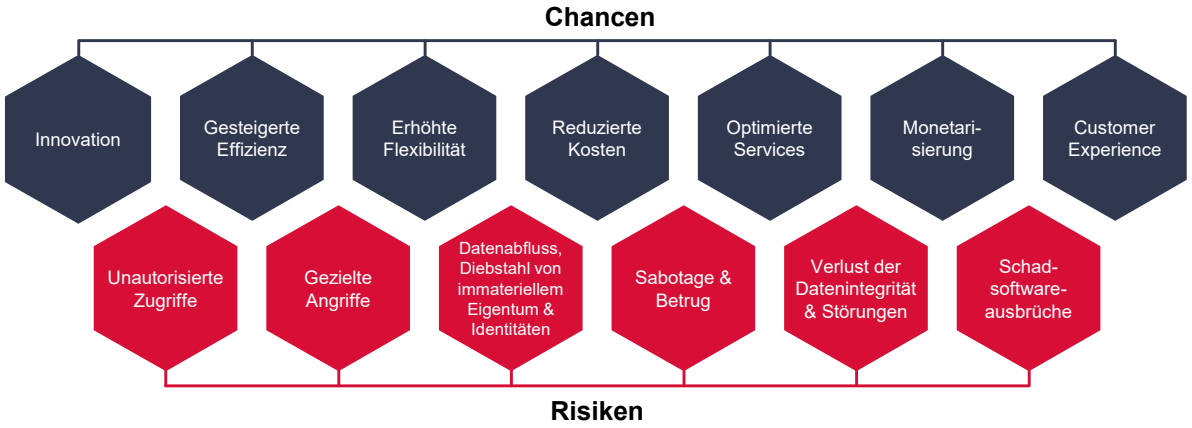


Abbildung 2: Chancen und Risiken der digitalen Transformation

wendigen und gewollten Zugriffen meist auch ungewollte Zugriffe oder sogar gezielte Angriffe möglich sind. Wie diverse Fälle von Industriespionage demonstriert haben, steigt mit einer zunehmenden Vernetzung auch das Risiko von Datenabflüssen mit Diebstahl von immateriellem Eigentum oder mit kriminellem Missbrauch personenbezogener Daten. Welche Schäden hier lauern, muss nach dem erneuten Cybervorfall-Rekordjahr 2020 nicht erneut im Detail aufbereitet werden.

Im Bereich der Kritischen Infrastrukturen wiegen die Folgen von Cyberangriffen, Sabotage und Störungen besonders schwer. Insbesondere aus der Büro-IT bekannte Verschlüsselungstrojaner können mit zunehmender Anbindung auch die Produktionswelt erreichen und dort auf ungeschützten Systemen ausbrechen.

3. Technologische Basis für digitale Transformation

Die digitale Transformation erfordert einen passenden Technologie-Stack. Ein Technologie-Stack beschreibt die übereinander liegenden Komponenten-Schichten und ihre Elemente, auf deren Basis Anwendungsfälle realisiert werden können. Die enthaltene Technologie muss angesichts der Chancen und Risiken nicht nur erforderliche Funktionalität bieten, sondern auch notwendige Cybersicherheit.

Der Wunsch nach digitaler Transformation trifft oft auf bestehende, heterogene Betriebs- und Produktionsmittel mit geringem digitalem Reifegrad, so genannte Brownfield-Umgebungen. Vollständig neu aufgesetzte Umgebungen, die beliebig mit neuer Technologie ausgestattet werden können (sog. Greenfields), sind gerade in der bereits seit längerem erfolgreichen deutschen Industrie die Ausnahme. Im Regelfall stößt die digitale Transformation auf einen Mix aus älterer, unzureichend digitalisierter und moderner, ab Werk vernetzter Technologie. Der Technologie-Stack muss alle diese Varianten integrieren und schützen können.

Ob die digitale Transformation zur Industrie 4.0 funktional erfolgreich und resistent vor Cyberangriffen sein kann, entscheidet sich insbesondere in der Brandungszone an den Rändern der Produktionsnetzwerke, der Grenze zwischen Physik und Cyberraum – den Cyber-Physikalische Systemen (CPS). CPS sind z. B. Maschinen, Sensoren, Kameras oder Anlagen, die sowohl Mechanismen ausführen oder Messwerte erfassen, als auch

mit digitaler Funktionalität wie Rechenleistung und Konnektivität ausgestattet sind. Aus dem Blickwinkel der Informationstechnologie handelt es sich dabei um den Rand der Netzwerke, das „Edge“. Für „Edge“ bzw. „Edge Computing“ gibt es bereits in mehreren Bereichen Definitionen – z. B. in der Telekommunikationsbranche. In diesem Beitrag ist mit „Edge“ das „Industrial Edge“ gemeint, also der Rand der Produktionsnetze direkt vor den CPS und der physischen Welt (vgl. Abb. 3).

An dieser Stelle kommt industriellen Edge-Gateways besondere Bedeutung zu. Diese Systeme stellen Verbindungen zwischen den IT-Domänen und der cyberphysikalischen Welt her. Um die erforderliche Flexibilität und Offenheit für Erweiterungen durch Dritte zu ermöglichen, empfiehlt es sich, die Architektur dieses Edge-Gateways als multifunktionale und sicherheitsgehärtete Plattform zu gestalten.

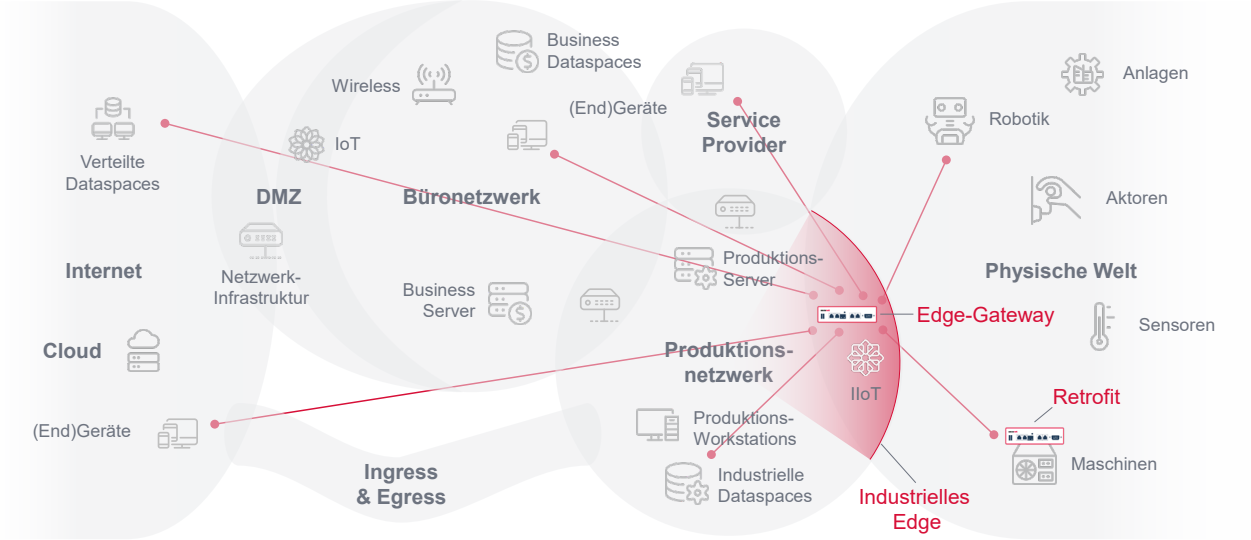


Abbildung 3: Platzierung der Industriellen Edge- und Retrofit-Gateways in einer stark vereinfachten Netzwerkdarstellung der digitalisierten Industrie[10]

4. Digitalisierungs- und Cybersecurity-Retrofit

Eine zur digitalen Transformation passende Architektur muss somit Funktionalität bieten, die eine Digitalmachung, Vernetzung und Optimierung ermöglicht, und sie muss Sicherheit vor Digitalisierungsrisiken bieten. Da sie dies nicht nur auf der grünen Wiese (Greenfield), sondern auch in gewachsenen Umgebungen (Brownfields) können muss, kommt dem Nachrüstungsfall (Retrofit) besondere Bedeutung zu.

Über einen funktionalen Digitalisierungs-Retrofit werden analoge, nicht vernetzte Geräte zu Cyber-Physikalischen Systemen (CPS) aufgerüstet:

[10] Eine direkte Verbindung von Produktionsnetzwerk und Internet ist oft nicht oder nur sehr eingeschränkt vorgesehen. Der Aspekt „Ingress & Egress“ (Kommunikation vom Internet ins Produktionsnetz und umgekehrt) soll sowohl diese Ausnahmen illustrieren, als auch aufzeigen, dass der Ist-Zustand manchmal vom Soll-Zustand abweicht.

- „Digitalmachung ermöglichen“ heißt, dass der bestehende Gerätepark (das Brownfield) entweder digital ausgestattet wird (z. B. indem digitale, vernetzte Sensoren an eine analoge Maschine angebracht werden) oder dass seine digitalen Fähigkeiten auf einen aktuellen Stand gehoben werden, z. B. weil die bisherige Rechenleistung für neue Funktionalität wie maschinelles Lernen nicht ausreicht.
- „Verbindung bzw. Vernetzung ermöglichen“ bedeutet, existierende Technik, die nur industriespezifische oder veraltete Busse und Protokolle unterstützt, mit aktueller Konnektivität nachzurüsten. Zudem müssen die diversen Industrieprotokolle, Maschinen-Befehlssätze und Datentypen in standardisierte, kompatible Formate übersetzt und herstellerübergreifend konsolidiert werden, damit sie als homogene Daten („Dataspace“) ausgewertet werden können.
- Aus der „Optimierung“ folgt, dass eine Edge Architektur eine flexible Basis für heutige und künftige Anwendungen bieten soll.

Auch ein Sicherheits-Retrofit muss möglich sein. Anders als bei IT-Anwendungen, bei denen Cyberangriffe vorrangig wirtschaftliche Schäden verursachen, können fehlgesteuerte CPS potenziell katastrophale Folgen für Leib und Leben von Menschen haben und müssen daher über eine Reihe dedizierter Cybersicherheitsmaßnahmen vor Missbrauch und Sabotage geschützt werden. Insbesondere langlebige Industrieanlagen wurden zum Herstellungszeitpunkt oft noch nicht ab Werk vor den heutigen Gefahren geschützt. Für den Sicherheits-Retrofit werden heute meist dedizierte Industrie-Firewall-Appliances oder VPN-Gateways direkt vor den CPS eingesetzt. Für neu hinzukommende Sicherheitsfunktionalität und bei höheren Leistungsanforderungen muss dann zusätzliche Hardware ins Feld verbracht werden. Hier bieten sich Edge-Gateways als multifunktionale, industrietaugliche und erweiterbare Systeme für zukunftssichere, effiziente und langlebige Lösungen an.

5. Anforderungen an eine multifunktionale und sichere Architektur

Eine Edge-Architektur, die Retrofits ermöglichen soll, den Stand der Technik abdeckt und zukunftssicher ist, muss verschiedene Anforderungen erfüllen. Grundlage für die Entwicklung unserer Architektur waren die in der folgenden Tabelle 1 dargestellten, unterschiedlichen Anforderungen aus den Bereichen Digitalisierung und Cybersicherheit im Industriellen Edge bzw. für das Industrielle Internet der Dinge (IIoT). Diese Anforderungen ergaben sich aus den in den vorherigen Kapiteln dargestellten Überlegungen und wurden durch Recherchen und Kundenworkshops verfeinert. Die Anforderungen sind nach den Kategorien der Digitalisierung (Digitalmachung, Vernetzung und Optimierung) und Sicherheit sowie den Horizonten „Aufholen“, „Mithalten“ und „Zukunftssicherheit“ eingeordnet. Einige der Anforderungen können sekundär auch in anderen Bereichen relevant sein, selbst wenn dies nicht explizit aufgeführt ist (insbesondere bezüglich der Sicherheitsaspekte).

	Aufholen **Retrofit**	*Mithalten* **Stand der Technik**	*Vorausgehen* **Zukunftssicherheit**
Digitalmachung	• General Purpose Computing Ressourcen zum Nachrüsten für alle Dinge • Anschluss an CPS-Sensoren und Aktoren	• High Performance Computing am Edge (im Vergleich zu sonst dort vorhandenen, ggf. begrenzten Ressourcen)	• Spezialisierte, anspruchsvolle Computing-Möglichkeiten am Edge (z. B. KI-Beschleunigung)
Vernetzung	• Ältere Interfaces und Protokolle anbinden (COM-Ports, Feldbusse) • Ethernet unterstützen • Wireless, wo keine Kabel möglich sind	• Neue Protokollstandards unterstützen (wie z. B. MQTT oder spezialisierte Industrieprotokolle) • Netzwerksegmentierung z. B. nach Geräteklassen ermöglichen	• Zero Trust Netzwerk-Architekturen (ZTNA) ermöglichen
Optimierung	• CPS mit Produktionssteuerungssystemen verbinden	• Anbindung an Business-Anwendungen (inkl. Cloud)	• Zukünftige Anwendungen oder Agenten vor Ort ausführen
Sicherheit	• Ältere oder „eingefrorene“ Systeme von neuen Bedrohungen abschirmen • Maschinenfirewall zum Nachrüsten • Sichere VPN-Zugriffe für Hersteller und Dienstleister ermöglichen • Verschlüsselungs-methoden, PKI-Infrastrukturen und Zertifikate unterstützen	• Network Access Control (NAC) ermöglichen • Überwachung auf bekannte Bedrohungen und Störungen • Konvertierung alter CPS-seitiger Protokolle in IT-seitig aktuelle, sichere Äquivalente • Sichere Enklaven schaffen • Sichere Cloudanbindung ermöglichen • Patching & Updates orchestrieren	• Traffic Monitoring und Anomalieerkennung auf komplexe und neuartige Störungen und Bedrohungen mit Alarmierung (System zur Angriffserkennung) • Krypto-Agilität (u.a. zur Nachrüstung von Post-Quanten-Kryptografie)

Tabelle 1: Anforderungen an eine multifunktionale und sichere Edge-Architektur (Auswahl)

Aus diesen Anforderungen lässt sich eine grundsätzliche Aufgabe ableiten. Im Sinne der agilen Entwicklung kann diese als User Story[11] formuliert werden:

„Als digitaler Transformierer will ich eine Edge-Architektur einsetzen, um funktional flexibel und cybersicher Digitalisierungs-Anwendungsfälle zu betreiben“.

„Digitale Transformierer“ sind in diesem Sinne alle Planer, Entscheider und Benutzer, die in Industrie und Kritischen Infrastrukturen mit der digitalen Transformation befasst sind. Üblicherweise sind dies Produktions- oder Anlagenmanager, Produktmanager, IT- und OT-Manager oder IT- bzw. OT-Sicherheitsverantwortliche sowie die Geschäftsleitung.

[11] Eine User Story ist eine kurze Beschreibung (Story) dessen, was ein Benutzer (User) will. (https://scrumguide.de/user-story/)

6. Edge-Architektur: Stack und Umsetzung

Die Grundlage unserer Architektur-Entwicklung war die Erfüllung der User Story mit aktuellen technologischen Mitteln. Als Vorbild diente uns der Cloud-Infrastruktur-Stack, der für die Edge-Architektur ein fast identisch einsetzbares Technologie-Modell von der Infrastruktur bis hin zu Anwendungsfällen liefert (vgl. Abb. 4).

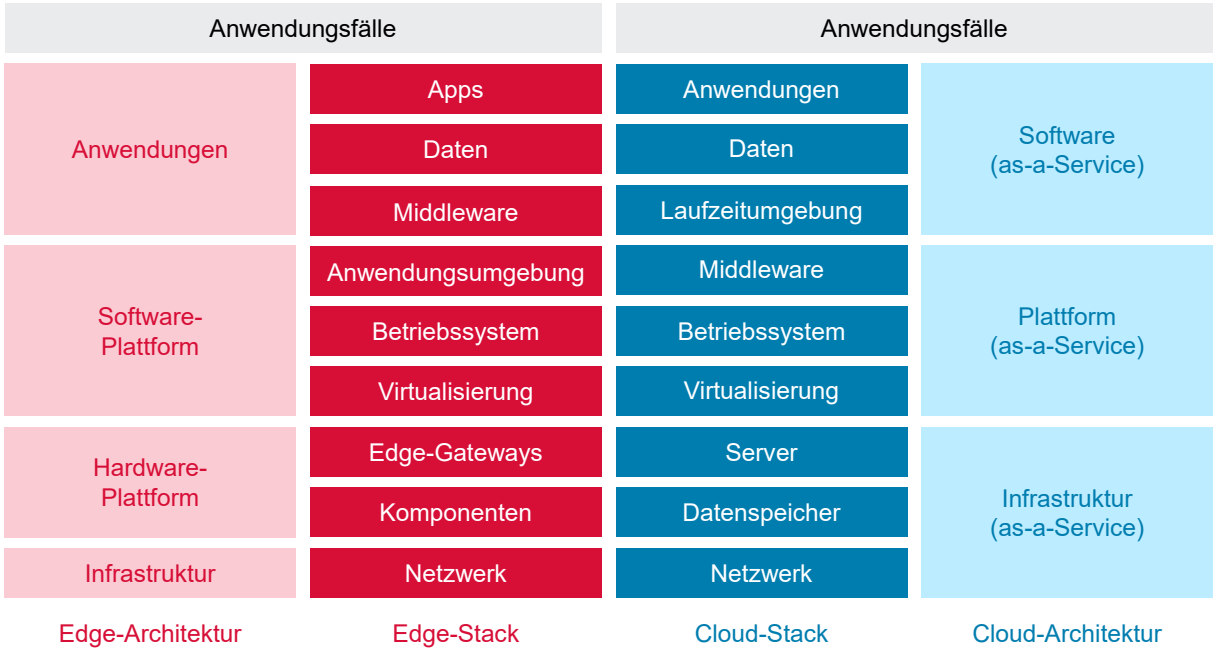


Abbildung 4: Architektur und Technologie-Stacks von Edge und Cloud im Vergleich[12]

Im Unterschied zur Cloud ist die **Infrastruktur** in Form des Netzwerks bei Edge-Anwendungsfällen meist bereits gegeben. Die Edge-Architektur muss damit in bestehende Netzwerk-Infrastrukturen eingebunden werden können. In bestimmten Fällen kann die Edge-Architektur aber auch selbst als Bridge, Router bzw. Gateway die Netzwerk-Infrastruktur (mit) aufbauen. Zu diesem Zweck ist eine Ausstattung spezifischer Edge-Gateways mit mehr als zwei Netzwerkschnittstellen hilfreich.

Die **Hardware-Plattform** der Edge-Architektur umfasst Edge-Gateways sowie spezielle Hardware-Komponenten zum Einbau oder in Verbindung mit diesen Gateways (im Cloud-Stack fällt dies noch unter Infrastruktur). In der Praxis werden als Gateways unterschiedlich leistungsfähige und robuste Hardwarevarianten eingesetzt. Wir haben uns in unserer Implementierung für einen lüfterlosen, kompakten Industrie-PC für Hutschienenmontage mit 10 Jahren Lieferfähigkeit und erweitertem Temperaturbereich entschie-

[12] Diese Darstellung des Cloud Technologie-Stack ist weit verbreitet, jedoch ohne klare Quelle. Beispiele finden sich z. B. über die folgende Suche in Google: https://www.google.com/search?q=cloud%20stack&tbm=isch&hl=en-US&tbs=rimg:CSeEJJz4jP7XYaYD1TKpmg8C. Zum Verständnis empfehlenswert ist die Anwendung auf das Gericht „Pizza“: https://www.linkedin.com/pulse/20140730172610-9679881-pizza-as-a-service/.

den. Er besitzt in einer von mehreren verfügbaren Varianten mehrfache Netzwerkschnittstellen und einen COM-Port zur Verbindung mit älteren Maschinen.

Als eine der Hardware-Komponenten ist in unserer Architektur unter anderem eine SSD-Festplatte enthalten, auf der ein dediziertes Secure Element angebracht und manipulationsresistent versiegelt ist. Dieses embedded Secure Element (eSE) übernimmt kryptografische Sicherheitsfunktionen wie die Verschlüsselung der Festplatte mit sicherer Ablage des zugehörigen Schlüsselmaterials.

Die durch uns entwickelte **Software-Plattform** als nächsthöhere Ebene der Architektur verknüpft Hardware und Anwendungen über Firmware und Treiber, ein Betriebssystem (OS) sowie eine eigene Anwendungsumgebung. An diesen Stellen findet ebenso wie an anderen Stellen im Gesamtkonzept eine für Industriezwecke passende Härtung der einzelnen Bestandteile statt. So ist beispielsweise ausschließlich die Installation von signierter und verschlüsselter Firmware möglich (Secure Update), und der Boot-Prozess ist dediziert abgesichert (Secure Boot)[13]. Das OS basiert auf einer weit verbreiteten Linux-Distribution. Da diese bereits für industrielle Zwecke eingesetzt wird, kann sowohl auf die umfangreichen Vorarbeiten in diesem Bereich, als auch auf die laufende Aktualisierung hinsichtlich Funktionalität und Sicherheit zurückgegriffen werden. Die Anwendungsumgebung (basierend auf einer OCI[14]-kompatiblen Container-Laufzeitumgebung) ist ebenfalls gehärtet und über passende Applikationsschnittstellen (APIs) an die weiteren Funktionen der Plattform angebunden.

Der Vorteil dieses Ansatzes ist, dass bereits viele durch Endkunden und Softwareanbieter entwickelte Anwendungen mit der Plattform verwendbar sind und über die Schnittstellen Funktionalität komfortabel nutzbar wird, die so nicht aufwendig eigens entwickelt werden muss. Neben geringeren Aufwänden für die Softwareentwicklung steigt so auch die Sicherheit. Dies zeigt sich insbesondere bei der Nutzung kryptografischer Funktionen. Werden diese durch nicht-spezialisierte Entwickler selbst implementiert, führt dies häufig zu Sicherheitsproblemen.[15] Im Falle der beschriebenen Plattform muss nun z. B. die Generierung eines Schlüssels nicht mehr selbst implementiert oder mittels potenziell unsicherer Bibliotheken eingebunden werden, sondern kann über eine „Generiere Schlüssel“ Funktion in der Plattform effizient und sicher genutzt werden. Die dahinter liegende Software wird durch uns als Plattformanbieter nach den Grundsätzen Security-by-Design und Security-by-Default implementiert und gepflegt.

Aus unseren Erfahrungen im Edge- und Industrial IoT-Markt hat sich gezeigt, dass die Betrachtung der **Anwendungsebene** (vergleichbar der Softwareebene im Cloud-Stack) im Edge-Stack in Middleware, Daten und Apps aufzuteilen ist. Middleware besteht aus spezialisierten Anwendungen, die zwischen der Anwendungsumgebung und den Apps arbeiten.[16] Sie stellen spezielle Funktionalität oder Dienste bereit, die entweder Apps

---

[13] Siehe dazu auch die entsprechenden empfohlenen Best Practice „Mitigations“ im Mitre Att&ck Framework (https://attack.mitre.org/mitigations/M1046/)

[14] Siehe Open Container Initiative (https://opencontainers.org/)

[15] https://www.schneier.com/essays/archives/1998/01/security_pitfalls_in.html

[16] Abweichend zum Cloud-Stack ist die Anwendungsumgebung Teil der Software-Plattform und die Middleware Teil der Anwendungsebene.

nicht selbst implementieren müssen oder die sich mehrere Anwendungen teilen. Middleware übernimmt beispielsweise die Übersetzung von höchst unterschiedlichen Industrieprotokollen auf einen einheitlichen Standard. Aufgrund ihrer Komplexität kann Middleware nicht immer durch den Anbieter der Edge-Plattform selbst angeboten werden. Es handelt sich dann um ein eigenes Produkt, welches in die Edge-Plattform integriert wird („Plattform-auf-Plattform“) und optional lizenzierbar ist. Manche Middleware ist jedoch integraler Bestandteil der Architektur, wie etwa jene für die Anbindung der Edge-Gateways an bestimmte Public Cloud Plattform-Dienste.

Die notwendigen Daten werden meist durch den Endanwender bzw. Integrator eines Anwendungsfalls eingebracht. Teilweise entstehen relevante Daten aber auch aus der Architektur selbst, wie z. B. über das integrierte Sicherheitsmonitoring. Die durch Endanwender zur Umsetzung ihrer Digitalisierungsprojekte nutzbare Funktionalität wird wiederum durch die hier Apps genannten Anwendungen flexibel abgedeckt. Ein Teil der möglichen Apps ist bereits gemeinsam mit der Plattform verfügbar. Dabei handelt es sich insbesondere um Basis-Funktionalität wie z. B. eine VPN-Client-App und eine App zur Umwandlung von Standard-IT-Protokollen in sicherere Varianten.[17] Aber auch eine App zur attributbasierten Authentifizierung und Autorisierung gehört zu den zukunftsträchtigen Anwendungen. Apps und Middleware können hochkomplex sein und tiefes Branchenwissen erfordern und umsetzen. Daher unterstützt die Plattform ein lebhaftes Ökosystem von Partner-Apps und -Middleware für viele **Anwendungsfälle**. Siehe dazu in Abbildung 5 die teilweise Abdeckung durch unsere Umsetzung oder durch separate und komplementäre Teile (schraffiert dargestellt).

Bezüglich Plattform, Middleware und Anwendungen befindet sich die Architektur in einem konstanten Spannungsverhältnis. Funktionalität kann Teil der Plattform sein, über Middleware als Plattformen-auf-Plattform oder über Apps bereitgestellt werden. Neben der vor allem durch die Wirtschaftlichkeit getriebenen Herausforderung, dies auszutarieren, ergibt sich eine große Chance. Denn für die Bereitstellung und Pflege innovativer Sicherheits- und Connectivity-Funktionalitäten existiert durch den Stack-Ansatz eine breite Forschungs- und Innovationsbasis. Sind die einzelnen Komponenten austauschbar oder interoperabel ausgelegt, kann im Einzelfall die am besten geeignete Kombination der jeweiligen Alternativen gewählt werden.

Abbildung 5 stellt dar, wie die beschriebene Architektur durch uns praktisch umgesetzt wurde. Software- und Hardware-Plattform sowie der überwiegende Teil der Infrastruktur liegen in der funktionalen und Sicherheits-Verantwortung des Anbieters. Die Anwendungen (Apps, Daten und Middleware) sind offen für Erweiterungen durch Drittanbieter, Forschungseinrichtungen oder die Anwender selbst (schraffierte Bereiche).

[17] Z. B. mit einer Umwandlung des unsicheren und durch Angreifer bereits seit einiger Zeit aktiv ausgenutzten SMBv1 in eine nicht-angreifbare, verschlüsselte Verbindung nach SMBv3.

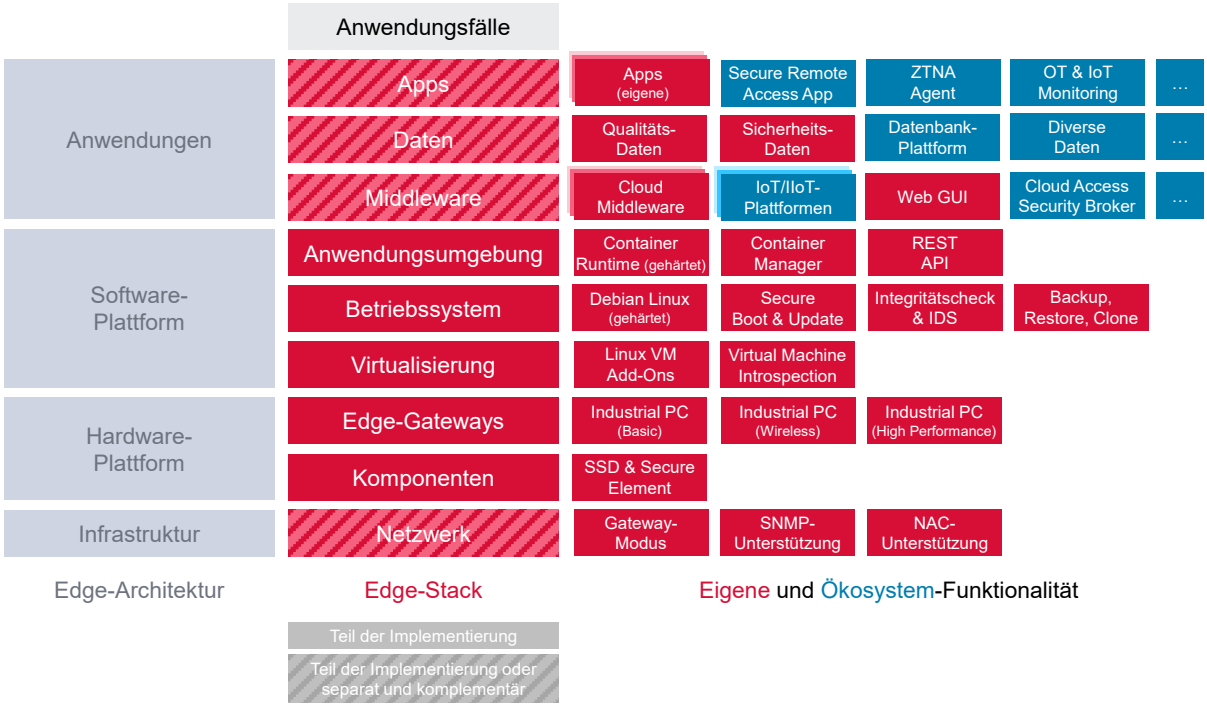


Abbildung 5: Umsetzung der multifunktionalen und sicheren Edge-Architektur

7. Erfahrungen und Lerneffekte aus dem praktischen Einsatz

Die vorgestellte Edge-Architektur hat bereits die Konzeptphase verlassen und ist bei ausgewählten Anwendern aus Industrie und Kritischen Infrastrukturen im praktischen Einsatz. In diesen Proofs-of-Concept (PoC) und Produktiveinsätzen wurden wichtige Erfahrungen gemacht und haben sich konkrete Lerneffekte für die Weiterentwicklung der Architektur ergeben.

7.1 Projekt „Netzwerkmikrosegmentierung“

In einem Projekt im Bereich der Automatisierungstechnik mussten durch ein Unternehmen laufend neue bzw. angepasste Netzwerksegmente und insbesondere deren Firewall-Systeme konfiguriert werden. In der Vergangenheit war dies mit einem Aufwand von mehreren Wochen verbunden, da die Konfiguration jeweils individuell und manuell durch Mitarbeiter über Zugriffe in das jeweilige Netzsegment erfolgte. Ziel des Proof-of-Concept mit der hier vorgestellten Edge-Architektur war es, dies zu automatisieren und die Aufwände signifikant zu reduzieren. Dazu wurden Edge Gateways in den Segmenten platziert und mit einer zentralen Instanz über ein abgesichertes Overlay-Netzwerk verbunden. Die zentral verwaltete Konfiguration konnte anschließend effizient in die jeweiligen Bereiche übertragen werden.

Im Zuge des Projekts zeigte sich, dass bei einem Netzwerk mit bereits vorhandener Infrastruktur („Brownfield“) ein vergleichsweise hoher Aufwand für die Bestimmung der passenden Firewall-Regeln notwendig war. Aus dieser Erfahrung wird die Edge-Architektur nun um eine Firewall-Anlernfähigkeit ergänzt. Bei dieser wird ein Edge-Gateway bzw. eine Edge-Computing-Appliance bereits einige Zeit vor dem Stichtag zur Anpas-

sung der Firewall am Übergang zum entsprechenden Netzsegment platziert. Durch die Integration der plattformeigenen Firewall-Funktionalität mit einer lernenden Anwendung kann nun das übliche Verhalten des Netzwerks erlernt und in Firewall-Regelwerke kodiert werden. Diese Regelwerke können anschließend (direkt oder nach einer manuellen Sichtung und ggf. Anpassung) in der Firewall des Gateways aktiviert oder auch in eine separate, dedizierte Firewall übertragen werden.

7.2 Projekt „Industrie 4.0 Dataspace“

In einem anderen Projekt sollte eine große Anzahl und Vielfalt an Maschinen zum zentralen Datenaustausch im Rahmen eines Industrie 4.0-Projekts angebunden werden. Aufgrund der großen Heterogenität der vorhandenen Industrieprotokolle wurde dieses Projekt gemeinsam mit einem Hersteller einer Middleware-Lösung für Maschinenkonnektivität durchgeführt. Da neben der Heterogenität auch eine große Anzahl geografisch verteilter Systeme anzubinden war, wurde zudem eine Public Cloud Komponente implementiert.

Über das eigentliche Projektziel hinaus konnte über die multifunktionale Architektur als zusätzlicher Lerneffekt auch die Sicherheit der Maschinen gesteigert werden. Dies war möglich, da die Gateways die zuvor kaum geschützten Maschinen nun sicher über eine Firewall anbinden sowie sicheren Hersteller-Fernzugriff ermöglichen. Die Edge-Gateways wurden zudem mit einer Netzwerkverkehrs-Überwachung auf Angriffe und Anomalien ausgestattet (siehe auch den folgenden Abschnitt).

7.3 Projekt „Sicherheitsüberwachung“

In einem dritten Projekt war die Sicherheitsfunktionalität das primäre Projektziel. Zuvor bestand bei dem Unternehmen aus dem Bereich der Kritischen Infrastrukturen die Herausforderung, dass entfernte und oft nur mit niedriger Bandbreite angebundene Netzsegmente nicht oder nur begrenzt überwacht werden konnten. Die Weiterleitung aller sicherheitsrelevanten Daten aus diesen Netzsegmenten war z. B. wegen einer Anbindung über das Mobilfunknetz nicht praktikabel. Als Lösung wurden daher Gateways nach der vorgestellten Architektur als verteilte Sensoren in den lokalen Edge-Netzwerken platziert. Die Erfassung und Analyse sicherheitsrelevanter Daten konnte so möglichst nahe am Ort des Entstehens erfolgen. Die aggregierten und deutlich reduzierten Daten konnten anschließend auch über schmalbandige Verbindungen sicher übertragen und zentral für ein übergreifendes Netzwerk-Monitoring eingesetzt werden.

Im Zuge des Projekts stellte sich früh heraus, dass neben einer Überwachung auf Schwachstellen, mögliche Angriffe und Anomalien insbesondere auch eine Überwachung auf die Verletzung von vorgegebenen Zuständen im Netzwerk (sog. „Policies“) sinnvoll wäre. Diese Funktionalität wurde entsprechend auf Anwendungsebene ergänzt. Netzverantwortliche können nun erwartetes oder zugelassenes Verhalten als Regeln in einer Whitelist eintragen. Jede Abweichung von diesen Regeln wird anschließend erkannt und gemeldet. Die Abweichungen können mittlerweile auch direkt an die Firewall-Konfiguration übertragen und durch diese blockiert werden.

Aus diesen beispielhaften Projekterfahrungen sowie aus Trends im industriellen Umfeld haben sich zudem mehrere grundsätzliche Lerneffekte und Anstöße zur Weiterentwicklung der vorgestellten Architektur ergeben.

7.4 Lerneffekt „Managementvarianten“

Bedingt durch die Vielfalt der KRITIS- und Industrie-Use Cases, stellte sich bald heraus, dass eine einzige Management-Lösung für die eingesetzten Edge-Gateways den vielfältigen Praxisanforderungen der (insbesondere mittelständischen) Betriebe nicht gerecht werden würde. So entstanden insgesamt vier Varianten zentralen Managements, jeweils optimiert für ihre Anwendungsfälle und Sicherheitslevel. Die erste Variante ist die individuelle, manuelle Administration einzelner Appliances für Pilotinstallationen. Die zweite Möglichkeit ist ein dediziertes Management-System für geschlossene, kritische Anwendungsfälle mit geringer Geräteanzahl, welches als Anwendung selbst auf einer Management-Edge-Appliance betrieben werden kann. Die dritte Option ist in der Form einer Gateway-Administration über spezialisierte IoT-Plattformen insbesondere relevant für Organisationen, die bereits eine IoT-Plattform im Einsatz haben und viele Gateways administrieren müssen. Die vierte Variante ist die Administration über Public Cloud-Dienste führender Hyperscaler-Plattformen die auch in sehr hohe Appliance-Stückzahlen skalierbar und besonders flexibel erreichbar ist.

7.5 Lerneffekt „Iterative Implementierung“

Eine weitere Erkenntnis aus den Praxiserfahrungen betrifft die jeweils unterschiedlichen Reifegrade von KRITIS- und Industrieunternehmen. Anders als in Digitalisierungsprojekten, die auf der grünen Wiese geplant werden, ist bei laufendem Betrieb kein Sprung von 0 (kaum Vernetzung, keine Security) auf 1 (vollständige Digitalisierung und Zero-Trust-Netzwerk-Architekturen) möglich. Anwender wünschen ein schrittweises Vorgehen. So soll etwa im ersten Schritt oft „nur“ die Konnektivität in Verbindung mit statischer Netzwerk-Segmentierung umgesetzt werden. In weiteren Schritten kommen nach und nach applikationsspezifische Geräteverbindungen über VPN-Tunnel hinzu, wiederum gefolgt von höherwertigen Sicherheits-Funktionalitäten. Die beschriebene Architektur erwies sich für dieses schrittweise Vorgehen durch die Möglichkeit des Ausrollens neuer App-Container in die bereits installierten Edge-Security-Gateways als sehr gut geeignet.

7.6 Lerneffekt „IT-Sicherheitsgesetz“

Parallel zur Weiterentwicklung der Edge-Architektur waren auch die Gesetzgebung und der Regulator aktiv. Auch wenn es bis zum finalen Stand mit Kabinettsbeschluss am 16. Dezember 2020 diverse Änderungen gab, zeichneten sich einige zentrale Punkte des IT-Sicherheitsgesetzes 2.0 schon vorher ab. Darunter fallen insbesondere der Einsatz von vertrauenswürdigen und nach Stand der Technik entwickelten Komponenten in den Kritischen Infrastrukturen, die Berücksichtigung der Lieferkette bei der Herstellung dieser Komponenten und der Betrieb von Systemen zur Angriffserkennung bei regulierten Unternehmen.

Viele dieser bald gesetzlichen Anforderungen wurden (wie auch bereits die Aspekte des IT-Sicherheitsgesetzes 1.0) bereits seit Beginn von der vorgestellten Edge-Architektur

berücksichtigt. So werden beispielsweise die Industrie-PCs aus dem Technologie-Stack mit einem deutschen Hersteller und Marktführer für vertrauenswürdige industrielle IT entwickelt und von diesem in Deutschland gefertigt. Einzelne, besonders kritische Komponenten wie die CryptoCore-SSD und das darauf verbaute embedded Secure Element werden im eigenen Konzern entwickelt und sind bereits über ihren Einsatz im Finanzsektor bewährt und zertifiziert.[18] Bezüglich der Systeme zur Angriffserkennung kam uns zugute, dass das Netzwerksicherheitsmonitoring-Produkt aus eigenem Haus in der Kombination von Edge-Architektur und Monitoring nun auch als ein verteiltes System eingesetzt wird, welches die Edge-Gateways als Sensoren und CPS-nahe Firewalls nutzen kann.

8. Fazit und Ausblick

Die digitale Transformation steht trotz großer medialer Präsenz an sehr vielen Punkten erst am Anfang ihrer Möglichkeiten. So ist es aber auch noch nicht zu spät, bei der Einführung oder der Nachrüstung (Retrofit) auf Multifunktionalität und Sicherheit zu achten.

Die hier vorgestellte Architektur stellt eine Möglichkeit zur Umsetzung der Vision von innovativer Multifunktionalität und Cybersicherheit dar. Neben dem Themenbereich Edge Computing sind die Autoren dieses Beitrags auch im Thema Cloud Computing aktiv. Dort gibt es mit der europäischen Initiative Gaia-X bereits ein Leuchtturmprojekt für eine souveräne und nachhaltige Technologieentwicklung in Europa. Viele der dort behandelten Aspekte sind auch für das Technologiefeld Edge relevant. So wäre es zu begrüßen, wenn die Prinzipien von Gaia-X auch zu Maximen des Edge Computing würden. Konkret sind dies[19]:

1. Europäischer Datenschutz
2. Offenheit und Transparenz
3. Authentizität und Vertrauen
4. Souveränität und Selbstbestimmtheit
5. Freier Marktzugang und europäische Wertschöpfung
6. Modularität und Interoperabilität
7. Nutzerfreundlichkeit

In der Entwicklung der präsentierten Edge-Architektur wurden diese Punkte berücksichtigt. So ist etwa durch die Architektur der Anwendungsplattform hohes Interoperabilitätspotenzial mit anderen Infrastrukturen und Architekturen gegeben. Die Autoren blicken daher in freudiger Erwartung in die Zukunft der gemeinsamen und sicheren digitalen Transformation in Industrie und Gesellschaft und bieten die vorgestellte Architektur als ihren Beitrag für einen wesentlichen Teilbereich dieses Prozesses an.

---

[18] Je nach Anwendungsbereich nach FIPS 140-2 L3 oder BSI CC L3 EAL5

[19] https://www.bmwi.de/Redaktion/DE/Publikationen/Digitale-Welt/das-projekt-gaia-x-executive-summary.pdf?blob=publicationFile&v=16

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

IT-Sicherheitsupdates: Pflichten für Hersteller und Verkäufer

Dr. Dennis-Kenji Kipker[1]

Kurzfassung:
Mit dem steten technischen Wandel stellt sich Software nicht selten erst im Nachhinein als unsicher dar. Auch nach der ursprünglichen Vertragserfüllung sind Anbieter von digitalen Produkten deshalb gehalten, die IT-Sicherheit durch regelmäßige Software-Aktualisierungen sicherzustellen. Eine derartige Rechtspflicht kann sich zum einen aus vertraglichen Nebenpflichten, zum anderen aus dem Deliktsrecht ergeben, und wird überdies zukünftig über das europäische Recht Eingang in den nationalen Verbraucherschutz finden.

Stichworte: Aktualisierungen, Digitale Produkte, Haftung, Patches, Softwareaktualisierungen, Updates, Updatepflicht, Verantwortlichkeit, Vertrag

IT-Sicherheit ist eine Daueraufgabe, denn nahezu tagtäglich offenbaren sich neue Risiken für Softwareprodukte. Dies erschwert es Herstellern und Verkäufern von IT-Produkten, die technische Ausgangslage angemessen zu überblicken und unter juristischen Gesichtspunkten zu ermitteln, in welchen Fällen, in welchem Umfang und für wie lange sie ihren Kunden IT-sicherheitsbezogene Softwareupdates zur Verfügung stellen müssen.

Der Beitrag beleuchtet die gegenwärtige Situation im Vertrags- und Deliktsrecht für die nicht immer eindeutig geregelten Updatepflichten für die IT-Sicherheit – insbesondere auch jenseits individuell vereinbarter Garantieverträge. Einbezogen werden dabei außerdem die aktuellen Entwicklungen im EU-Verbraucherschutzrecht aus der Herstellerperspektive. Ausgangspunkt der juristischen Betrachtung ist die Situation, dass ein digitales Produkt bei seinem Rollout den seinerzeit gängigen technischen Anforderungen entsprach, die sich im Nachhinein aufgrund weiterer Entwicklungen jedoch nicht mehr als ausreichend erweisen.

1. Rechtsquellen

Bisher wurde im deutschen Recht auf die ausdrückliche Regulierung einer regelmäßigen Aufrechterhaltung der IT-Sicherheit von digitalen Produkten weitestgehend verzichtet. Zwar lassen sich mögliche Pflichten aus einer Auslegung bestehender vertrags- und deliktsrechtlicher Vorschriften unter Zuhilfenahme der Rechtsprechung herleiten, eine solche Vorgehensweise ist jedoch alles andere als transparent und rechtssicher. Abhilfe könnten hier Regulierungsansätze aus dem EU-Recht schaffen, die zurzeit zwar primär nur für das B2C-Segment gelten und deshalb lediglich mittelbar über Regressansprüche auch Pflichten im B2B-Bereich entfalten. Nicht ausgeschlossen ist aber, dass eine derartige Entwicklung rechtspolitisch schon jetzt die Weichen auch für eine künftige er-

[1] Institut für Informations-, Gesundheits- und Medizinrecht (IGMR), Universität Bremen. Der Verfasser dankt Michael Walkusz von der Universität Bremen für seine Unterstützung bei der Erstellung dieses Beitrags.

weiterte Verantwortlichkeit von Herstellern und Verkäufern außerhalb von Verbraucherverträgen stellt. In diesem Zusammenhang zu nennen sind die EU-Richtlinie über bestimmte vertragsrechtliche Aspekte der Bereitstellung digitaler Inhalte und digitaler Dienstleistungen (DID-RL)[2], die zurzeit mit einem aktuellen Referentenentwurf (DID-RefE)[3] in das BGB umgesetzt wird, und die EU-Richtlinie über bestimmte vertragsrechtliche Aspekte des Warenkaufs (WK-RL)[4], für die ebenfalls ein im Dezember 2020 veröffentlichter Referentenentwurf vorliegt (WK-RefE)[5]. Die neuen EU-Richtlinien und die nationalen Umsetzungsgesetze sehen im Hinblick auf die IT-Sicherheit verschiedene Anforderungen und Aktualisierungspflichten im Verbraucherschutzrecht vor.

2. Maßgaben aus dem Vertragsrecht

2.1. Verantwortlichkeiten für Softwareherstellung und Vertrieb

Die Herstellung von Software und entsprechende Vertriebswege sind mittlerweile keine „Einbahnstraße“ mehr, verstanden als einseitige Leistungsbeziehung zwischen Hersteller, ggf. Zwischenvertrieb und Endnutzer. Im Gegenteil, entsprechende Herstellungs- und Vertriebswege stellen sich infolge der heutzutage in einer Vielzahl beteiligten Akteure als ein hochkomplexes Netzwerk dar. So können beispielsweise die Grundkomponenten und Bibliotheken einer fertigen Software aus Open Source-Quellen stammen, bevor sie von einem Zulieferer an den entsprechenden Hersteller gelangen, der dann unter Umständen nach individuellen Kundenwünschen weitere Anpassungen am Produkt vornimmt. Weitere technische und damit ebenso rechtliche Schwierigkeiten können sich daraus ergeben, dass eine Software nicht als selbstständiges Produkt ausgeliefert wird, sondern vielmehr im Sinne eines Embedded Systems untrennbar mit einer Hardwarekomponente verbunden ist, um ihre Funktionsfähigkeit zu entfalten.

2.2. Updatepflichten und Vertragstypologisierung

Die Art und der Umfang von möglichen Updatepflichten zur IT-Sicherheit sind von der jeweils zwischen den Parteien gewählten Vertragsform abhängig. Allen Regelungen gemeinsam ist aber, dass das digitale Produkt frei von Mängeln bereitzustellen ist.[6] Das gilt unabhängig davon, ob es sich um eine selbstständige Software oder um ein Embedded System handelt. Sieht der Vertrag vor, dass das digitale Produkt zu einem punktuellen Zeitpunkt bereitzustellen ist, so kommen dem Grunde nach kauf- und werkvertragsrechtliche Regelungen in Betracht, um die Mangelfreiheit zu bewerten. Bei laufenden und regelmäßig wiederkehrenden Leistungen dürfte ein Dauerschuldverhältnis vorliegen, für das der Rückgriff auf das Miet- und Dienstvertragsrecht geboten ist. Soweit

---

[2] EU-RL 2019/770.

[3] Referentenentwurf eines Gesetzes zur Umsetzung der Richtlinie über bestimmte vertragsrechtliche Aspekte der Bereitstellung digitaler Inhalte und digitaler Dienstleistungen, Stand 05.10.2020, abrufbar unter: https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RefE_BereitstellungdigitalerInhalte.pdf (01.01.2021).

[4] EU-RL 2019/771.

[5] Referentenentwurf eines Gesetzes zur Regelung des Verkaufs von Sachen mit digitalen Elementen und anderer Aspekte des Kaufvertrags, abrufbar unter: https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RefE_Warenkaufrichtlinie.pdf (01.01.2021).

[6] Vgl. §§ 433 Abs. 1 S. 2, 535 Abs. 1 S. 2, 633 Abs. 1 BGB.

es Verbraucherverträge über digitale Produkte und digitale Dienste anbelangt, setzt der DID-RefE neue Maßstäbe, denn er schlägt die Schaffung eines eigenständigen Gewährleistungsrechts im Allgemeinen Teil des Schuldrechts vor. Ein so verstandener „digitaler Verbraucherschutz“ durchzieht mithin die unterschiedlichen Vertragstypen. Innerhalb dieses Regelungskomplexes differenziert der Entwurf innerhalb einzelner Vorschriften zwischen punktuellen und dauerhaften Leistungsbeziehungen und schließt in seinem Anwendungsbereich den Rückgriff auf sonstige vertragsrechtliche Regelungen des BGB aus. Der WK-RefE sieht vor, dass die vorgesehenen Neuerungen in das bestehende Kaufrecht übernommen werden.

Soweit es die Vertragstypologisierung zur Bestimmung von Updatepflichten für sichere Software anbelangt, kann der juristische Streit, ob es sich bei Software um eine Sache im Rechtssinne handelt, mangels nennenswerter rechtlicher Auswirkungen dahinstehen: Im Kaufrecht erklärt § 453 BGB für den Rechtskauf die Vorschriften über den Sachkauf für entsprechend anwendbar,[7] im Werkvertragsrecht wird schon keine Sache, sondern ein Erfolg geschuldet,[8] und bei der Überlassung auf Zeit ist entweder das Mietrecht bei bejahter Sacheigenschaft[9] oder ansonsten das Pachtrecht anwendbar[10], wobei § 582 Abs. 2 BGB rechtlich wiederum Rückgriff auf das Mietrecht nimmt.

2.3. Pflichten aus Kauf- und Werkverträgen

Aufgrund oft nicht gesondert getroffener Beschaffenheitsvereinbarungen zur IT-Sicherheit richtet sich die Mangelfreiheit für Softwareprodukte nach § 434 Abs. 1 S. 2 Nr. 2 BGB für das Kaufrecht bzw. nach § 633 Abs. 2 S. 2 Nr. 2 BGB für das Werkvertragsrecht. Hiernach ist das Produkt in einer Weise zu leisten, die bei Sachen gleicher Art üblich ist und die der Kunde nach der Art der Sache erwarten kann. Diese Eigenschaft bestimmt sich nach dem Erwartungshorizont eines Durchschnittskäufers. Der Durchschnittshorizont ist im Wege der Auslegung festzustellen, wobei vor allem die Verkehrsauffassung bei einem Vergleich mit Produkten der gleichen Gattung[11], der Zweck des Produkts, aber auch wesentliche Sicherheitsvorschriften, sowie technische Normen und Standards von Bedeutung sein können.[12] Ein genereller Rückgriff auf den vielzitierten „Stand der Technik“ ist jedoch nicht statthaft, soweit keine gesonderte Vereinbarung getroffen wurde. Insbesondere kann dieser unbestimmte Rechtsbegriff nicht ohne Weiteres auf verbraucherschutzrechtliche Sachverhalte bezogen werden, da er sich an der „Front des technischen Fortschritts“[13] bemisst, der andere Beurteilungsmaßstäbe zugrunde liegen dürften. Für die Auslegung und Bestimmung der Mangelfreiheit können aber die Definitionen des BSIG einzubeziehen sein. So findet sich eine Bestimmung zur „Sicherheit in der Informationstechnik“ in § 2 Abs. 2 BSIG. § 2 Abs. 6 BSIG enthält

7 Stresemann in: MüKoBGB, § 90, Rn. 25.
8 Busche in: MüKoBGB, § 631, Rn. 1.
9 BGH NJW 2007, 2394; von dem Bussche/Schelinski, Münchner Anwaltshandbuch zum IT-Recht, Rn. 340.
10 Roth-Neuschild in: Auer-Reinsdorff/Conrad, Handbuch IT- und Datenschutzrecht, § 13, Rn. 23; zur Megede, NJW 1989, 2580, 2583.
11 Faust in: BeckOKBGB, § 434, Rn. 66.
12 BGH NJW 1985, 1769, 1770; Berger in: Jauernig, BGB, § 434, Rn. 30.
13 BVerfG NJW 1979, 359, 362.

eine Legaldefinition der IT-Sicherheitslücke. Aus den beiden vorgenannten Vorschriften geht hervor, dass eine Software jedenfalls hinreichend vor äußeren Einflüssen geschützt sein muss, um als sicher zu gelten.[14] Eine Schwachstelle ist somit bereits dann gegeben, wenn das bloße Risiko eines an sich vermeidbaren Fremdeinflusses besteht.[15] War die geleistete Software somit zum Zeitpunkt des Gefahrübergangs, also bei Übergabe bzw. Abnahme oder Bereitstellung nicht hinreichend vor unzulässigen Fremdeinflüssen geschützt, so kann von einem rechtlich relevanten Mangel ausgegangen werden, der den Anbieter zu einer Nacherfüllung in Form der Bereitstellung einer sicheren Version des Programms verpflichtet.

Juristisch schwieriger zu beurteilen ist der Fall, in dem die Software bei Gefahrübergang im oben beschriebenen Sinne zwar mangelfrei war, sich allerdings durch den vom Anbieter zwar unverschuldeten, aber allgemein vorhersehbaren technischen Wandel im weiteren zeitlichen Nutzungsverlauf als unsicher herausstellt. Um für diesen Fall für den Käufer bzw. Verwender unbillige Härten zu vermeiden, wäre es daher denkbar, unter dem Gesichtspunkt der IT-Sicherheit rechtliche Ausnahmen von den allgemeinen Gefahrtragungsregeln zu machen. In dem Zusammenhang werden verschiedene Theorien diskutiert, die u.a. von einer Verschiebung des Gefahrübergangs oder einem Mangelverdacht ausgehen,[16] im Ergebnis aber juristisch nicht überzeugend sind. Denn die bloße, wenn auch nahezu sichere Erwartbarkeit einer IT-Sicherheitslücke durch Zeitablauf kann keinesfalls konkret genug sein, um Handlungspflichten für Hersteller und Verkäufer auszulösen. Anderenfalls bestünde eine nicht mehr vertretbare Verantwortlichkeit für IT-Produkte, die einer theoretisch unbegrenzten Erwartung an die Nutzungsdauer gleichkäme. Ist der Mangel nicht bereits bei Gefahrübergang angelegt, so scheint es unbillig, das allgemeine Betriebsrisiko, das auch die IT-Sicherheit umfasst, auf Hersteller und Verkäufer abzuwälzen.

Allerdings kann sich eine Verpflichtung, durch nachträgliche und regelmäßige Updates für eine IT-sichere Software zu sorgen, als Bestandteil einer vertraglichen Nebenpflicht ergeben. So hat die Rechtsprechung schon in den 1990er-Jahren für Verwaltungsprogramme „Wartungspflichten“ zur Sicherstellung ihrer reibungslosen Funktionsfähigkeit auferlegt. Dies wurde mit dem berechtigten Interesse des Nutzers begründet, das Produkt für einen gewöhnlichen und seiner Preisklasse angemessenen Lebenszyklus verwenden zu können.[17] Freilich ist damit noch keine Aussage darüber getroffen, was unter einem „angemessenen Lebenszyklus“ zu verstehen ist. Die Rechtsprechung zieht hier Vergleiche zur Bereitstellung von physischen Ersatzteilen.[18] So soll mindestens für den Zeitraum, innerhalb dessen die Software noch auf dem Markt angeboten wird, und für weitere fünf Jahre, nachdem das Produkt vom Markt genommen wurde, eine solche

14 *Rockstroh/Kunkel*, MMR 2017, 77, 78.

15 *Redeker*, IT-Recht, Rn. 342.

16 Vgl. *Schimmer*, DuD 2006, 616, 617.

17 OLG Koblenz NJW 1993, 3144, 3145; LG Köln NJW-RR 1999, 1285, 1286.

18 LG Köln NJW-RR 1999, 1285, 1286; AG Rüsselsheim DAR 2004, 280.

Pflicht zur Bereitstellung von Updates bestehen.[19] In Anbetracht des Alters dieser Entscheidung scheint heutzutage eine andere Wertung angemessen – nicht nur, weil Software einen deutlich größeren Stellenwert als vor 20 Jahren hat, sondern auch, weil die physische Lebensdauer hochpreisiger IT-Produkte die der Software (ohne Updates) weit überschreitet.

2.4. Pflichten aus Dauerschuldverhältnissen

Mittlerweile wird Software bevorzugt nicht mehr verkauft und dem Kunden sodann unverändert zum Gebrauch überlassen, sondern wird ihm mehr und mehr in der Form von Serviceverträgen (bspw. SaaS als eine Form des Cloud Computings) angeboten. Das gilt vor allem für komplexe Embedded Systems aus dem Bereich Industrie 4.0. Für diese Fälle von Dauerschuldverhältnissen beurteilt sich die Vertragsmäßigkeit der angebotenen Leistung am gesamten Zeitraum der vertraglichen Beziehung. Beispielsweise vom Mietrecht ausgehend (es wäre auch an Pachtverträge oder gemischte Verträge *sui generis* zu denken) ist die Mangelhaftigkeit eines IT-Produkts nach § 536 Abs. 1 BGB zu bemessen. Ein Mangel ist jede nachteilige Abweichung des tatsächlichen Zustands von dem vertraglich vereinbarten Zustand.[20] Hierbei lässt sich eine neu auftretende IT-Sicherheitslücke einer gemieteten Software zweifelsfrei unter diesen Mangelbegriff subsumieren. Auch wenn dadurch die Gebrauchstauglichkeit der Mietsache freilich nicht unmittelbar aufgehoben wird, so lässt sich aus § 535 Abs. 1 S. 2 BGB eine Instandhaltungspflicht des Vermieters ableiten, die ihn verpflichtet, Vorsorgemaßnahmen zu treffen, um Eingriffe in die Rechte des Mieters zu vermeiden, die kausal aus der unsicheren Software resultieren könnten.[21] Ähnliche rechtliche Erwägungen müssen für Dienstverträge zur Instandhaltung von Software gelten.

2.5. Verbraucherschutz: Neue Pflichten aus DID-RL und WK-RL

Mit der DID-RL und der WK-RL greift der europäische Gesetzgeber erstmals die Sicherheit als Maßstab für die Vertragsmäßigkeit von digitalen Produkten ausdrücklich auf – hierzu gehören ebenfalls entsprechende Pflichten zur Aktualisierung. Die europäischen Richtlinien sind bis zum 01.07.2021 in das nationale Recht umzusetzen, entsprechende nationale Gesetze sollen zum 01.01.2022 Wirkkraft entfalten. Die Richtlinien gelten in ihrem Anwendungsbereich zwar grundsätzlich nur für das B2C-Segment,[22] eine rechtspolitische Übertragung auch auf B2B scheint aber für die Zukunft nicht ausgeschlossen. Überdies sehen beide Richtlinien in Art. 20 DID-RL, Art. 18 WK-RL Regressansprüche des betroffenen Unternehmers gegen seinen jeweiligen Zulieferer vor. Somit reichen die verbraucherschützenden Bestimmungen auch mittelbar in B2B-Leistungsbeziehungen hinein, wenn vorgesehen ist, das Produkt am Ende der Lieferkette einem Verbraucher anzubieten. Zur Gewährleistung einer umfassenden Harmonisierung belassen die Richtlinien keine weiteren nationalen Umsetzungsspielräume. Art. 4 DID-

---

19 LG Köln NJW-RR 1999, 1285, 1286.

20 Dazu BGH NJW 2000, 1714, 1715; BGH NJW 2005, 2152; BGH NJW 2010, 3152; BGH NJW 2011, 514, 515.

21 BGH NJW 1987, 831, 833; *Hübner/Griesbach/Fuerst* in: Lindner-Figura/Oprée/Stellmann, Geschäftsraummiete, Rn. 93.

22 Art. 3 Abs. 1 DID-RL, Art. 3 Abs. 1 WK-RL.

RL und Art. 4 WK-RL verbieten den Mitgliedstaaten sogar die Einführung eines strengeren Verbraucherschutzniveaus, sodass der deutsche Gesetzgeber gehalten ist, den Regelungsgehalt der Richtlinien identisch in das geltende nationale Recht zu übertragen.

Bei beiden Richtlinien war der EU-Gesetzgeber um größtmögliche Technologieoffenheit bemüht. Daher enthält die DID-RL eine weit gefasste Begriffsdefinition[23] der für ihre Anwendbarkeit notwendigen digitalen Inhalte und digitalen Dienstleistungen, die als „digitale Produkte" zusammengefasst werden. Digitale Inhalte sind gem. Art. 2 Nr. 1 DID-RL solche, die digital erstellt und bereitgestellt werden. Dabei kommt es auf das kumulierte Vorliegen des digitalen Herstellungs- und Bereitstellungsprozesses an. Digital verfasste, aber in Papierform bereitgestellte Bücher oder analog erzeugte, und digital bereitgestellte Bilder fallen nicht in den Anwendungsbereich der Vorschriften. Typische Beispiele für digitale Produkte sind daher Computerproramme, Apps, Videos, Audio, Bilder, digitale Spiele oder elektronische Bücher, gleichgültig, ob diese heruntergeladen oder gestreamt werden. Digitale Dienstleistungen hingegen sind nach Legaldefinition des Art. 2 Nr. 2 DID-RL solche, die dem Verbraucher die Erstellung, Verarbeitung oder Speicherung von Daten in digitaler Form oder den Zugang zu solchen Daten ermöglichen, oder die gemeinsame Nutzung der vom Verbraucher oder von anderen Nutzern der entsprechenden Dienstleistung in digitaler Form hochgeladenen oder erstellten Daten oder sonstige Interaktionen mit diesen Daten ermöglichen. Dies umfasst Software-as-a-Service (SaaS) und andere Formen des Datei-Hostings, die in einer Cloud Computing-Umgebung und in sozialen Medien angeboten werden.[24] Die DID-RL findet auch Anwendung, wenn die digitalen Produkte auf einem physischen Datenträger bereitgestellt werden, der ausschließlich der Bereitstellung dient, Art. 3 Abs. 3 DID-RL. Für „Waren mit digitalen Elementen", also bewegliche körperliche Gegenstände, die in einer Weise digitale Produkte enthalten oder mit ihnen verbunden sind, dass die Waren ihre Funktionen ohne diese digitalen Inhalte oder digitalen Dienstleistungen nicht erfüllen könnten, ist der Anwendungsbereich der WK-Richtlinie eröffnet, vgl. Art. 2 Nr. 5 lit. b WK-RL. Mit dieser Definition werden durch die WK-RL vor allem Produkte des IoT umfasst.

2.5.1. Mangelbegriff nach DID-RL und WK-RL

Beide EU-Richtlinien bestimmen für den digitalen Mangelbegriff neue Anforderungen. Dabei wird in Art. 6 ff. DID-RL und Art. 5 ff. WK-RL zwischen der vertraglich vereinbarten Beschaffenheit (Art. 7 DID-RL und Art. 6 WK-RL – Subjektive Anforderungen) und einer solchen Beschaffenheit unterschieden, die aus objektiver Sicht, namentlich auch am Maßstab der Sicherheit, marktüblich und vernünftigerweise zu erwarten ist (Art. 8 DID-RL und Art. 7 WK-RL – Objektive Anforderungen). Die Richtlinien sehen jedoch keinen Vorrang von vertraglichen Beschaffenheitsvereinbarungen gegenüber objektiven Anforderungen vor. Dies ist aber unschädlich, da in den allermeisten Fällen bei Verbraucherverträgen als tägliches Massengeschäft ohnehin nicht von individuellen Abreden auszugehen sein dürfte. Deshalb dürften im Anwendungsbereich beider EU-

23 *Staudenmayer*, NJW 2019, 2497, 2497 f.

24 Erwägungsgrund 19 DID-RL.

Richtlinien die objektiven Anforderungen an die Vertragsmäßigkeit die weitaus größere Rolle spielen. Zur konkreten Beurteilung eines Mangels kommt es jedoch darauf an, was begrifflich unter der weit gefassten „Sicherheit" zu verstehen ist. Eine Auslegung ergibt in diesem Zusammenhang Folgendes: Einerseits soll der Nutzer eines informationstechnischen Systems bzw. von Maschinen, in denen ein solches System eingebettet ist, vor Schäden an eigenen Rechtsgütern geschützt werden – die sog. „safety". Andererseits ist ein digitales Produkt so herzustellen, dass es vor unbefugten Eingriffen und Manipulationen Dritter geschützt ist – die sog. „security". Im englischen Sprachgebrauch werden in der Regel für die jeweilige Schutzrichtung beide genannten Begriffe verwendet, wohingegen im Deutschen der Unterschied beider Schutzrichtungen nur im Synonym „Sicherheit" nicht eindeutig zum Ausdruck kommt.[25] Die englische Sprachfassung der DID-RL hingegen spricht von „security", was für den Schutz eines digitalen Produktes von außen spricht. Dieses Verständnis einer Aufnahme der IT-Sicherheit wird durch Erwägungsgrund 48 der DID-RL bestätigt, denn hier wird festgeschrieben, dass Zahlungsdaten eines Verbrauchers, der online einen Kauf tätigt, vor Schad- und Spähsoftware zu schützen sind.

Die beiden EU-Richtlinien verwenden im Rahmen der objektiven Anforderungen an die Vertragsmäßigkeit im Hinblick auf die „übliche" und „zu erwartende" Beschaffenheit die gleiche Formulierung wie in Art. 2 Abs. 2 lit. d der EU-Verbrauchsgüterkauf-Richtlinie[26], die bereits in das deutsche Kaufrecht umgesetzt wurde. Deshalb können auch im Bereich der IT-Sicherheit für die Beurteilung der üblichen bzw. zu erwartenden Anforderungen die bereits dargestellten, aktuellen Bestimmungen des Kaufrechts herangezogen werden,[27] sodass im Ergebnis für die Bestimmung der Anforderungen an die IT-Sicherheit keine wesentlichen Änderungen zu erwarten sind.

Entgegen dem in Art. 22 DID-RL und Art. 21 WK-RL vorgesehenen zwingenden Charakter des neuen EU-Verbraucherschutzes sehen Art. 8 Abs. 5 DID-RL und Art. 7 WK-RL vor, den Parteien die Möglichkeit zu eröffnen, von den objektiven Anforderungen der Vertragsmäßigkeit des Produkts durch Vereinbarung abzuweichen, wenn der Verbraucher der Abweichung ausdrücklich und gesondert zugestimmt hat. Obwohl diese Bestimmung dem grundsätzlich und zuvor gesetzlich festgeschriebenen gleichrangigen Charakter von subjektiven und objektiven Anforderungen widerspricht, soll durch die Abweichungsmöglichkeit der Privatautonomie der Parteien Rechnung getragen werden. Die in diesem Zusammenhang geforderten höheren Anforderungen des Verbrauchers an seine Erklärung sollen ihn vor überraschenden Folgen seiner Entscheidung schützen. Zwar reicht somit beispielsweise ein bloßer Verweis auf den Servicevertrag nicht aus, allerdings sind die Anforderungen an die Erklärung – der gesonderten Einwilligung in Datenschutzbestimmungen ähnlich – bereits bei einer besonders gekennzeichneten Opt-in-Lösung erfüllt. Die praktische Wirkung dieser gesetzlichen Abweichungsmöglichkeit ist kritisch zu hinterfragen, vor allem deshalb, weil sie eine Absenkung der Anforderungen an die Vertragsmäßigkeit ermöglicht, und sich der Unternehmer so seiner an sich

25 Zur Differenzierung zwischen „safety" und „security" siehe *Kipker* in: ders., Rechtshandbuch Cybersecurity, 2020, S. 4.
26 EU-RL 1999/44.
27 *Staudenmayer*, ZEuP 2019, 663, 680 f.

geforderten objektiven gesetzlichen IT-Sicherheitsverpflichtung verhältnismäßig leicht entziehen kann. Ebenfalls dürfte schon jetzt davon auszugehen sein, dass von derlei Klauseln in naher Zukunft inflationär Gebrauch gemacht wird. Ob überdies die erhöhten Anforderungen des Verbrauchers an seine Erklärung sinnstiftend sind, sei dahingestellt. Der zumeist rechtsunkundige Verbraucher dürfte sich in vielen Fällen faktisch dazu genötigt fühlen, zum Leistungserhalt derlei Klauseln zuzustimmen,[28] denn keineswegs anders ist die Situation auch bei der Bestätigung Allgemeiner Geschäftsbedingungen oder bei der Abgabe datenschutzrechtlicher Einwilligungserklärungen. Durch diese individuelle Abweichungsmöglichkeit wird es Unternehmen – vor allem solchen, die auf ihrem Gebiet ohnehin eine Monopolstellung besitzen – ermöglicht, die Anforderungen an die Qualität ihres digitalen Produkts faktisch einseitig zu bestimmen.[29] So werden dann auch die gesetzlichen Regelungen zur objektiven Vertragsmäßigkeit ausgehebelt, was nicht im Sinne der allgemeinen Sicherheit digitaler Produkte sein kann.

2.5.2. Aktualisierungspflicht nach DID-RL und WK-RL

Für digitale Produkte und Dienste sowie für Waren mit digitalen Elementen bestimmen Art. 8 Abs. 2 DID-RL und Art. 7 Abs. 3 WK-RL eine ausdrückliche Aktualisierungspflicht, das umfasst ebenso Sicherheitsaktualisierungen. Mit der Aktualisierung soll bei Verträgen mit punktuellen Leistungen die Vertragsmäßigkeit auch nach der erstmaligen Bereitstellung erhalten bleiben. Das soll insbesondere für den Fall gelten, wenn das Produkt unsicher wird, obwohl dessen Funktionsfähigkeit weiterhin gegeben ist.[30] Hieraus folgt mithin die grundsätzliche gesetzliche Pflicht des Unternehmers, das digitale Produkt regelmäßig auf die Vertragsmäßigkeit, insbesondere auf dessen Sicherheit, hin zu prüfen, und im Bedarfsfalle durch entsprechende Aktualisierungen (Updates) nachzubessern. Dabei deutet die passive Formulierung im Gesetz an, dass der Unternehmer diesen Pflichten nicht persönlich nachkommen muss, er kann sich somit eines Erfüllungsgehilfen bedienen. Auch muss der Unternehmer die Aktualisierung nur „bereitstellen“, die Installation obliegt gem. Art. 8 Abs. 3 DID-RL und Art. 7 Abs. 4 WK-RL dem Verbraucher selbst. Dieser ist zur Durchführung der Installation zwar nicht verpflichtet, allerdings kann er seine Gewährleistungsrechte verlieren, wenn er eine rechtzeitige oder instruktionsmäßige Installation unterlässt. Da aber vom Verbraucher als technischem Laien eine selbstständige Einbettung neuer Software in ein digitales Produkt nicht erwartet werden kann, trägt der Unternehmer weiterhin die sinngemäße Verantwortung, den Installationsprozess für den Verbraucher einfach und verständlich zu gestalten, so durch Hinzufügen einer Anleitung[31] oder mit einem Installationsprogramm, das den Verbraucher durch den Prozess begleitet. Die gesetzliche Aktualisierungspflicht umfasst auch, dass der Unternehmer dafür Sorge trägt, dass der Verbraucher auf einem geeigneten Weg über die Bereitstellung der Aktualisierung informiert wird.

Mit dem Bestehen einer Aktualisierungspflicht über den Zeitpunkt der Bereitstellung des digitalen Produkts hinaus stellt sich die Frage ihrer zeitlichen Dauer. Art. 8 Abs. 2

---

28 Vgl. *Kumkar*, ZfPW 2020, 306, 318.

29 A.A. *Ehle/Kreß*, CR 2019, 723 (726); *Kumkar,* ZfPW 2020, 306 (312).

30 Siehe auch Begründung des DID-RefE zu § 327f Abs. 1 BGB-E, 64.

31 Vgl. Art. 8 Abs. 1 lit. c DID-RL.

lit. b DID-RL und Art. 7 Aba. 3 lit. a WK-RL legen bei Verträgen mit punktueller Leistung die Art und den Zweck des digitalen Produkts, sowie die Umstände und die Art des Vertrages als Maßstab fest. Diese offen formulierten Kriterien und unterschiedlichen Anknüpfungspunkte tragen durch ihre Flexibilität zwar dem raschen technologischen Wandel Rechnung, machen es aber im Sinne der Rechtssicherheit umso schwieriger, einen verlässlichen Zeitraum für die Updatepflicht zu bestimmen. Erwägungsgrund 47 DID-RL nennt hier den Zeitraum der Gewährleistung als Mindestdauer, die Aktualisierungspflicht kann sich zeitlich indes auf die zu erwartende Lebensdauer eines Produkts erweitern.

3. Maßgaben aus dem Deliktsrecht

Bei der Betrachtung der Pflichten zu IT-Sicherheitsupdates spielen nicht nur vertragliche, sondern auch deliktische Anforderungen eine Rolle. Soweit eine Software zum Zeitpunkt ihres Inverkehrbringens den zu erwartenden Sicherheitsmaßstäben entspricht, scheidet zwar eine Haftung nach § 1 Abs. 1 ProdHaftG aus, da schon kein Fehler gem. § 3 ProdHaftG vorliegt. Nichtsdestotrotz ergeben sich aus dem allgemeinen Deliktsrecht, namentlich § 823 Abs. 1 BGB, bestimmte Pflichten, die auch die IT-Sicherheit von Produkten umfassen. Insbesondere kann hier aus der deliktsrechtlich anerkannten Produzentenhaftung[32] eine Produktbeobachtungspflicht[33] erwachsen. Generell gilt, dass die Verantwortung des Herstellers für ein Produkt nicht mit dem Rollout endet, sondern aufgrund der fortlaufenden, durch den Betrieb entstehenden Gefährdung und den wirtschaftlichen Vorteilen für den Hersteller fortbesteht. Dementsprechend haftet er für Schäden, die bei IT-Sicherheitslücken kausal dadurch verursacht werden, dass Dritte (gezielt) in informatische Funktionsabläufe des Produkts eingreifen, wenn der Hersteller zumutbare Maßnahmen zum Schutz der Rechte und Rechtsgüter des Betroffenen unterlässt.[34] Aus dieser Verantwortlichkeit lassen sich verschiedene IT-sicherheitsbezogene Pflichten ableiten, die unabhängig von den Festsetzungen aus dem Vertragsrecht gelten. Zuvorderst treffen den Hersteller umfassende Erkundigungspflichten nach möglichen Gefahrenquellen des Produkts. Dabei stehen im Bereich der IT-Sicherheit vielfältige Möglichkeiten zur Verfügung, so bspw. in Form von User Groups, Mailinglisten, Produktfeedback, der Beteiligung an CERTs/PSIRTs, sowie bei Konferenzen und in Fachbeiträgen. Soweit infolge der Informationsbeschaffung Risiken festgestellt werden, ist der Hersteller gehalten, gefahrmindernd einzugreifen.[35] Mangels einschlägiger Rechtsprechung speziell in der IT-Sicherheit sind jedoch nur wenige konkrete rechtliche Vorgaben zur Ausgestaltung der Gefahrbeseitigungspflicht vorhanden. Generell dürfte aber anzunehmen sein, dass professionell eingesetzte und hochwertige Produkte aus dem B2B-Bereich umfassende Produktbeobachtungspflichten des Herstellers auch für die

[32] BGH NJW 1969, 269, 272 ff.
[33] BGH NJW 1981, 1606, 1607.
[34] BGH NJW 1990, 1236, 1237.
[35] RGZ 163, 21, 26.

IT-Sicherheit nach sich ziehen. Dem Hersteller steht in diesem Zusammenhang ein abgestufter Maßnahmenkatalog zur Verfügung, der von einer bloßen Warnpflicht[36] bis hin zur (kostenpflichtigen) Beseitigung der IT-Sicherheitslücke durch entsprechende Updates reicht.

4. Fazit und Ausblick

Die Rechtspflichten zu IT-Sicherheitsupdates sind vielfältig und komplex, und gehen in verschiedenen Fällen nicht deutlich aus den gesetzlichen Vorschriften hervor, sondern sind erst im Wege extensiver Auslegung zu ermitteln. Gleichwohl steht fest, dass Hersteller und Verkäufer digitale Produkte nach dem Rollout für einen angemessenen Zeitraum mit Sicherheitspatches zu versorgen haben. Damit ist jedoch nicht zwangsläufig eine kostenfreie Zurverfügungstellung der IT-Sicherheitsaktualisierungen verbunden, sondern diese können im Sinne der Gefahrtragungsregeln bei einem bei Auslieferung zunächst einwandfreien Produkt auch entgeltpflichtig sein. Verhältnismäßig einfach ist hier die rechtliche Handhabung bei Dauerschuldverhältnissen, schwieriger wird die rechtliche Würdigung für vertraglich festgelegte punktuelle Leistungserbringungen, beispielsweise im Kauf- und Werkvertragsrecht. Selbst ohne vertragliche Beziehungen ergeben sich aus dem Deliktsrecht im Rahmen der Produktbeobachtung Sorgfaltspflichten des Herstellers für die IT-Sicherheit. Durch neue Regelungen im europäischen Recht wird in Zukunft überdies der digitale Verbraucherschutz größer denn je geschrieben. Im Zeitalter von IoT gehört hierzu auch die Sicherheit vernetzter Produkte, auf denen nicht zuletzt auch eine Vielzahl (sensibler) personenbezogener Daten gespeichert ist. Dieser EU-Regulierungsansatz, der in naher Zukunft auch rechtspolitische Wirkung auf B2B-Angebote entfalten könnte, ist grundsätzlich zu begrüßen – trotz der Tatsache, dass sich die Regelungen zu Updatepflichten aus der DID-RL und der WK-RL umfassend vertraglich abbedingen lassen. Allen Regelungen zu IT-sicherheitsbezogenen Updatepflichten gemein ist der bisher noch unbestimmte Zeitraum, für den ein Support zu gewährleisten ist. Geltende Rechtsvorschriften treffen hier (noch) keine konkreten Aussagen. Daher ist es nun an den Herstellern, vertretbare und transparente Nutzungszeiträume für ihre IT-Produkte bis zur Abkündigung und darüber hinaus zu bestimmen, um sowohl für professionelle wie auch für private Anwender eine bessere Planbarkeit der IT-Sicherheit und der zu ihrer Aufrechterhaltung entstehenden Kosten zu ermöglichen.

[36] BGH NJW 1981, 1603, 1604; *Förster* in: BeckOKBGB, § 823, Rn. 740.

Literaturhinweise

[1] Auer-Reinsdorff, Astrid/Conrad, Isabell (Hrsg.), Handbuch IT- und Datenschutzrecht, 2. Auflage, München 2016.

[2] BeckOKBGB, Hau, Wolfgang/Poseck, Roman (Hrsg.), 56. Edition, München 2020.

[3] *Ehle, Kristina/Kreß, Stephan*, Neues IT-Vertragsrecht für digitale Inhalte und Dienste gegenüber Verbrauchern, CR 2019, 723.

[4] Jauernig, Stürner, Rolf (Hrsg.), Bürgerliches Gesetzbuch, 18. Auflage, München 2021.

[5] Lindner-Figura, Jan/Oprée, Frank/Stellmann, Frank (Hrsg.), Geschäftsraummiete, 4. Auflage, München 2017.

[6] Kipker, Dennis-Kenji (Hrsg.), Cybersecurity, München 2020.

[7] *Kumkar, Lea Katharina*, Herausforderungen eines Gewährleistungsrechts im digitalen Zeitalter, ZfPW 2020, 306.

[8] Münchener Anwaltshandbuch zum IT-Recht, Leupold, Andreas/Glossner, Silke (Hrsg.), 3. Auflage, München 2013.

[9] Münchener Kommentar zum Bürgerlichen Gesetzbuch, Säcker, Franz Jürgen/Rixecker, Roland/Oetker, Hartmut/Limperg, Bettina (Hrsg.), Band I und Band VI, jeweils in der 8. Auflage, München 2018/2020.

[10] *Megede, Ekkehard zur*, Bemerkungen zu Rechtsfragen im Bereich der EDV, NJW 1989, 2580.

[11] *Redeker, Helmut*, IT-Recht, 7. Auflage, München 2020.

[12] *Rockstroh, Sebastian/Kunkel, Hanno*, IT-Sicherheit in Produktionsumgebungen, MMR 2017, 77.

[13] *Schimmer, Klaus*, Bewusst sicher Programmieren, DuD 2006, 616.

[14] *Staudemeyer, Dirk*, Auf dem Weg zum digitalen Privatrecht – Verträge über digitale Inhalte, NJW 2019, 2497.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Technischer und juristischer Umgang mit Datenschutz-Katastrophen: Vorbereitung, Krisenmanagement und „Lessons learned“

Joerg Heidrich[1], Dr. Christoph Wegener[2]

Kurzfassung:
Lange wurden IT-(Sicherheits-)Vorfälle als rein technisches Problem betrachtet. Dies hat sich spätestens mit der Einführung der DSGVO geändert, denn deren Anforderungen gehen aus technischer Sicht deutlich über die bisherigen Vorgaben des BDSG hinaus. Vielmehr sind nunmehr durchaus hohe Anforderungen an die Informationssicherheit ein elementarer Bestandteil des Datenschutzes geworden. Dies schlägt sich auch beim Umgang mit IT- und Datenschutzkatastrophen nieder. Was dabei aus technischer und juristischer Sicht zu beachten ist, zeigt der nachfolgende Beitrag, der zugleich praktische Tipps für eine angemessene Vorbereitung auf den Ernstfall gibt.[3]

Stichworte: Aufsichtsbehörden (Datenschutz), BDSG, Bußgelder (Datenschutz), Datenschutz, Datenschutzvorfall, DSGVO, Informationssicherheit, IT-Sicherheit, Meldepflichten, personenbezogene Daten, Privacy by Default, Privacy by Design, Recht auf Vergessen, Risikoabwägung, Schadensersatz, Stand der Technik

1. Einleitung

Das Verhältnis zwischen dem klassischen Datenschutz und der Informationssicherheit wird in der DSGVO neu geregelt. Die jetzigen Bestimmungen gehen weit über die alte Rechtslage hinaus und machen die Informationssicherheit der zur Datenverarbeitung genutzten IT-Systeme zu einem elementaren Bestandteil des Schutzes personenbezogener Daten. Die Änderung steht damit auch in einer Linie mit anderen Regulierungsvorhaben, wie etwa dem (zukünftigen) IT-Sicherheitsgesetz[4]. Sinn und Zweck dieser Regulierungen insgesamt ist es, mit zum Teil sehr detaillierten Vorgaben für eine Erhöhung der Informationssicherheit technischer Systeme zu sorgen.

Im Kern der neuen Regelungen stehen die Anforderungen des Art. 32 DSGVO für die Verarbeitung personenbezogener Daten. Die Vorgaben entsprechen dem Bild einer Waage: Auf der einen Seite (der Waage) stehen *Stand der Technik[5], Implementierungskosten, Zweck und Durchführung der Verarbeitung, sowie Eintrittswahrscheinlichkeit und Schwere des Risikos für die Rechte und Freiheiten der Betroffenen*. Auf Basis dieser Voraussetzungen treffen der Verantwortliche und der Auftragsverarbeiter geeignete technische und organisatorische Maßnahmen, um ein diesem Risiko angemessenes

---

[1] Heidrich Rechtsanwälte Partnerschaftsgesellschaft mbB, Hannover

[2] wecon.it-consulting, Gevelsberg

[3] Dieser Beitrag basiert auf einem Vortrag auf der 28. DFN-Konferenz „Sicherheit in vernetzten Systemen“.

[4] Referentenentwurf zum IT-SiG vom 19. November 2020, abrufbar unter: https://intrapol.org/wp-content/uploads/2020/11/Entwurf-IT-SiG-2.0-19.11.2020.pdf (abgerufen am 6. Januar 2021).

[5] Zur näheren Erläuterung zum „Stand der Technik“ siehe beispielsweise die Handreichung des *Bundesverbands der IT-Sicherheit (TeleTrusT)*, Stand Oktober 2020, abrufbar unter: https://www.teletrust.de/publikationen/broschueren/stand-der-technik/ (abgerufen am 6. Januar 2021).

Schutzniveau (das bildlich auf der anderen Seite der Waage steht) zu gewährleisten – es mithin auszubalancieren.

Praktisch höchst relevant sind dabei die Meldepflichten der Art. 33, 34 DSGVO[6] bei Verstößen gegen die Informationssicherheit und daraus resultierenden Sicherheitspannen. Neben der zentralen Vorschrift des Art. 32 DSGVO enthalten noch weitere Regelungen aus der DSGVO direkt oder indirekt Vorgaben für die Informationssicherheit. Hervorzuheben sind hier etwa die Löschvorgaben des Art. 17 DSGVO[7] sowie die bislang immer noch zu wenig beachteten Anforderungen an „Privacy by Design“ und „Privacy by Default“ des Art. 25 DSGVO[8].

2. Technische und organisatorische Maßnahmen der DSGVO im Überblick

Die Auswahl von angemessenen „technischen und organisatorischen Maßnahmen“[9] nach Stand der Technik ist in der DSGVO an unterschiedlichen Stellen relevant. Grundsätzlich sind diese in Abhängigkeit des Risikos für den Betroffenen zu wählen. Dabei gibt das potenzielle Risiko für den Betroffenen im Fall des Missbrauchs seiner personenbezogenen Daten die Messlatte vor, nicht etwa das Risiko für die die personenbezogenen Daten verarbeitende Organisation. Damit unterscheidet sich die Bewertung beispielsweise von der klassischen Vorgehensweise im Rahmen der ISO 27000, bei der typischerweise die Risiken der Organisation im Fokus stehen.

Bei der Auswahl der technischen und organisatorischen Maßnahmen (TOM) stellt Art. 32 DSGVO auf die drei Kriterien *„Stand der Technik“*[10], *„geeignet“* sowie *„angemessen“* ab, die entsprechend zu erfüllen sind:

- Geeignet ist ein Mittel dann, wenn der damit verfolgte Zweck überhaupt erreicht oder zumindest gefördert werden kann. Konkret bedeutet dies, dass die Ziele des Art. 32 DSGVO mit den Maßnahmen realisiert werden können und dadurch ein angemessenes Schutzniveau – auch langfristig – realisiert werden kann.
- Eine Maßnahme ist im juristischen Sprachgebrauch „angemessen“, wenn „der beabsichtigte Zweck nicht außer Verhältnis zu der Schwere des Eingriffs steht“. Inhaltlich geht es also um die Verhältnismäßigkeit der Mittel. Dabei wird das Schutzniveau nicht absolut, sondern immer im Kontext der jeweiligen Verarbeitung gegenüber dem Risiko für die Rechte und Freiheiten der Betroffenen bewertet. Dadurch kommt auch zum Ausdruck, dass Restrisiken weder ausgeschlossen werden können, noch ausgeschlossen werden sollen.[11]

In den folgenden drei Abschnitten dieses Beitrags wollen wir die aus unserer Sicht wesentlichen Bereiche im Zusammenhang mit der Auswahl von TOM betrachten.

---

6 Eine gute Übersicht dazu bietet *Paal*, ZD 2020, 119-124.

7 In der Praxis höchst relevant durch die Fragen nach der Dauer von Protokollierung und Löschfristen für Backups und Archivierung, siehe dazu etwa *Keppeler/ Berning*, ZD 2017, 314.

8 Dazu: *Hansen,* in: *Simitis/Hornung/Spiecker gen. Döhmann*, DSGVO, Art. 25 Rn. 5-9.

9 Diese werden im Folgenden auch kurz als „TOM“ bezeichnet.

10 Zur weiteren Erläuterung des Begriffs siehe beispielsweise Fn 5.

11 *Pilz* in *Gola*, Art. 32, Rn. 9 ff.

2.1. Maßnahmen zum Löschen von Daten

Das Löschen von Daten spielt aus mehreren Gründen eine zentrale Rolle. Daten sind zu löschen, wenn diese nicht mehr benötigt werden, dies ergibt sich unter anderem aus den Grundsätzen der Verarbeitung nach Art. 5 DSGVO. Speziellere Vorgaben bzgl. des Löschens finden sich in Art. 17 DSGVO. Dort ist unter den genannten Bedingungen ein explizites Recht des Betroffenen auf Löschung der Daten vorgesehen, wörtlich heißt es: *„Die betroffene Person hat das Recht, von dem Verantwortlichen zu verlangen, dass sie betreffende personenbezogene Daten unverzüglich gelöscht werden[…]“*.

Für den Betrieb der die personenbezogenen Daten verarbeitenden IT-Systeme bzw. die entsprechenden IT-Prozesse bedeutet dies vor allem, dass die Möglichkeit der Löschung von Daten von vorneherein zwingend zu berücksichtigen ist. Insbesondere ist durch passende *Löschkonzepte*[12] bereits im Vorfeld zu definieren, unter welchen Umständen welche Daten wann und wie gelöscht werden (müssen). Die Erstellung eines Löschkonzepts ist umso wichtiger, da die DSGVO anders als das BDSG a.F. kein „Sperren“ von Daten mehr vorsieht, beispielsweise für den Fall, dass ein Löschen aus technischen Gründen nicht möglich ist.[13] Technisch ist zudem zu berücksichtigen, dass ein Löschen mit den üblichen Löschkommandos auf Betriebssystemebene wohl nicht die Anforderungen der DSGVO erfüllen dürfte.[14]

Neben der Umsetzung dieser technischen Anforderungen entstehen in der Praxis zudem Probleme dadurch, dass die nach DSGVO zu löschenden Daten gegebenenfalls nach anderen Regularien noch aufzubewahren sind. Hierbei kommt es häufig zu *Interessenskonflikten* mit den Vorgaben des Datenschutzes, etwa im Rahmen von Aufbewahrungspflichten für Handelsbriefe (beispielsweise auch in Form von E-Mails), die sich aus dem Steuerrecht oder vertraglichen Vorgaben ergeben. Auch dieser Aspekt muss in einem Löschkonzept frühzeitig Berücksichtigung finden.

2.2. Maßnahmen zur Datenminimierung

Der Grundsatz der *Datenminimierung* geht weiter als der im BDSG enthaltene Grundsatz der *Datensparsamkeit*. Er besagt, dass die jeweiligen Daten hinsichtlich ihrer Verarbeitungszwecke *„angemessen und erheblich sowie auf das notwendige Maß beschränkt sein müssen“*. Erwägungsgrund 39 führt hierzu aus, dass die Verarbeitung personenbezogener Daten ausschließlich unter der Voraussetzung vorgenommen werden sollte, dass *„die Verarbeitungszwecke nicht mit gleichermaßen wirksamen Methoden auf zumutbare Weise erreicht werden können“*.

Hieraus ergibt sich eine zwingende Beschränkung auf das notwendige Maß – und zwar hinsichtlich der erhobenen Daten sowohl in quantitativer wie auch qualitativer Hinsicht.[15] Konkret bedeutet das, dass die Pflicht zur Datenminimierung sich auf alle Mo-

---

[12] Dazu detailliert: *Faas /Henseler*, BB 2018, S. 2292; *Gründel*, ZD 2019, S. 493

[13] *Keppeler/Berning*, ZD 2017, 314

[14] Vielmehr wird hier ein „sicheres Löschen“ erforderlich sein, das je nach den konkreten technischen Voraussetzungen beispielsweise durch gezieltes Überschreiben der Daten realisiert werden kann.

[15] *Spindler/Dalby* in *Spindler/Schuster*, DSGVO, Art. 5, Rn. 12

dalitäten der Datenverarbeitung (Art und Menge der Daten, Zahl der Verarbeitungsvorgänge und Anzahl der Betroffenen) bezieht und durch Maßnahmen des Datenschutzes in Form einer entsprechenden Technikgestaltung (z. B. Pseudonymisierung und Anonymisierung) ergänzt werden muss.[16]

2.3. Maßnahmen für die Sicherheit der Verarbeitung

Art. 32 DSGVO enthält die zentralen Vorgaben an die Inhalte der TOM. Leider sind diese nicht in allen Punkten entsprechend konkret. So geben Art. 32 Abs. 1 lit. a-d DSGVO zwar einige ausgewählte, mehr oder weniger konkrete Maßnahmen an, die umgesetzt werden sollen, eigentlich versteckt sich dahinter aber viel mehr.

Zum besseren Verständnis gehen wir daher zunächst kurz auf die expliziten Vorgaben des Art. 32 Abs. 1 DSGVO ein:

- Art. 32 Abs. 1 lit. a DSGVO fordert, dass personenbezogene Daten – wenn immer möglich – *pseudonymisiert* und *verschlüsselt* werden. Damit wird zum einen das Prinzip der Datensparsamkeit wieder aufgenommen, zum anderen auch direkt eine technische Maßnahme genannt, die zum Schutz der personenbezogenen Daten regelmäßig umzusetzen ist.
- Art. 32 Abs. 1 lit. b DSGVO fordert, dass die klassischen Schutzziele (der Informationssicherheit) *Vertraulichkeit, Integrität und Verfügbarkeit* der entsprechenden IT-Systeme und Dienste sichergestellt werden und dass diese IT-Systeme und Dienste entsprechend den Anforderungen *„belastbar“* sind, also auch bei einer „intensiven“ Nutzung ordnungsgemäß funktionieren.
- Art. 32 Abs. 1 lit. c DSGVO fordert, dass ein *Notfallkonzept* zur schnellen Wiederherstellung der Verfügbarkeit der und des Zugangs zu den personenbezogenen Daten existiert, inklusive eines Backup-Konzepts für die Daten.
- Art. 32 Abs. 1 lit. d DSGVO fordert schließlich, dass ein *regelmäßiger Prozess zur Überprüfung der getroffenen Maßnahmen* existiert. Dadurch soll zum einen sichergestellt werden, dass die Vorgaben des Art. 32 DSGVO kontinuierlich erfüllt werden, zum anderen ermöglicht dieser Prozess auch einen entsprechenden Nachweis der Umsetzung der Vorgaben gegenüber internen und externen Stellen.

In dieser recht knappen Beschreibung sind allerdings eine Reihe von umzusetzenden Vorgaben versteckt. Letztendlich adressiert Art. 32 Abs. 1 DSGVO nämlich nahezu alle denkbaren Maßnahmen der Informationssicherheit, soweit diese im konkreten Fall geeignet und angemessen sind (vgl. dazu auch Abschnitt 2). Die dort genannten Punkte sind nur Beispiele, die von den Verantwortlichen zu einem Gesamtkonzept für die Informationssicherheit zusammenzuführen sind.

Dazu gehören beispielsweise der *Schutz vor unbefugtem Zugriff* – sei es bei einer Anwendung auf einem Desktop-System oder bei einer Webanwendung – und ein Schutz durch *Verschlüsselung sowohl bei der Speicherung als auch der Kommunikation* von

[16] Spindler/Dalby in Spindler/Schuster, DSGVO, Art. 5, Rn. 12

personenbezogenen Daten. Auch die konsequente *Aktualisierung der verwendeten Systeme und Anwendungen mittels Patches und Updates* sowie die *Vermeidung von Standard-Konfigurationen* – insbesondere im Hinblick auf Passwörter und andere Authentisierungstoken – zählen zu den Anforderungen. Nicht zuletzt fallen auch *regelmäßige Backups inkl. Test derselben* oder die *Durchführung von Pentests*[17] zur frühzeitigen Erkennung von Sicherheitslücken sowie eine *regelmäßige Überprüfung auf Aktualität der Maßnahmen* (Stichwort: Stand der Technik) in die Liste der Anforderungen nach Art. 32 DSGVO. Ergänzend wird beispielsweise auch ein *Change-Management* erforderlich sein, um Änderungen an den IT-Systemen jederzeit nachzuvollziehen und gegebenenfalls auch wieder rückgängig zu machen.[18]

Wie bereits erwähnt, spielt ein risikobasierter Ansatz im Rahmen der Umsetzung von Maßnahmen nach Art. 32 DSGVO eine entscheidende Rolle. Daher wird dieser Aspekt im folgenden Abschnitt dieses Beitrags zusammen mit der Vorgehensweise bei einer Datenschutz-Folgeabschätzung näher beleuchtet.

3. Risk Assessment im Datenschutz aka „Datenschutz-Folgenabschätzung“

Die DSGVO basiert in wesentlichen Punkten auf einem risikobasierten Ansatz.[19] Das bedeutet: Ausgehend von der Gefahr, die den Betroffenen durch die Verarbeitung ihrer Daten jeweils droht, ist ein angemessenes Schutzniveau zu ermitteln und technisch und organisatorisch umzusetzen. Um diese Risiken in einem speziellen Projekt oder Verfahren zu erfassen, abzuschätzen und zu minimieren, enthält die DSGVO das Instrument der *„Datenschutz-Folgenabschätzung (DSFA)“*.[20]

Diese ist ein spezielles Instrument zur Beschreibung, Bewertung und Eindämmung von Risiken. Hierbei müssen typische Schadenszenarien und daraus abgeleitet der Schutzbedarf der Daten überprüft und dokumentiert werden. Eine DSFA ist nach Art. 35 DSGVO durchzuführen, wenn die Form der Verarbeitung, *„insbesondere bei der Verwendung neuer Technologien, aufgrund der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung“* voraussichtlich ein hohes Risiko zur Folge hat. Wichtig ist: Es dreht sich dabei nicht um die Risiken der die Daten verarbeitenden Organisation, sondern um die der betroffenen natürlichen Personen, deren Daten verarbeitet werden.

Notwendig ist eine Folgenabschätzung zum Beispiel stets bei der Verarbeitung von besonderen Kategorien personenbezogener Daten, bei einer Datenverarbeitung in großem Umfang oder beim Zusammenführen und Kombinieren von Daten, etwa bei Big Data-Prozessen. [21] Je größer der Umfang der verarbeiteten Daten auf der einen und das Risiko für die Betroffenen auf der anderen Seite, desto wahrscheinlicher ist die Erforderlichkeit dieser Einschätzung.

---

[17] *Deusch /Eggendorfer*, K&R 2018, S. 223

[18] Einen Überblick über die zu ergreifenden Maßnahmen bieten: *Martini* in *Paal/Pauly*, DSGVO Art. 32 Rn. 43 ff; *Laue* in *Spindler/Schuster*, DSGVO, Art. 32 Rn. 17 ff.

[19] Dazu: *Alt*, DS 2020, 169 mit weiteren Nachweisen

[20] *Laue* in *Spindler/Schuster* Elektron. Medien, DS-GVO Art. 35

[21] Hierfür gibt es sogenannte *Blacklists* der Aufsichtsbehörden, siehe dazu beispielsweise in ZD-Aktuell 2019, 06539; sowie Leitlinien der Art. 29-Datenschutzgruppe, ZD-Aktuell 2017, 05587

Eine DSFA führt man idealerweise in einem *interdisziplinären Team* durch, in dem neben Juristen und Technikern auch Vertreter der jeweiligen Fachabteilungen und natürlich der zuständige Datenschutzbeauftragte beteiligt sind. Inhaltlich identifiziert man im ersten Schritt die Risikoquelle und bestimmt die damit verbundenen Eintrittswahrscheinlichkeiten. Für die so ermittelten Risiken muss man geeignete Abhilfemaßnahmen finden und diese auf den jeweiligen Sachverhalt anpassen.

Kommen die Verantwortlichen bei einer DSFA zu dem Ergebnis, dass die geplante Datenverarbeitung ein zu hohes Risiko für die Betroffenen beinhaltet und dies auch nicht ausreichend mit wirksamen technischen und organisatorischen Maßnahmen einzugrenzen ist, so müssen sie gemäß Art. 36 DSGVO die Aufsichtsbehörde konsultieren. Diese unterbreitet daraufhin schriftliche Empfehlungen oder kann von ihren Untersuchungsbefugnissen nach Art. 58 DSGVO Gebrauch machen.

4. Von der IT-Katastrophe zum Datenschutz-GAU

Da in den vielen IT-Prozessen direkt oder indirekt personenbezogene Daten verarbeitet werden, endet ein IT-Sicherheitsvorfall meist auch in einem Datenschutzvorfall. Besonders problematisch wird dies immer dann, wenn beispielsweise Unberechtigte große Mengen von Daten abgreifen können oder langfristigen Zugriff erhalten haben oder wenn besondere Kategorien personenbezogener Daten betroffen sind.

Technisch sind häufig unzureichende Zugriffskontrollen in unterschiedlicher Ausprägung die Ursache für etwaige Datenschutzverstöße. Dazu gehören beispielsweise vergessene Netzwerkfreigaben, User-Accounts mit fehlerhaften Berechtigungen, aber auch durch einfache Session-ID nur unzureichend umgesetzte Zugriffsbeschränkungen im Web-Bereich.

Werden solche Sicherheitsprobleme bekannt oder gibt es gar konkrete Angriffe, beispielsweise auf die IT-Systeme einer Organisation, gilt es schnellstmöglich zu handeln. Dabei ist wie bei jeder Art von Sicherheitsvorfall grundsätzlich zwischen den beiden Aspekten „Erst-Analyse des Vorfalls“ und „Schnellstmögliches Ergreifen von Gegenmaßnahmen“ abzuwägen. Sind von dem Sicherheitsvorfall potenziell auch personenbezogene Daten betroffen, ist bei allen Schritten zwingend zu berücksichtigen, dass die entsprechende Meldung an die zuständige Aufsichtsbehörde in der Regel innerhalb von 72 Stunden zu erfolgen hat.

4.1. Für einen Datenschutz-GAU besonders relevante Aspekte

Neben den üblichen Maßnahmen des Intrusion-Response-Prozesses sind im Kontext dieses Beitrags dabei vor allem folgenden Aspekte relevant:

- Jedes Unternehmen muss sich bewusst sein, dass eine solche Datenpanne jederzeit geschehen kann und dass die Frage nicht lautet „Kann uns das passieren?“, sondern eher „Wann wird es passieren?“. Dementsprechend gilt es, sich *auf den Ernstfall vorzubereiten*. Hierzu gehört es insbesondere, vorab ein *interdisziplinäres Team* zu bilden, welches „in der Materie“ steckt, also insbesondere die relevanten Geschäftsprozesse kennt, und sich kurzfristig zusammenfinden kann.

- Die *Geschäftsleitung* und der *zuständige Datenschutzbeauftragte* sind unverzüglich zu informieren, sobald die Möglichkeit besteht, dass auch personenbezogene Daten von dem Sicherheitsvorfall betroffen sein können. Die weitere (technische) Analyse des Vorfalls erfolgt dann grundsätzlich in enger Abstimmung mit dem Datenschutzbeauftragten.
- Je nach Sachlage sind gegebenenfalls frühzeitig weitere, *externe Experten aus dem Bereich der rechtlichen und technischen Aspekte des Datenschutzes* hinzuzuziehen, um Kollateralschäden bei der Analyse und Kommunikation zu vermeiden.
- Es ist dann schnellstmöglich festzustellen, ob, seit wann und in welchem Umfang personenbezogene Daten von dem Sicherheitsvorfall betroffen sind und um welche konkreten Datenarten es sich dabei handelt. Dabei ist auch zu ermitteln, ob und gegebenenfalls (seit) wann personenbezogene Daten durch den Sicherheitsvorfall abgeflossen sind, um welche konkreten Datenarten es sich dabei handelt und wer davon betroffen ist.
- In Abstimmung mit dem Datenschutzbeauftragten und den möglicherweise hinzugezogenen externen Experten sind erste bzw. weitere Gegenmaßnahmen technischer und organisatorischer Art zu ergreifen. In jedem Fall sollte ein *enges Monitoring der betroffenen IT-Systeme* erfolgen, um einen weiteren Missbrauch – insbesondere von personenbezogenen Daten – schnellstmöglich erkennen und entsprechend minimieren zu können.
- Aus den Ergebnissen sind entsprechende Listen und Dokumente zur Kommunikation mit den Aufsichtsbehörden zu erstellen. Die Kommunikation mit den Aufsichtsbehörden erfolgt in der Regel ausschließlich durch den vorab *festgelegten Experten*, etwa den zuständigen Datenschutzbeauftragten oder einen externen Berater. Dabei ist die Frist von nur 72 Stunden ab Kenntnis des Vorfalls zu berücksichtigen.
- In Abstimmung mit der Geschäftsleitung erfolgt zudem die *Kommunikation mit allen zu informierenden Stellen*. Dazu gehört die bereits erwähnte Kommunikation zur zuständigen Aufsichtsbehörde, aber auch die zu externen und internen Partnern und häufig natürlich auch den eigenen Mitarbeitern.
- Je nach Ausmaß des Vorfalls kann das Hinzuziehen eines *Experten für Krisenkommunikation* ratsam sein. Dies gilt insbesondere dann, wenn eine große Anzahl von Kunden zu informieren ist und damit zu rechnen ist, dass der Vorfall die Medien beschäftigen wird.
- Gerade bei größeren Vorfällen, einer Vielzahl von Betroffenen oder wenn höchst sensible Daten (insbesondere also Daten nach Art. 9 DSGVO) öffentlich wurden, werden die Aufsichtsbehörden im Nachgang der Meldung detaillierte Nachfragen zum Ablauf des Vorfalls, den bestehenden TOM und den ergriffenen Schutzmaßnahmen stellen. Hierauf muss das Unternehmen inhaltlich wie personell vorbereitet sein.
- Ist der Vorfall als besonders gravierend zu bewerten, besteht auch das Risiko von Bußgeldern oder Schadensersatzansprüchen durch die Betroffenen. Soweit dies

im Raum steht, sollte sich das Unternehmen rechtzeitig mit entsprechend *spezialisierten juristischen Beratern* in Verbindung setzen.

5. Meldepflichten für Sicherheitsvorfälle

Die DSGVO sieht in Art. 33 und Art. 34 Meldepflichten für Datenschutzpannen vor. Diese beziehen sich in erstgenannter Vorschrift zunächst auf eine Benachrichtigung der zuständigen Landesdatenschutzbehörde.[22] Art 34 DSGVO geht noch weiter und fordert eine Mitteilung an alle durch die Datenpanne Betroffenen, also im Regelfall an Kunden und/oder an Mitarbeiter.

5.1. Meldepflichten an die Behörde

Art. 33 DSGVO ist bei einer Verletzung des Schutzes personenbezogener Daten anwendbar, die *„voraussichtlich zu einem Risiko für die Rechte und Freiheiten natürlicher Personen führt“*. Damit beschränken sich die Anforderungen also nicht auf besonders schwere Datenlecks, sondern umfassen ohne größere Einschränkung nahezu jeden Verlust persönlicher Informationen. So gehören zu den meldepflichtigen Vorgängen nach Ansicht der niedersächsische Datenschutzbehörde auch Verstöße wie die unrichtige Entsorgung personenbezogener Daten auf Papier, das Anzeigen der Daten falscher Betroffener in einem Kundenportal oder die mündliche Bekanntgabe personenbezogener Daten an die falsche Person.

Die Meldung muss unverzüglich und *„möglichst binnen 72 Stunden“* ab Kenntnis des Data Breach erfolgen. Diese Frist ist extrem kurz und lässt Unternehmen kaum Zeit für eine entsprechende Aufarbeitung der Geschehnisse. Welche Inhalte die zu übermittelnde Meldung aufweisen muss, regelt Art. 33 Abs. 3 der DSGVO.

5.2. Meldepflichten an die Betroffenen

Noch weiter gehen die Pflichten des Art. 34 DSGVO. Die Vorschrift setzt voraus, dass die Sicherheitspanne *„voraussichtlich ein hohes Risiko für die persönlichen Rechte und Freiheiten natürlicher Personen zur Folge“* hat. Diese Herangehensweise stellt eine weitere direkte Umsetzung des risikobasierten Ansatzes[23] der DSGVO dar.

Wann genau ein solches hohes Risiko für die Betroffenen anzunehmen ist, definiert die DSGVO nur im Ansatz und auf Basis einer Prognose (*„voraussichtlich“*). Bei hoher drohender Schadensschwere genügt bereits eine geringe Eintrittswahrscheinlichkeit. Umgekehrt überschreitet auch ein geringer zu erwartender Schaden die Risikoschwelle, wenn er mit hoher Wahrscheinlichkeit eintritt.[24] Dieser Grundsatz ergibt sich bereits aus Erwägungsgrund 75 der DSGVO. Ansatzpunkte für ein hohes Risiko liegen auch dann vor, wenn besondere Kategorien personenbezogener Daten nach Art. 9 DSGVO Gegenstand der Verletzung sind[25].

---

[22] Dazu: *Paal*, ZD 2020, 119; *Becker*, ZD 2020, 175.

[23] Dazu ausführlich: *Schröder*, ZD 2019, 503.

[24] *Martini,* in: *Paal/Pauly*, Art. 34 Rn. 30

[25] *Laue,* in: *Spindler/Schuster*, Art. 34, Rn. 6.

Die Benachrichtigung sollte nach Erwägungsgrund 86 eine in klarer und einfacher Sprache verfasste Beschreibung der Art der Datenverletzung sowie an die betroffene natürliche Person gerichtete Empfehlungen zur Minderung etwaiger nachteiliger Auswirkungen des Data Breach enthalten. Solche Benachrichtigungen der betroffenen Person sollten stets *„unverzüglich“,* also so rasch wie möglich, in Absprache mit der Aufsichtsbehörde und nach Maßgabe der von dieser oder von anderen zuständigen Behörden wie beispielsweise Strafverfolgungsbehörden erteilten Weisungen erfolgen.

6. Sanktionen

Die Aufsichtsbehörden haben verschiedene Sanktionsmöglichkeiten, mit denen sie auf Verstöße gegen die DSGVO reagieren können. Diese reichen vom Aussprechen von Verwarnungen oder Rügen über die Anordnung der Aussetzung der Datenverarbeitung bis hin zu Bußgeldern.[26]

6.1. Bußgelder

Das Schreckgespenst der DSGVO sind unzweifelhaft die Bußgelder, die von den Behörden nach Art. 83 DSGVO verhängt werden können. Diese Geldstrafen müssen *„in jedem Einzelfall wirksam, verhältnismäßig und abschreckend“* sein. Vor allem das Merkmal der Abschreckung ist dabei ein neuer Bestandteil des Datenschutzes. Auch die Höhe der Geldbußen hat sich dramatisch verändert. Diese können nun bis zu 20 Mio. EUR oder bis zu 4% des weltweit erzielten Jahresumsatzes des vorangegangenen Geschäftsjahrs eines Unternehmens betragen.

6.2. Schadensersatzansprüche

Neben einem hohen Bußgeld droht bei Datenschutzpannen auch weiterer, in der öffentlichen Diskussion nach wie vor wenig bekannter „Ärger“ nach Art. 82 DSGVO. Danach haben potenziell Betroffene eines Data Breach wegen eines Verstoßes gegen DSGVO-Vorgaben einen Schadensersatzanspruch für materielle oder immaterielle Schäden.[27] Erforderlich ist ein Verschulden des Verantwortlichen oder auch des Auftragsverarbeiters.

7. „Lessons learned“ aka „Checkliste zur Vorbereitung“

In diesem Abschnitt wollen wir nun nochmals zusammenfassen, welche Aspekte bereits proaktiv zu berücksichtigen sind, um Datenschutzvorfälle möglichst zu vermeiden und im Falle des Falles angemessen reagieren zu können. Basierend auf den Erfahrungen aus Praxisfällen und unseren allgemeinen Empfehlungen entsteht so eine praxisbezogene Checkliste für die richtige Vorbereitung.

1. *Grundsätzliche Beschränkung auf wirklich notwendige Daten*
   Das klingt zunächst simpel, wird aber immer wieder vernachlässigt. Häufig kommt es zu einer Sammlung von Daten, die für den Geschäftsprozess nicht notwendig und daher unbedingt zu vermeiden sind.

[26] Einen Überblick dazu bietet *Laue,* in: *Spindler/Schuster*, Art. 32, Rn. 28ff

[27] Dazu im Überblick: *Wybitul*, NJW 2019, 3265; *Paal*, MMR 2020, S. 14.

2. *Berücksichtigung und regelmäßige Überprüfung der Löschfristen*
   Daten, die noch in den Produktionsprozessen genutzt werden, müssen regelmäßig auf etwaige Löschanforderungen überprüft werden. Dazu ist es zum einen notwendig, die entsprechenden Fristen überhaupt zu kennen, und zum anderen dann technisch die Voraussetzung zu schaffen, die Datenbestände regelmäßig und gegebenenfalls automatisiert gegen diese Fristen abgleichen zu können. Auch dieser Punkt wird immer wieder vernachlässigt, insbesondere werden Löschanforderungen und -prozesse häufig in den – soweit überhaupt vorhandenen – Backup-Konzepten nicht angemessen berücksichtigt.
3. *Angemessener Zugriffsschutz nach Stand der Technik*
   Es sind geeignete und angemessene technische und organisatorische Maßnahmen nach Stand der Technik zu ergreifen, um das Risiko für einen Missbrauch der personenbezogenen Daten bestmöglich zu reduzieren. Für die konkrete Ausgestaltung verweisen wir hier auf die Inhalte aus Abschnitt 2. Die korrekte Umsetzung und Aktualität dieser Maßnahmen sind zudem regelmäßig zu kontrollieren und zu überprüfen. Dies sollte vorzugsweise nicht nur durch interne Mitarbeiter erfolgen, sondern beispielsweise auch durch externe Fachleute ergänzt werden, etwa durch Pentests. Nur dadurch kann nachgewiesen und zudem langfristig sichergestellt werden, dass die Vorgaben – allen voran der *„Stand der Technik“* – tatsächlich eingehalten werden.
4. *Gepflegtes Verzeichnis der Verarbeitungstätigkeiten*
   Ein *detailliertes Verarbeitungsverzeichnis* nach Art. 30 DSGVO, das auf dem aktuellen Stand gehalten wird, ist die Grundvoraussetzung für die Beantwortung der Frage, wo und wie die Daten in der Organisation verarbeitet werden. Es ist Voraussetzung, um bei einem potenziellen IT-Sicherheitsvorfall zeitnah abschätzen zu können, ob und in welchem Umfang überhaupt personenbezogene Daten betroffen sind. Nur so lassen sich die Auskünfte an die Aufsichtsbehörden innerhalb der knappen Frist von 72 Stunden erteilen. Auch die Auswahl von möglichen Gegenmaßnahmen vereinfacht sich, wenn die aktuell vorhandenen TOM ausreichend dokumentiert sind und dann bewertet werden können.
5. *Proaktive Kontaktpflege zu Experten und den Aufsichtsbehörden*
   Für größere Zwischenfälle sind in den meisten Unternehmen die Bordmittel nicht ausreichend. Es besteht vorab ein Kontakt zu *Fachleuten aus dem technischen und juristischen Bereich* und diese können kurzfristig aktiviert werden. Stehen Bußgelder oder Schadensersatzansprüche im Raum, ist auch der rechtzeitige Kontakt zu *spezialisierten Datenschutz-Anwälten* unerlässlich. Im konkreten Fall ist es auch zu spät, *Kontakte zu den Aufsichtsbehörden* zu knüpfen. Eine proaktive Kommunikation schafft auch hier Vertrauen für den Fall der Fälle und reduziert entsprechende Probleme in der Kommunikation.
6. *Vorbereitet sein, Teams bilden und Übungen durchführen*
   Es ist ein *interdisziplinäres Team* notwendig, das sich mit der Thematik auskennt und weiß, wie kurzfristig reagiert werden muss und kann. Es sollte dabei auch einen Ansprechpartner – sowohl innerhalb des Unternehmens in Form eines *Ko-*

ordinators als auch einen *Fachmann zur Kommunikation mit den Behörden* – geben, der vorab benannt ist. Sofern damit zu rechnen ist, dass die Medien auf den Vorfall aufmerksam werden, sollten auch für diesen Bereich entsprechende *Kommunikationsprofis* benannt werden. Hier kann – gerade bei in diesem Bereich wenig erfahrenen Unternehmen – sehr schnell eine schlechte Außenwirkung entstehen, die sich richtig vorbereitet vermeiden lässt.

8. Fazit

„Richtige Vorbereitung ist die halbe Miete“ – diese Aussage gilt uneingeschränkt im Querschnittsfeld von Datenschutz und Informationssicherheit. Nur wenn die Anforderungen des Datenschutzes auch mit Blick auf einen möglichen Datenschutzvorfall bereits im Vorfeld angemessen berücksichtigt werden, können die Vorgaben der DSGVO sinnvoll umgesetzt und erfüllt werden. Dies gilt schon allein hinsichtlich der enorm kurzen Zeitspanne von 72 Stunden, innerhalb der eine Meldung an die zuständige Aufsichtsbehörde zu erfolgen hat.

Dazu ist es notwendig, neben den klassischen Anforderungen des Datenschutzes auch die korrespondierenden Anforderungen an die Informationssicherheit (proaktiv) umzusetzen und zu dokumentieren; die entsprechenden Vorgaben finden sich in Art. 32 DSGVO. Da sich dabei in der Regel zahlreiche Überschneidungen ergeben, sollten die beiden Themen aber auch grundsätzlich als „Hand-in-Hand“-Prozess verstanden werden. Denn nur dann können die zahlreichen Synergien, die Datenschutz und Informationssicherheit haben, sinnvoll genutzt, Prozesse entsprechend verzahnt und damit letztendlich umfangreich Ressourcen eingespart werden.

Literaturhinweise

[1] Alt, Ulrich, Datensicherheit, Datenschutz und Technik – ein risikoorientierter Ansatz: DS 2020, 169.

[2] Becker, Franz: Meldungen nach Art. 33 DS-GVO, Voraussetzungen der Meldepflicht und die Doppelrolle der Aufsichtsbehörden, ZD 2020, S. 175.

[3] Deusch, Florian/Eggendorfer, Tobias: Penetrationstest bei Auftragsverarbeitung, K&R 2018, S. 223.

[4] Faas, Thomas/Henseler, Maren: Speicherdauer und Aufbewahrungsfristen unter der DSGVO, BB 2018, S. 2292.

[5] Forgó, Nikolaus/Helfrich, Marcus/Schneider, Jochen: Betrieblicher Datenschutz, 3. Auflage, München 2019.

[6] Gola, Peter: DS-GVO, 1. Auflage, München 2017.

[7] Gründel Achim: Ermittlung des Löschbedarfs bei unstrukturierten Datenbeständen, ZD 2019, S. 493.

[8] Jandt, Silke/Steidle, Roland: Datenschutz im Internet, 1. Auflage, München 2018.

[9] Keppeler, Lutz Martin/Berning, Wilhelm: Technische und rechtliche Probleme bei der Umsetzung der DS-GVO-Löschpflichten, ZD 2017, S. 314.

[10] Paal, Boris: Meldepflicht bei Datenschutzverstößen nach Art. 33 DS-GVO, ZD 2020, S. 119

[11] Paal, Boris: Schadensersatzansprüche bei Datenschutzverstößen, Voraussetzungen und Probleme des Art. 82 DS-GVO, MMR 2020, S. 14.

[12] Paal, Boris/Pauly, Daniel: Datenschutzgrundverordnung, Kommentar, 2. Auflage, München 2018.

[13] Simitis, Spiros/Hornung, Gerrit/Spiecker gen. Döhmann, Indra: Datenschutzrecht, Datenschutzrecht: DSGVO mit BDSG, 1. Auflage, Baden-Baden 2019.

[14] Spindler, Gerald/Schuster, Fabian: Recht der elektronischen Medien, 4. Auflage, München 2019.

[15] Schröder, Markus: Der risikobasierte Ansatz in der DS-GVO, ZD 2019, S. 503.

[16] Wybitul, Tim: Immaterieller Schadensersatz wegen Datenschutzverstößen – Erste Rechtsprechung der Instanzgerichte, NJW 2019, S. 3265.

[17] Wybitul, Tim: Vermeidung von DS-GVO-Risiken nach Datenpannen und Cyberangriffen, NJW 2020, 2577.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Staatliche Regulierung des Internets in Russland

Verena Wingerter[1]

Kurzfassung:
Die Debatte über ein souveränes russisches Internet, kurz RuNet, wird spätestens seit dem Inkrafttreten des Gesetzes über ein Autonomes Internet im November 2019 vermehrt öffentlich geführt. Um diese Debatte über RuNet verstehen zu können, ist es wichtig, die Entwicklung der russischen Internetgesetze nachzuvollziehen und zu kontextualisieren. Seit den politischen Protesten 2011 wurden verschiedene Gesetze zur staatlichen Regulierung des Internets erlassen, die weitreichende Konsequenzen für Privatpersonen und Unternehmen haben. Für internationale Unternehmen ist das Verständnis der russischen Rechtslage und der Auswirkungen auf Betriebsabläufe sowie Wirtschaftsbeziehungen von großer Bedeutung. Der vorliegende Beitrag stellt die wichtigsten Internetgesetze vor und ordnet sie in den politischen Kontext ein. Des Weiteren werden die technische Umsetzung der rechtlichen Anforderungen analysiert und Risiken für Unternehmen mit Russlandbezug evaluiert. Der Beitrag entwickelt drei Szenarien zur zukünftigen Risikoeinschätzung.

Stichworte: DPI (Deep Packet Inspection / Russland), Gesetz über ein Autonomes Internet (Russland), Internetkontrolle, Internetregulation, ISPs, Nationales Internet (Russland), Roskomnadzor, RuNet, Russland, SORM (Russland), Sovereign Internet Law, Yarovaya Gesetze

1. Einführung

Die Russische Föderation ist nicht nur ein wichtiger Akteur in der internationalen Staatengemeinschaft und in globalen Institutionen, sondern auch ein bedeutender Markt für viele europäische Unternehmen. Einen wichtigen Aspekt der internen und internationalen Kommunikation stellt das Internet dar, auf das viele Organisationen und Privatpersonen angewiesen sind. Im letzten Jahrzehnt hat der russische Staat viele Gesetze erlassen, die Internetnutzung und Datentransfer regulieren und die eine besondere Analyse und Betrachtung erfordern. Hierbei wird klar, dass sich Russland im internationalen Kontext stark für die Entwicklung neuer Internetnormen einsetzt, während Gesetze auf nationaler Ebene auf eine engere Kontrolle und auf die Etablierung eines nationalen Internets abzielen.

Zum besseren Verständnis der russischen Internetregulation ist es ratsam, die Gesetzesänderungen der vergangenen Jahre und die daraus folgenden technischen Auswirkungen zu evaluieren. In diesem Zusammenhang ist insbesondere das Gesetz über ein Autonomes Internet von 2019 von Bedeutung, allerdings haben auch die Antiterrorismus Gesetze von 2016, sowie die zahlreichen Gesetze zur De-Anonymisierung sowie dem Blocken von „extremistischen“ Inhalten die russische Rechtslage maßgeblich geprägt. Neben dem Rechtsrahmen ist auch eine Analyse der technischen Umsetzung durch staatliche Stellen und der zur Verfügung stehenden Kapazitäten wichtig, welche die Risiken für Unternehmen mit russischen Niederlassungen oder russischen Partnern prägen. Diese Risiken können grob als Betriebsstörungen, Wirtschaftsspionage und höhere Be-

[1] DCSO Deutsche Cyber-Sicherheitsorganisation | Hertie School

triebskosten klassifiziert werden. Zudem ist die Evaluierung von drei Szenarien zur weiteren Entwicklung des russischen Internets ratsam, um eine Einschätzung der zukünftigen Risiken zu erhalten. Während es möglich ist, dass Russland sich für eines von zwei „Extremen“ entscheidet, kann davon ausgegangen werden, dass das wahrscheinlichste Szenario eine russische Rechtslage ist, die mit dem aktuellen chinesischen Internet vergleichbar ist.

2. Gesetzeslage

Der russische Rechtsrahmen für das Internet hat sich über die vergangenen Jahre stark verändert. Im Folgenden werden die wichtigsten Entwicklungen und ihre Bedeutung für das russische Internet ausgeführt.

2.1. Staatliche Kontrolle durch SORM

Ein zentraler Baustein des russischen Internets ist SORM, das ‚System technischer Mittel zur Gewährleistung der Funktionen der operativ-untersuchenden Maßnahmen.‘ SORM wurde in der UdSSR entwickelt, um eine Überwachung der Telekommunikation zu erlauben. 1998 wurden die Berechtigungen für SORM-2 auf das Internet ausgeweitet,[2] wovon unter anderem der russische Inlandsgeheimdienst FSB unter dem damaligen Direktor Vladimir Putin profitierte, der Zugriff auf SORM-Daten bekam.[3]

In Vorbereitung der Olympischen Winterspiele in Sochi 2014 wurde SORM mit neuen Befugnissen ausgestattet: Während SORM-1 weiterhin für das Sammeln und Auswerten der Telefonkommunikation verantwortlich ist, bekam SORM-2 weitere Befugnisse, welche die inhaltliche Überwachung sowie die Auswertung von Internetverbindungen und Datennetzwerken beinhalten. Mit SORM-3 wurde zudem ein System geschaffen, welche das Nachverfolgen und Speichern von jeglichen Kommunikationsdaten für ausgewählte Internetnutzer und IP-Adressen zulässt. SORM-3 analysiert hierfür den Internetverkehr konstant und erstellt individuelle Nutzerprofile.[4] Seit 2014 ist SORM zudem berechtigt, auch soziale Netzwerke auszulesen.

Eine Technologie, die oft in Verbindung mit SORM diskutiert wird, ist Deep Packet Inspection (DPI). DPI ermöglicht die Auswertung von Datenpaketen auf Header und Content Ebene, was vergleichbar damit ist, die Adresse sowie den Inhalt eines Briefes zu lesen. Diese Monitoring- und Filtering-Technologie wird von Internet Service Providern (ISPs) seit Mitte der 2000er eingesetzt, um die Bandbreitennutzung zu optimieren. Seit 2012 wird DPI von den Providern allerdings auch für die Implementierung einer staatlich regulierten Blockliste genutzt.[5] Die Blockliste, die auf einem Gesetz zum Kin-

[2] Jen Tracy, “Beware FSB Surveillance Of Internet,” Moscow Times, March 16, 1999, http://www.libertarium.ru/l_sormpr_mtimes.

[3] Andrei Soldatov and Irina Borogan, The Red Web: The Kremlin’s War on the Internet (New York: PublicAffairs, 2017), 170.

[4] Vladimir Khazov, “COPM-1, COPM-2, COPM-3: особенности и отличия,” Vas Expert (blog), November 1, 2016, https://vasexperts.ru/blog/osobennosti-i-otlichiya-sorm.

[5] Andrei Soldatov and Irina Borogan, The Red Web: The Kremlin’s War on the Internet (New York: PublicAffairs, 2017), 212.

desschutz basiert, erfolgt auf IP-Adressen und URL-Level, was für Plattformen wie Youtube bedeutet, dass diese aufgrund eines einzelnen Beitrags oder Videos komplett geblockt werden würden. Aus diesem Grund wird, auch wenn gesetzlich nicht vorgeschrieben, DPI zum gezielten Blocken von Inhalten seitens der ISPs verwendet.[6] Im Zuge neuer Antiterrorismus-Gesetze 2016 wurden weitere Bestimmungen geschaffen, welche die Speicherung von Kommunikationsdaten von bis zu drei Jahren vorsehen, weswegen Provider auf DPI zurückgreifen, um die Menge an Daten akkumulieren zu können. Dies ist jedoch nicht gesetzlich vorgeschrieben.

Erst das neue Gesetz über ein Autonomes Internet von 2019 sieht die Installation von „technischen Mitteln" vor, „um Bedrohungen der Stabilität, Sicherheit und Integrität" des Internets zu begegnen, welche allgemeinhin als Blackboxes bezeichnet werden.[7] Diese Blackboxes sollen Berichten zufolge DPI Kapazitäten enthalten, was durch staatliche Tests der bestehenden DPI Optionen gestützt wird.[8] Auch die Tatsache, dass das technische Equipment vom Staat gezahlt wird anstatt wie bisher üblich durch die Provider selbst, lässt Rückschlüsse darauf zu, dass mit dem neuen Gesetz von 2019 DPI Kapazitäten in das staatliche System SORM eingegliedert werden sollen.

Das Gesetz über ein Autonomes Internet ist auch aus weiteren Gründen interessant: Neben der Notwendigkeit der Installation der Blackboxes etabliert es die rechtliche Grundlage für ein nationales Domain Name System (DNS). Des Weiteren schreibt es vor, dass ISPs Internetverkehr nur noch durch staatlich zugelassene Internet Exchange Points (IXPs) routen dürfen.[9] Diese Bestimmungen zielen darauf ab, Russland die Möglichkeit zu geben, das russische Internet vom globalen Internet abzukoppeln und zu isolieren. Die Isolation des russischen Internets ist als Notfallmaßnahme bei besonderen Bedrohungen vorgesehen. Erste Tests dazu fanden bereits im Dezember 2019 statt.[10]

2.2. Roskomnadzor und der staatliche Einfluss auf ISPs und IT-Firmen

Zentral für Russlands Bestreben, das Internet zu regulieren, ist der Föderale Dienst für die Aufsicht im Bereich der Kommunikation, Informationstechnologie und Massenkommunikation, Roskomnadzor. Roskomnadzor wurde als eigenständige Behörde gegründet und ist verantwortlich für die Implementation von Gesetzen bezüglich des Internets. Seit 2012 ist Roskomnadzor für die Umsetzung der Blocklist zuständig. Wird ein Verstoß gegen das Gesetz festgestellt, benachrichtigt Roskomnadzor die betroffenen Verantwortlichen, welche verpflichtet sind, die Inhalte innerhalb von drei Tagen zu löschen. Geschieht dies nicht, fügt Roskomnadzor die IP Adresse und URL der betroffenen Seite der Blockliste hinzu, welche von ISPs zweimal täglich heruntergeladen und

---

[6] Ibid, 167-169.

[7] Bundesgesetz vom 1. Mai 2019 Nr. 90-FZ, https://rg.ru/2019/05/07/fz90-dok.html.

[8] „Как Роскомнадзор тестирует DPI," TelecomTimes, March 29, 2019, https://telecomtimes.ru/2019/03/roskomnadzor-dpi/.

[9] Alena Epifanova, "Deciphering Russia's 'Sovereign Internet Law': Tightening Control and Accelerating the Splinternet," DGAP Analysis, January 16, 2020, https://dgap.org/en/research/publications/deciphering-russias-sovereign-internet-law.

[10] "Учения по устойчивости Рунета прошли успешно," Interfax, December 23, 2019, https://www.interfax.ru/russia/689098.

geblockt werden muss.[11] Bekannt geworden ist der gescheiterte Versuch, die Messaging App Telegram zu blockieren, welche dem Bann aber durch geschicktes Domain Fronting ausweichen konnte bis Roskomnadzor diese Entscheidung 2020 revidierte.[12] Durch die Verwaltung der Blockliste ist Roskomnadzor eine zentrale Institution, die staatlichen Einfluss auf ISPs und IT-Firmen in Russland ausüben kann.

Die Blockliste ist jedoch nicht die einzige staatliche Anforderung an Provider. Um neue Auflagen zu erfüllen, mussten ISPs 2012 und 2014 technisches Equipment auf eigene Kosten installieren, das ihnen keinerlei Einblick gewährt. Mit den Antiterrorgesetzen von 2016 wurden Investitionen in Höhe von mehr als 60 Milliarden Euro notwendig, die nicht vom Staat übernommen wurden.[13] Während ISPs diese hohen Investitionen für den Zugriff des Staates auf die Internetkommunikation tätigen, genießen sie jedoch zu wenig Vertrauen, um selbst Einblick in die Überwachung zu erhalten.[14] Grundsätzlich lässt sich festhalten, dass staatliche Auflagen starke Konsequenzen für den russischen IT-Markt haben. Als Beispiel lassen sich das Gesetz zur lokalen Datenspeicherung von 2015 oder das Gesetz von 2019, welches vorinstallierte russische Software auf technischen Geräten vorsieht, nennen. Die Politik des russischen Staates wird insbesondere am Beispiel von Yandex deutlich: Der größte russische Internetkonzern, der Services von Google und Amazon, sowie Uber und Lieferando für Russland anbietet, hat nur begrenzten Einfluss und beugt sich dem Druck der Politik.[15] Andere Unternehmen werden zur Geschäftsübernahme gedrängt oder gründen „Konglomerate, kontrolliert entweder direkt durch die Regierung oder durch andere Akteure, welchen der Kreml die Weiterentwicklung der russischen Internetpolitik zutraut."[16]

2.3. Regulierung von Verschlüsselungstechnologien

Eine weitere Entwicklung ist die Regulierung von Verschlüsselungstechnologien, die auf ein Verbot von Anonymisierungstechnologien abzielt. Schon 2016 wurde im Rahmen der Antiterrorgesetze beschlossen, dass alle verschlüsselten Daten entweder durch Government Backdoors oder durch extra bereitgestellte Encryption keys für staatliche Behörden zugänglich sein müssen.[17] Seit 2017 ist es zudem gesetzlich verboten, anonym

---

11 Daniil Turovsky, "This is how Russian Internet censorship works," Meduza, August 13, 2015, https://meduza.io/en/feature/2015/08/13/this-is-how-russian-internet-censorship-works.

12 Isabelle Khurshudyan, "How the founder of the Telegram messaging app stood up to the Kremlin – and won," The Washington Post, June 28, 2020, https://www.washingtonpost.com/world/europe/russia-telegram-kremlin-pavel-durov/2020/06/27/4928ddd4-b161-11ea-98b5-279a6479a1e4_story.html.

13 "Yarovaya Law: One Year After," Digital.Report Analytica, April 24, 2017, https://analytica.digital.report/en/2017/04/24/yarovaya-law-one-year-after/.

14 Andrei Soldatov and Irina Borogan, The Red Web: The Kremlin's War on the Internet (New York: PublicAffairs, 2017), 73.

15 Gideon Lichfield and Evan Gershkovich, "Podcast: How Russia's everything company works with the Kremlin," MIT Technology Review, September 30, 2020, https://www.technologyreview.com/2020/09/30/1009174/podcast-russia-yandex-kremlin/.

16 Original: "conglomerates, controlled either directly by the government or by other actors trusted to advance Kremlin internet policy." Dylan Myles-Primakoff, Justin Sherman, "Russia's Internet Freedom Shrinks as Kremlin Seizes Control of Homegrown Tech," Foreign Policy, October 26, 2020, https://foreignpolicy.com/2020/10/26/russia-internet-freedom-kremlin-tech/.

17 Oleg Schindler, "The Yarovaya Law: One Year After," Digital.Report Analytica, 24 April 2017, https://analytica.digital.report/en/2017/04/24/yarovaya-law-one-year-after/.

geblockte Inhalte aufzurufen, was als Verbot von Virtual Private Networks (VPNs) sowie von Tor verstanden wird. Die Nutzung von VPNs ist allerdings weiterhin gestattet, wenn Nutzer vorherbestimmt werden, die Zugriff auf geblockte Inhalte erhalten, und wenn VPNs für die technische Unterstützung des Betriebs genutzt werden.[18] Zudem wurde 2018 das anonyme Surfen im Internet und das Bewerben von VPNs unter Strafe gestellt. Erst 2019 wurden die zehn größten VPNs vor die Wahl gestellt, die russische Blockliste in ihre Systeme zu integrieren oder gebannt zu werden. Einzig der russische IT-Sicherheitsdienstleister Kaspersky-Lab entschied, der staatlichen Anweisung Folge zu leisten.

2020 wurde ein neuer Gesetzesvorschlag diskutiert, welcher jedoch noch nicht verabschiedet wurde (Stand: Januar 2021). Der neue Vorschlag sieht ein Verbot der Verschlüsselungsstandards TLS 1.3, ESNI, DNS over HTTPS (DoH) und DNS over TLS (DoT) vor.[19] Der Grund für diesen Gesetzesvorstoß ist, dass Roskomnadzors aktuelles Filtersystem durch diese neuen Verschlüsselungstechniken ausgehebelt wird, da das Filtern auf Klartext URLs basiert. Schon das HTTPS Protokoll sorgt für Probleme: Das Filtern funktioniert über eine Notlösung, bei der ‚Leaks' des Internetverkehrs, wie etwa die Server Name Identification (SNI) Felder und DNS Queries, ausgelesen werden. Anhand der Probleme, welche durch die neuen Verschlüsselungstechniken für Roskomnadzor erzeugt würden, ist davon auszugehen, dass dieser Gesetzesvorschlag angenommen und implementiert wird.

2.4. Kontext

Um die Regulation des russischen Internets zu verstehen, ist es hilfreich, diese in den Kontext von politischen und geopolitischen Entwicklungen zu setzen. Die Verschärfung der Regulation des Internets mit der Einführung der Blocklist 2012 lässt sich als eine Konsequenz der politischen Unruhen 2011 sehen, die maßgeblich durch Austausch und Kommunikation in sozialen Medien organisiert wurden. Es wird vermutet, dass auch der politische Protest auf dem Maidan und der Umsturz der russisch-gestützten Führung in der Ukraine einen weiteren Grund für die Ausweitung der Befugnisse für SORM neben den offiziell genannten Sicherheitsvorkehrungen für die Winterolympiade 2014 in Sochi darstellt. Auch andere Gesetze folgen auf internationale Ereignisse: das Gesetz von 2015, welches die lokale Speicherung von persönlichen Daten vorschreibt, folgt auf Edward Snowdens Enthüllungen über die weitreichenden Überwachungsprogramme der USA.[20] Die Antiterrorgesetze von 2016 sind im internationalen Kontext der Erfolge des IS im Nahen Osten sowie der Terroranschläge in Paris zu sehen. Das Gesetz über

[18] Michael Kan, "Russia Demands 10 Major VPNs Censor Content or Face Ban," PCMag, March 28, 2019, https://uk.pcmag.com/old-news/120304/russia-demands-10-major-vpns-censor-content-or-face-ban.

[19] Catalin Cimpanu, "Russia wants to ban the use of secure protocols such as TLS1.3, DoH, DoT, ESNI," ZDNet, September 22, 2020, https://www.zdnet.com/article/russia-wants-to-ban-the-use-of-secure-protocols-such-as-tls-1-3-doh-dot-esni/.

[20] Andrei Soldatov and Irina Borogan, The Red Web: The Kremlin's War on the Internet (New York: PublicAffairs, 2017), 210.

ein Autonomes Internet von 2019, welches die Isolation des russischen Internets ermöglicht, wird mitunter als Reaktion auf besonders schädliche Cyberkampagnen wie WannaCry und NotPetya gewertet.[21]

Russland reagiert allerdings nicht nur auf äußere Einflüsse, sondern gestaltet seine Internetpolitik auch proaktiv. Ein Aspekt der russischen Internetpolitik ist die 2016 unterzeichnete Doktrin für Informationssicherheit der Russischen Föderation, welche Russlands Ansatz bezüglich der "Information Sphere", wie Russland den Cyberraum bezeichnet, und der strategischen nationalen Prioritäten aufzeigt. Ein wichtiger Fokus ist die Konzentration auf russische Informationstechnologien, um die Abhängigkeit von externen Akteuren zu verringern. Diese Entwicklung könnte auf lange Sicht negative Konsequenzen für die Sicherheit der Technologien bedeuten, wenn der russische Markt losgelöst von der internationalen Sicherheitsforschung und durch eine kleinere Nutzerbasis geringere Kontrollen und Standards entwickelt. Allgemein lässt dich die russische Doktrin für Informationssicherheit als „Russia First in Cyberspace" zusammenfassen. Dieser nationale Fokus auf staatliche Souveränität im Internet zieht sich als roter Faden durch Russlands Internetpolitik, welche auf die Anpassung globaler Internetnormen abzielt. Ein Beispiel hierfür ist die Resolution der Generalversammlung der Vereinten Nationen zur Bekämpfung der Nutzung von Informations- und Kommunikationstechnologien zu kriminellen Zwecken.[22] Menschenrechtsaktivist*innen werfen der Resolution vor, dass sie zu unspezifisch für das Erreichen der erklärten Ziele sei, dabei aber eine weitere Kriminalisierung von normalen Onlinetätigkeiten in Autoritären Regime zulasse.[23] Ein wichtiger Partner in den Bestrebungen, Souveränität zu einer grundlegenden Norm im Informationsraum zu machen, ist die Volksrepublik China, mit der Russland 2015 ein bilaterales Abkommen zur Kooperation im Bereich der internationalen Informationssicherheit unterzeichnete.[24]

3. Technische Umsetzung: Tooling und Technologie

Die Russische Föderation setzt auf verschiedene Technologien, um die Gesetze zu implementieren und den Internetverkehr zu beobachten und zu überprüfen. Ein zentraler Baustein ist SORM, welches seit 2014 aus drei Teilen besteht und eine umfassende Überwachung zulässt und in Referenz auf vergleichbare US-Amerikanische Programme als „PRISM on Steroids" bezeichnet wurde.[25] SORM-1 ist ein Relikt der Sowjetunion

[21] Max Seddon and Henry Foy, "Russian technology: can the Kremlin control the internet?" Financial Times, June 5, 2019, https://www.ft.com/content/93be9242-85e0-11e9-a028-86cea8523dc2.

[22] „Resolution der Generalversammlung, verabschiedet am 27. Dezember 2019 [74/247]" Vereine Nationen, December 27, 2019, https://digitallibrary.un.org/record/3847855/files/A_RES_74_247-DE.pdf.

[23] "Open letter to UN General Assembly: Proposed international convention on cybercrime poses a threat to human rights online," Association for Progressive Communications, November 2019, https://www.apc.org/en/pubs/open-letter-un-general-assembly-proposed-international-convention-cybercrime-poses-threat-human.

[24] James Andre Lewis, "Sino-Russian Cybersecurity Agreement 2015," Center for Strategic & International Studies, May 11, 2015, http://cyber-peace.org/wp-content/uploads/2013/05/RUS-CHN_CyberSecurityAgreement201504_InofficialTranslation.pdf.

[25] Shaun Walker, "Russia to monitor 'all communications' at Winter Olympics in Sochi," The Guardian, October 6, 2013, https://www.theguardian.com/world/2013/oct/06/russia-monitor-communications-sochi-winter-olympics

und wird zur Telefonüberwachung genutzt; SORM-2 ist für das Monitoring von Internetverkehr zuständig, während SORM-3 ein Datenbanksystem darstellt, aus dessen Informationen jederzeit Nutzerprofile erstellt werden können. Einem Erlass des Ministeriums für Digitale Entwicklung, Kommunikation und Massenmedien zufolge erhebt SORM folgende Daten: IP Adressen (IPv4 und IPv6), Benutzer IDs in Systemen von Telekommunikationsbetreibern, E-Mail Adressen auf SMTP, POP3 und IMAP4 Protokollen sowie Webmail - verschlüsselte Kommunikation hierbei ausgenommen, fernerdie International Mobile Subscriber Identity (IMSI), die International Mobile Equipment Identity (IMEI), MAC Adresse von Nutzerendgeräten, Identifikatoren von Nachrichten wie ICQ und die Mobile Identification Number der Mobile Subscriber Station (MIN).[26] Durch Leaks ist zudem bekannt, dass SORM auch die Handynummer, Login Details, Netzwerkadressen, Messenger Nummern, GPS Koordinaten, die Home Location Register (HLR) Datenbank sowie Betreffzeilen von Emails sammelt.[27]

Spätestens seit dem neuen Gesetz über ein Autonomes Internet ist Deep Packet Inspection (DPI) eine Technologie, die in SORM integriert wird und staatlichen Autoritäten weitere Fähigkeiten zur Verfügung stellt, allerdings zeigen Tests vom Dezember 2019, dass die Integration nur teilweise erfolgreich ist: In manchen Fällen funktionierte das Internet weiterhin ohne Probleme, in anderen Fällen führte es zu langsamerer Internetgeschwindigkeit, schwächeren Signalen oder gar lokalen Ausfällen.[28] Ein Beispiel für die Nutzung von DPI in Kombination mit SORM stellt Belarus dar. Als Reaktion auf die Proteste gegen die Wahl Lukashenkos blockte das Regime über 10.000 Schlüsselworte, um den Internetverkehr zu filtern und Zugang zu Informationen blocken, was neben Internetausfällen auch zum Blocken von internationalen Unternehmensseiten wie beispielsweise Walmart, Amazon oder Disney führte.[29] Belarus ist nicht das einzige post-sowjetische Land, welches auf SORM setzt: auch die Ukraine und Kyrgyztan haben SORM Technologien von Russland erworben.[30]

Dem Gesetz über ein Autonomes Internet nach dürfen ISPs den Internetverkehr in Gefahrensituationen nur durch staatlich genehmigte Internet Exchange Points (IXPs) routen. Darüber hinaus sieht das neue Gesetz vor, dass transnationaler Internetverkehr durch Kontrollpunkte geroutet werden muss, die als Kill-Switches fungieren können.[31]

---

[26] "Order No. 83 On the approval of the Rules for the use of equipment for switching systems, including software that ensures the implementation of established actions when conducting operational-search measures. Part III. Rules for the use of equipment for switching and routing information packets of data transmission networks, including software that ensures the implementation of established actions during operational-search activities," Ministry of Communications and Mass Communications of the Russian Federation, April 16, 2014, https://digital.gov.ru/ru/documents/4249/.

[27] Petr Lokhov, "Suspicious sniffer: Programmer discovers thousands of phone numbers, addresses, and geolocations apparently leaked by Russia's 'SORM' surveillance tech," Meduza, August 27, 2019, https://meduza.io/en/feature/2019/08/27/suspicious-sniffers.

[28] "Russia: Growing Internet Isolation, Control, Censorship," Human Rights Watch, June 18, 2020, https://www.hrw.org/news/2020/06/18/russia-growing-internet-isolation-control-censorship.

[29] David Gilbert, "Belarus Cut Off the Internet and Tried to Make It Look Like an Accident," Vice, August 11, 2020, https://www.vice.com/en_us/article/z3e8v3/belarus-cut-off-the-internet-and-tried-to-make-it-look-like-an-accident.

[30] Andrei Soldatov and Irina Borogan, "In Ex-Soviet States, Russian Spy Tech Still Watches You," Wired, December 21, 2012, https://www.wired.com/2012/12/russias-hand/.

[31] Alena Epifanova, "Deciphering Russia's 'Sovereign Internet Law': Tightening Control and Accelerating the Splinternet," DGAP Analysis, January 16, 2020, https://dgap.org/en/research/publications/deciphering-russias-sovereign-internet-law.

Eine Konsequenz der erfolgreichen, technischen Implementierung des Gesetzes ist, dass Russland ein nationales Intranet einrichten könnte, welches sich theoretisch vom globalen Internet ablösen und nur Internetverbindungen innerhalb Russlands zulassen könnte. Dies wäre ein grundlegender Eingriff in die existierende Struktur des russischen Internets, die auf einer Vielzahl von ISPs basiert. Eine Zentralisierung auf wenige Routing Punkte könnte gravierende Auswirkungen auf die Funktionsfähigkeit des Internets haben und für instabile Verbindungen sorgen.[32]

Das Gesetz von 2019 birgt noch eine weitere technische Neuerung: ein nationales Domain Name System (DNS). Das DNS ist essentieller Grundstein des Internets, da es die einmalige Zuordnung von IP-Adressen und deren Auflösung verwaltet und dadurch das Internet zugänglich und funktionsfähig macht. Es ist unklar, welche Auswirkungen ein nationales DNS auf die lokale Funktionsweise, das globale Internet und die Interaktion dieser Komponenten hätte. Grundsätzlich lassen sich zwei Optionen festhalten: der Versuch ein alternatives Root-System zu schaffen und dieses an das universelle DNS anzuhängen oder die Entwicklung eines parallelen DNS, welche höchstwahrscheinlich nicht mit dem Rest der Welt vereinbar wäre.[33] Aber auch bisherige Experimente mit alternativen Root Systemen, die an das universelle DNS angeglichen wurden, um eine Doppelung von IP-Adressen auszuschließen, garantieren keinen Erfolg. Es ist unklar, wie Roskomnadzor in der Entwicklung eines nationalen DNS vorgehen wird. Schon im Dezember 2019 wurden Tests durchgeführt, um zu evaluieren, ob die russische Internetinfrastruktur ohne das DNS funktioniert. Die Ergebnisse dieser Tests sind der Öffentlichkeit jedoch nicht zugänglich.[34]

4. Relevante Risiken für Unternehmen

Unternehmen sehen sich einer Vielzahl von Risiken ausgesetzt, welche sich für in Russland ansässige Unternehmen und Unternehmen mit russischen Geschäftspartnern unterscheiden. In Russland ansässige Unternehmen unterstehen der russischen Gerichtsbarkeit und unterliegen somit russischen Gesetzen, weswegen die Risiken eine höhere Intensität für sie haben. Die Risiken lassen sich grundsätzlich in drei Kategorien einteilen: Betriebsstörungen, Wirtschaftsspionage und gestiegene Betriebskosten.

Betriebsstörungen können in Form von Internetstörungen oder gezielter Zensur auftreten, auch wenn letzteres bisher im Zusammenhang mit Nicht-Regierungsorganisationen (NGOs) und nicht mit unpolitischen Unternehmen beobachtet wurde. Während die Webseiten von NGOs wie dem German Marshall Fund, dem Atlantic Council oder der Open Society Foundation von Roskomnadzor geblockt werden, ist das Blockieren von Unternehmen bisher nur in Fällen von Nichteinhaltung russischer Gesetze erfolgt. Als

[32] Ibid.

[33] Samantha Bradshaw and Laura DeNardis, "The Politicisation of the Internet's Domain Name System – Implications for Internet Security, Universality, and Freedom," European Consortium for Political Research, August 2015, https://ecpr.eu/Events/PaperDetails.aspx?PaperID=25211&EventID=94.

[34] Catalin Cimpanu, „Russia successfully disconnected from the internet," ZDNet, December 23, 2019, https://www.zdnet.com/article/russia-successfully-disconnected-from-the-internet/.

Beispiele gelten LinkedIn, welches gesperrt wurde, da die Daten nicht auf lokalen Servern gespeichert werden, aber auch verschiedene VPN-Anbieter, welche die Bedingungen von Roskomnadzor nicht akzeptieren wollten.[35] Das größte Risiko für Betriebsstörungen bei internationalen Unternehmen sind Internetausfälle. Der Bann von 10.000 Stichwörtern in Belarus blockte nicht nur die Bürger*innen in ihrer politischen Organisation, sondern auch die Webseiten von Walmart, Amazon und Disney. Auch in Russland gab es etliche ungewollte Ausfälle in Folge von Roskomnadzors Versuch, die Messenger App Telegram 2018 zu blocken.[36] Unternehmen mit russischen Niederlassungen oder Partnern sollten demnach Betriebsstörungen in ihre Risikokalkulation aufnehmen.

Ein weiteres Risiko stellt Wirtschaftsspionage dar, welche durch die technischen Möglichkeiten von SORM und DPI möglich ist. Es ist kein Fall öffentlich bekannt, in dem russische Behörden diese technischen Kapazitäten gegen internationale Firmen aktiv im russischen Markt eingesetzt hätte, allerdings ist es rechtlich weder den ISPs noch den Privatpersonen oder Organisationen gestattet, Einblick in die Überwachung von SORM zu erhalten. Bekannt ist jedoch die Überwachung von Oppositionellen oder Menschenrechtsaktivist*innen zum Beispiel durch die Veröffentlichungen von Kompromat, also von privatem, kompromittierendem Material, das durch rechtmäßige oder unrechtmäßige Überwachung gesammelt wurde.[37] Theoretisch ist es möglich, dieses Überwachungssystem auch auf wirtschaftliche Unternehmen zu fokussieren, insbesondere dann, wenn deren Interessen konträr zu denen des Staates laufen. Jedoch ist eine abschließende Aussage zu Wirtschaftsspionage durch SORM und DPI nicht möglich.

Ein letztes zu beachtendes Risiko sind gesteigerte Betriebskosten durch die Präsenz in Russland sowie Handelsbeziehungen mit russischen Partnern. Den Unternehmen ist anzuraten, Rechtsexpert*innen für die russische Internetgesetzeslage anzustellen. Zum einen können diese schnell auf neue Entwicklungen reagieren und Unternehmen vor Verstößen gegen die Gesetze schützen. Beispiele für die Notwendigkeit von Rechtsberatung sind zum einen das Gesetz über die lokale Speicherung persönlicher Daten aber auch das Gesetz, welches allgemein als Verbot von VPNs verstanden wird. Nichteinhaltung der Gesetze kann zu Strafen führen, wie die wiederholten Bußgelder für Google seit 2018 zeigen. Zwar beträgt die Summe insgesamt nur rund 5,7 Millionen Rubel (ca. 76.900 Euro) und ist im Vergleich zu Googles Umsatz gering, dennoch sollten solche Geldbußen in der Risikokalkulation berücksichtigt werden. Auch das Gesetz über ein Autonomes Internet stellt eine Herausforderung für Unternehmen dar: Falls es Russland gelänge, ein nationales und autonomes Internet zu kreieren, dann würde das die bisherige IT-Infrastruktur und transnationale Handelsbeziehungen in Frage stellen. Es ist unklar, inwieweit Russland das autonome Internet nutzen möchte, wenn es über die tech-

[35] John Faulds, "Which websites and services are banned in Russia?" Techradar, August 24, 2020, https://www.techradar.com/vpn/which-websites-and-services-are-banned-in-russia.

[36] Matt Burgess, „This is why Russia's attempts to block Telegram have failed," Wired, April 28, 2018, https://www.wired.co.uk/article/telegram-in-russia-blocked-web-app-ban-facebook-twitter-google.

[37] Roman Dobrokhotov, "Under surveillance in Russia," Al Jazeera, November 8, 2016, https://www.aljazeera.com/indepth/opinion/2016/11/surveillance-russia-161107133103258.html.

nischen Möglichkeiten verfügt. Aus diesem Grund sollten Unternehmen die Entwicklungen in Russland sorgfältig verfolgen, die rechtlichen und technischen Unsicherheiten in der Risikokalkulation berücksichtigen und gegebenenfalls ihre IT-Infrastruktur und Strategie auf ein isoliertes russisches Internet anpassen.

5. Szenarien

Aus dem Gesetz über ein Autonomes Internet von 2019 ergeben sich mehrere Szenarien für die Zukunft des russischen Internets und seiner Eingliederung im globalen Kontext.

5.1. Status Quo mit Emergency Exit

Der Status Quo mit Emergency Exit sieht eine Implementierung der technischen Anforderungen des Gesetzes vor, die allerdings nur in Notfällen genutzt wird. Dies würde bedeuten, dass Russland an das globale Internet angegliedert bleibt und die Kontrolle über sogenannte Kill-Switches sowie ein nationales DNS nicht aktiv genutzt wird. Nur in besonderen Notfällen, welche die Integrität des Netzwerkes, die Stabilität des Netzwerkes und die Funktionssicherheit des Netzwerkes betreffen, würde dann der Kill-Switch aktiviert und das Internet isoliert werden.[38] Dieses Szenario birgt die wenigsten Risiken für Unternehmen, da im Normalfall alles wie gewöhnlich funktionierte und es nur in Ausnahmesituationen zu temporären Ausfällen käme.

5.2. RuNet

Das zweite Szenario sieht ein komplett abgeschottetes russisches Internet, kurz RuNet, vor. Russland würde auf eigene Infrastruktur setzen und ein nationales DNS nutzen, welches losgekoppelt vom universellen DNS funktionieren würde. In einem solchen Szenario könnten russische Mitarbeiter*innen nicht mehr auf globale Netzwerke zugreifen und IT-Expert*innen hätten keinen Zugriff auf die russischen Firmennetze. In einem solchen Szenario ist den Unternehmen anzuraten, eine Kosten-Nutzen-Kalkulation aufzustellen und die Risiken durch vergleichsweise schlechter gesicherte Systeme in Russland oder doppelte Betriebskosten für die russische Infrastruktur zu berücksichtigen. Ein weiteres Risiko in diesem Szenario ist die Notwendigkeit von transnationaler Kommunikation, die durch ein isoliertes RuNet nicht möglich wäre. Dieses Szenario würde zweifelsohne die größte Unsicherheit für Unternehmen erzeugen. Allerdings gibt es zurzeit keinerlei Hinweise, dass Russland langfristig ein isoliertes russisches Internet plant.

5.3. Russische ‚Great Firewall' mit Emergency Exit

Das letzte Szenario sieht eine Verschärfung der technischen Überwachungskapazitäten vor sowie die Implementierung der Kill-Switches, um Russland in Notfällen vom globalen Internet isolieren zu können. Russland würde im Normalfall mit dem Internet verbunden bleiben, allerdings striktere Regeln und Zensurmechanismen vergleichbar mit China einführen. Die beiden Länder teilen eine ähnliche Ansicht über Souveränität im Cyberspace und kooperieren nicht nur bei internationalen Resolutionen, sondern auch bei 5G. Demzufolge scheint es naheliegend, dass Russland sich auch in seiner nationalen

[38] Alena Epifanova, "Deciphering Russia's 'Sovereign Internet Law': Tightening Control and Accelerating the Splinternet," DGAP Analysis, January 16, 2020, https://dgap.org/en/research/publications/deciphering-russias-sovereign-internet-law.

Herangehensweise ans Internet dem chinesischen System angleicht. Dies würde einen Ausbau der Filter- und Blockiermöglichkeiten bedeuten, dem sich Unternehmen beugen müssen, in dem der Betrieb aber weiterhin möglich ist. Auch in diesem Szenario ist spezialisierter Rechtsbeistand und lokale IT-Expertise empfehlenswert.

6. Fazit

Die russische Internetrechtslage ist in der Entwicklung, und das Resultat ist noch nicht sicher. Neue Gesetze aus dem letzten Jahrzehnt ermöglichen jedoch einige Erkenntnisse und Empfehlungen.

Russland hat schrittweise zunehmend repressivere Maßnahmen in seine Gesetzgebung und seine Sicherheitssysteme implementiert. Seit 2012 hat der Staat seinen Druck auf die Privatsphäre und Unabhängigkeit von ISPs und Telekommunikationsunternehmen sowie Einzelpersonen und privaten Organisationen kontinuierlich erhöht. Während die meisten Maßnahmen auf politischen Protest abzielen, können dieselben rechtlichen und technischen Rahmenbedingungen auch gegen wirtschaftliche Rivalen angewendet werden. Technologien wie SORM und DPI sowie eine Rechtsgrundlage für intransparentes, staatlich angeordnetes Abhören bieten Strafverfolgungsbehörden eine breite Palette von Kontrollmaßnahmen, die auch Auswirkungen auf Unternehmen haben können.

Ein wichtiger Schritt ist das Gesetz über ein Autonomes Internet von 2019, welches die Möglichkeit eines vom globalen Internet isolierten russischen Internets vorsieht. Das Gesetz schreibt vor, dass der Internetverkehr über von der Regierung genehmigte Internet Exchange Points (IXPs) umgeleitet wird, staatlich kontrollierte Kill-Switches implementiert werden sowie ein nationales Domain Name Systems (DNS) eingerichtet wird. Dies würde Unternehmen mit russischen Niederlassungen oder russischen Partnern vor neue Herausforderungen stellen. Da es keinen Präzedenzfall für ein nationales Internet gibt, sind die vollständigen Konsequenzen eines „souveränen Internets“ unklar, sodass von Kompatibilitätsproblemen auszugehen ist.

Insgesamt ist der Fall des russischen Internets einzigartig und unterscheidet sich vom Ansatz der Internetregulierung in den meisten anderen technologisch fortgeschrittenen Ländern. Es ist wahrscheinlich, dass Russland die Kontrolle über das Internet verschärfen und seinen Weg zu einer Lösung fortsetzen wird, die der chinesischen Great Firewall ähnelt. Während es unwahrscheinlich ist, dass sich Russland auf unbestimmte Zeit vom globalen Internet abkoppelt, muss die Möglichkeit einer vorübergehenden Trennung in Zeiten politischer Unruhen, Naturkatastrophen oder noch nicht definierter „ausländischer Bedrohungen“ im russischen Internet als ernsthafte Möglichkeit angesehen werden. Unternehmen sollten sich dieser Möglichkeit und ihrer Folgen sowie der allgemeinen Auswirkungen und Folgen der staatlichen Internetregulierung in Russland für ihre Unternehmen und Mitarbeiter*innen gleichermaßen bewusst und darauf vorbereitet sein.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Virtuelle Hauptversammlungen: Ein sicherer Ersatz für Präsenzveranstaltungen?

Prof. Dr.-Ing. Andreas Mayer[1]

Kurzfassung:
Die virtuelle Hauptversammlung (HV) ist seit März 2020 rechtlich der Präsenz-HV gleichgestellt. Sie hat sich als gesellschaftsrechtliches Kriseninstrument, während der COVID-19-Pandemie, schnell und in voller Breite etabliert. In diesem Beitrag wurden 623 virtuelle HVs empirisch erfasst und die Sicherheit der zugrundeliegenden HV-Portale systematisch auf bekannte Schwachstellen und den Einsatz von bewährten Security Best Practices hin untersucht. Bei knapp 72 % der virtuellen HVs wurden kritische Schwachstellen gefunden, welche potenziell von Angreifern ohne Spezialkenntnisse und mit geringen Ressourcen ausgenützt werden konnten. Betroffen waren u. a. virtuelle HVs großer deutscher Aktiengesellschaften aus bekannten Börsensegmenten, wie z. B. DAX und MDAX.
Nach eigenen Recherchen ist dies die erste Veröffentlichung, welche das Sicherheitsniveau von virtuellen HVs systematisch und breit angelegt untersucht. Im Ergebnis konnten die Schwachstellen behoben und so die Sicherheit von virtuellen HVs maßgeblich verbessert werden.

Stichworte: Aktionärsportal, Bedrohungsanalyse, Hauptversammlung, HV-Portal, Schutzziele, Virtuelle Versammlungen

1. Einleitung

Am 28. April 2020 veranstaltete die Bayer AG mit über 5.000 teilnehmenden Aktionären für rund 1 Million Euro die erste rein virtuelle Hauptversammlung (HV) in Deutschland[2]. Die Grundlage für die Durchführung von virtuellen HVs ist das am 27. März 2020 vom Bundesrat im Eilverfahren verabschiedete Gesetz zur Abmilderung der Folgen der COVID-19-Pandemie im Zivil-, Insolvenz- und Strafverfahrensrecht [1]. Bisher war im Aktienrecht verankert, dass die jährlich verpflichtende HV von Aktiengesellschaften zwingend als physische Präsenzveranstaltung stattfinden muss. Die Ausnahmeregelung war zunächst bis zum 31. Dezember 2020 befristet, wurde aber unlängst bis Ende 2021 verlängert [2]. Obwohl die virtuelle HV laut dem verabschiedeten Gesetz ausdrücklich nicht verpflichtend ist, hat sich mit 78 % die überwiegende Mehrheit der deutschen Aktiengesellschaften in der HV-Saison 2020 für eine präsenzlose Durchführung entschieden.

Die HV als Organ einer Aktiengesellschaft (AG) dient zur Information aller Aktionäre inkl. Frage- und Auskunftsrecht und zum Beschluss über grundsätzliche Entscheidungen, wie z. B. Entlastung von Vorstand und Aufsichtsrat, Gewinnverwendung, Wahl des Abschlussprüfers, Genehmigung dringend notwendiger Kapitalmaßnahmen oder die Abstimmung über den zwangsweisen Ausschluss der Aktionäre in einem Squeeze-Out-Verfahren. In einer virtuellen HV werden somit kritische Unternehmensentscheidungen mit teils weitreichenden Folgen getroffen und zugleich sehr sensible personenbezogene

[1] Hochschule Heilbronn, E-Mail: andreas.mayer@hs-heilbronn.de

[2] https://www.juve.de/nachrichten/deals/2020/04/aktionaerstreffen-linklaters-mandantin-bayer-spart-mit-online-hv-25-millionen-euro

Daten von Aktionären verarbeitet. Für die Durchführung von virtuellen HVs bedienen sich die Aktiengesellschaften deshalb in aller Regel eines spezialisierten HV-Dienstleisters, welcher neben organisatorischer und rechtlicher Unterstützung auch ein HV-Portal zur praktischen Durchführung der virtuellen HV im Internet zur Verfügung stellt.

In diesem Artikel werden zwei Forschungsfragen adressiert:

1. Welche HV-Dienstleister zur Abwicklung von virtuellen HVs gibt es und wie ist deren Marktanteil?
2. Wie ist das Sicherheitsniveau der HV-Portale dieser HV-Dienstleister?

Zur Beantwortung dieser beiden Forschungsfragen liefert dieser Artikel den folgenden Beitrag:

In einer Bedrohungsanalyse werden zunächst exemplarisch kritische Bedrohungen im Kontext von virtuellen HVs erarbeitet und daraus abgeleitet Schwachstellen und denkbare Angriffe auf HV-Portale vorgestellt (Kapitel 2).

Auf Grundlage der Bedrohungsanalyse wird eine Methodik zur systematischen Untersuchung der Sicherheit von HV-Portalen und deren Angriffsfläche skizziert (Kapitel 3). Dabei werden Informationen über die verwendeten Technologien gewonnen, eingesetzte Security Best Practices analysiert und die HV-Portale im Rahmen von Blackbox-Penetrationstests auf typische und weit verbreitete Schwachstellen aus der OWASP Top 10-Liste [3] untersucht[3].

In einer empirischen Studie wurden 623 virtuelle HVs von deutschen Aktiengesellschaften im Zeitraum vom 28. April 2020 – 31. Dezember 2020 analysiert. Hierbei konnten insgesamt 15 HV-Dienstleister mit acht unterschiedlichen HV-Portalen identifiziert werden. Während der systematischen Sicherheitsanalyse nahm der Autor an 46 virtuellen HVs mit 71 unterschiedlichen Accounts teil.

Im Ergebnis wiesen 71,6 % der untersuchten virtuellen HVs kritische Schwachstellen auf, welche u. a. das unbemerkte Ändern der Stimmabgabe von Aktionären, die vollständige Übernahme des Aktionärs-Accounts durch den Angreifer, das gezielte Verhindern der Durchführung von virtuellen HVs oder das Auslesen der personenbezogenen Daten von Aktionären ermöglichten. Letztere Sicherheitslücke erlaubte es, mit dem Einsatz von ca. 20 €, die personenbezogenen Daten (Name, Adresse, Geburtsdatum, …) *aller* Aktionäre, inkl. Abstimmungsverhalten und deren Anteilsbesitz, von *allen* durch den Dienstleister durchgeführten HVs auszulesen (mehr als 100 virtuelle HVs waren hiervon betroffen). Von den gefundenen Sicherheitslücken waren u. a. auch virtuelle HVs von großen DAX- und MDAX-Konzernen betroffen. Insgesamt konnten in sechs von acht HV-Portalen konkrete Schwachstellen gefunden werden. Zusammen mit den Ergebnissen der empirischen Studie und der erfassten potenziellen Angriffsfläche, werden diese Schwachstellen in Kapitel 4 dargestellt und anschließend diskutiert (Kapitel 5).

[3] Es wurden nur passive Analysen und nicht-invasive Tests mit legitimen HV-Accounts durchgeführt. Zu keinem Zeitpunkt wurden unberechtigt Daten Dritter eingesehen, verändert oder die Funktionalität der HV-Portale beeinträchtigt.

Im Rahmen von Responsible Disclosure-Verfahren, wurden die gefundenen Schwachstellen den betroffenen HV-Dienstleistern offengelegt, von diesen bestätigt und mit Unterstützung des Autors behoben.

In Kapitel 6 wird der Stand der Forschung dargestellt. Nach den Recherchen des Autors, ist dies der erste Beitrag, welcher das Sicherheitsniveau von virtuellen HVs systematisch und breit angelegt untersucht und maßgeblich verbessert hat.

2. Bedrohungsanalyse und Angreifermodell

Virtuelle HVs werden in einem HV-Portal abgehalten, welches als Webanwendung über das Internet zur Verfügung gestellt wird. Im Folgenden werden ausgehend von den Vermögenswerten (Assets), den bereitgestellten Funktionalitäten (Use-Cases) und den drei grundlegenden Schutzzielen der Informationssicherheit (Vertraulichkeit, Integrität und Verfügbarkeit) kritische Bedrohungen für HV-Portale beispielhaft dargestellt. Diese Bedrohungen können durch Schwachstellen entstehen, welche in der Anwendung bzw. der zugehörigen Infrastruktur vorhanden sind. Basierend auf den Bedrohungen und den Schwachstellen werden anschließend bekannte Angriffe skizziert, welche die Schutzziele kompromittieren können.

2.1. Assets und Funktionalitäten

In einem HV-Portal werden als wesentliche Assets die personenbezogenen Daten der für eine HV angemeldeten Aktionäre gespeichert. Diese umfassen zumindest Vorname, Name, Wohnort, Anzahl der zum Nachweisstichtag gehaltenen Aktien, Aktionärsnummer, Aktiengattung und Besitzart der Aktien (Eigen-/Fremdbesitz). Diese Daten werden auch bei einer physischen HV für das Teilnehmerverzeichnis benötigt. Regelmäßig werden jedoch weitere sensible Daten, wie z. B. vollständige Anschrift, E-Mail-Adresse, Telefonnummer, Geburtsdatum, Nationalität und die Bank, welche den Aktienbestand gemeldet hat, abgespeichert. In HV-Portalen wird zudem das Abstimmungsverhalten der einzelnen Aktionäre festgehalten.

Zur Durchführung von virtuellen HVs stellt ein HV-Portal typischerweise die folgenden grundlegenden Use-Cases bereit:

- Bild- und Tonübertragung der HV
- Stimmabgabe per elektronischer Briefwahl inkl. Änderung/Widerruf
- Vollmacht und Weisung an die Stimmrechtsvertreter der Gesellschaft
- Vollmachterteilung an einen Dritten
- Einreichung von Fragen an die Gesellschaft (im Vorfeld der HV)
- Erklärung von Widersprüchen zu Hauptversammlungsbeschlüssen
- Einsicht von Dokumenten (z. B. in das Teilnehmerverzeichnis)
- Login-/Logout

2.2. Bedrohungen, Schwachstellen und Angreifermodell

Basierend auf den vorhandenen Assets und den Use-Cases entstehen reale Bedrohungen für HV-Portale. Damit eine Bedrohung existiert, müssen eine oder mehrere Schwachstellen in der Software und/oder Infrastruktur der HV-Portale vorhanden sein, die in

einem Angriff ausgenützt werden können. Für die nachfolgend dargestellten Angriffe, besitzt der Angreifer die folgenden, leicht zu erfüllenden, Fähigkeiten:

1. Zugriff auf das HV-Portal: Die Internetadresse zum betreffenden HV-Portal wird in der HV-Einladung bzw. auf der Webseite der Gesellschaft veröffentlicht und ist für jeden einsehbar.
2. Account zum Login in das HV-Portal: Der Angreifer kann jederzeit eigene Zugangsdaten für ein von einer AG genutztes HV-Portal erhalten. Voraussetzung ist, dass er Aktien der Gesellschaft erwirbt und den Anteilsbesitz fristgerecht nachweist. Hierfür reicht es aus, dass der Angreifer eine einzige Aktie besitzt.
3. Opfer klickt auf einen Link: Weiterhin kann der Angreifer einen legitimen HV-Teilnehmer (das Opfer) dazu bringen, auf einen Link zu klicken (z. B. durch einen Beitrag in einem Diskussionsforum oder durch eine E-Mail). Das Opfer muss gleichzeitig im HV-Portal eingeloggt sein.

In Tabelle 1 werden als Ergebnis der durchgeführten Bedrohungsanalyse drei real existierende Bedrohungen aufgeführt, welche jeweils eines der drei grundlegenden Schutzziele kompromittiert. Zu jeder aufgeführten Bedrohung werden korrespondierende Schwachstellen-Klassen genannt, welche der OWASP Top 10-Liste[4] entnommen wurden. Im Speziellen werden die Schwachstellen-Klassen „A2:2017 Broken Authentication“, „A5:2017 Broken Access Control“ und „A8:2013 Cross-Site Request Forgery“ verwendet. Diese werden als Grundlage für die manuelle Blackbox-Sicherheitsprüfung der HV-Portale verwendet.

In der letzten Spalte von Tabelle 1 werden beispielhaft korrespondierende und allgemein bekannte Angriffe aufgeführt. Im Folgenden werden diese kurz erläutert:

- Identitätsdiebstahl durch einen Brute Force-Angriff: Zur Nutzung eines HV-Portals muss sich der Teilnehmer authentifizieren. Bei allen untersuchten HV-Portalen geschieht dies durch Eingabe von Benutzername und Passwort. Werden hierbei schwache Passwörter verwendet, können diese durch einen automatisierten Brute Force-Angriff erraten werden.
- Identitätsdiebstahl durch einen Session Fixiation-Angriff [4]: Bei einem Session Fixiation-Angriff benutzt das Opfer eine vom Angreifer vorgegebene Session ID. Dadurch ist die Session ID auch dem Angreifer bekannt und er kann infolgedessen die Identität des Opfers im HV-Portal übernehmen.
- Fehlerhafte Zugriffskontrolle (Broken Access Control): Authentifizierte Benutzer können in einem HV-Portal evtl. auf sensible Daten (z. B. die persönlichen Daten anderer Aktionäre) oder Funktionen zugreifen, für die sie nicht berechtigt sind. Dies ist möglich, wenn die Zugriffskontrolle fehlerhaft implementiert ist und die Berechtigungen nicht oder nur unzureichend geprüft werden.

[4] Die Open Web Application Security Project (OWASP) Foundation veröffentlicht seit 2004 regelmäßig eine anerkannte Top 10-Liste der kritischsten und häufigsten Sicherheitsprobleme von Webanwendungen.

- Cross-Site Request Forgery (CSRF) [5]: Bei einem CSRF-Angriff führt das Opfer, meist unbewusst, vom Angreifer vorgegebene Aktionen aus. Dies kann bei einem HV-Portal z. B. das Ändern des Abstimmungsverhaltens sein.
- Sperren von Accounts durch einen Brute Force-Angriff: Eine häufig eingesetzte Gegenmaßnahme, um das automatisierte Erraten von Passwörtern zu verhindern, ist das Sperren von Benutzeraccounts nach mehrmaliger Falscheingabe. Jedoch kann dadurch auch ein legitimer HV-Teilnehmer nicht mehr an der HV teilnehmen. Wenn der Angreifer den Aufbau und die Vergaberichtlinien der Benutzeraccounts kennt, kann er gezielt einzelne oder alle Teilnehmer von einer HV ausschließen.

Kompromittiertes Schutzziel	Bedrohung	OWASP Schwachstellen-Klasse	Mögliche Angriffe
Vertraulichkeit	Angreifer kann die persönlichen Daten und/ oder das Abstimmungsverhalten von Aktionären einsehen	A2:2017 Broken Authentication A5:2017 Broken Access Control	Identitätsdiebstahl durch Brute Force- oder Session Fixiation-Angriff, fehlerhafte Zugriffskontrolle
Integrität	Angreifer kann die persönlichen Daten und/ oder das Abstimmungsverhalten von Aktionären ändern	A2:2017 Broken Authentication A5:2017 Broken Access Control A8:2013 Cross-Site Request Forgery	Identitätsdiebstahl durch Brute Force- oder Session Fixiation-Angriff, fehlerhafte Zugriffskontrolle, Manipulation von Daten durch CSRF-Angriff
Verfügbarkeit	Angreifer kann Accounts von Aktionären (gezielt) sperren und so die Teilnahme an der HV verhindern	A2:2017 Broken Authentication	Accounts durch Brute Force-Angriff sperren

Tabelle 1: Die Ergebnisse der Bedrohungsanalyse

3. Methodik

Nachfolgend wird die für die systematische Sicherheitsanalyse verwendete Methodik, welche in Abbildung 1 dargestellt wird, beschrieben.

Als Datenquelle wurden die von der Bundesanzeiger Verlag GmbH[5] veröffentlichten Einberufungen von HVs herangezogen. Anhand dieser Bekanntmachungen wurde geprüft, ob es sich um eine virtuelle HV handelt und wann diese stattfindet. Im Falle einer virtuellen HV wurden die folgenden Analyseschritte durchgeführt:

[5] https://www.bundesanzeiger.de/

1. HV-Dienstleister identifizieren: Mithilfe der im Bundesanzeiger bekanntgemachten HV-Einladung und den auf der Webseite der AG veröffentlichten Informationen wurde der HV-Dienstleister identifiziert und der Link (*HV-Portal URL*) zu dem HV-Portal extrahiert.
2. Informationsgewinnung: Um die potenzielle Angriffsfläche abzuschätzen und einen Eindruck über die verwendeten Security Best Practices zu gewinnen, wurde das HV-Portal in einem ersten Schritt untersucht. Hierbei wurden die folgenden öffentlich verfügbaren Informationen gewonnen:

    a) Verwendete Technologien: Die freie Browser Extension Wappalyzer [6] und die manuelle Untersuchung der HTTP-Header, lieferten die Datenbasis für die eingesetzten Softwarebibliotheken und die verwendeten Infrastrukturkomponenten. Weiterhin wurde untersucht, ob veraltete Softwarebibliotheken mit bekannten Schwachstellen eingesetzt werden.

    b) Eingesetzte Security Header: Es existieren heutzutage eine Vielzahl von anerkannten und bewährten HTTP Security Header (z. B. HSTS[6]), welche benutzt werden können, um die Sicherheit von Webanwendungen zu erhöhen. Der Einsatz dieser Security Header wurde mit dem kostenfreien Dienst Security Headers [7] untersucht. Als Ergebnis wird ein Rating von A+ (sehr gut) bis F (ungenügend) vergeben.

    c) Benutzte TLS-Konfiguration: Ein weiterer wichtiger Baustein für die Sicherheit des HV-Portals ist der Einsatz des TLS-Protokolls, um die Datenübertragung per HTTPS hinsichtlich Vertraulichkeit, Integrität und Authentizität abzusichern. Der kostenfreie Dienst SSL Labs [8] wurde eingesetzt, um die Sicherheit der übertragenen Daten zu beurteilen. Als Ergebnis wird ein Rating von A+ (sehr gut) bis F (ungenügend) vergeben.

3. Manuelle Sicherheitsprüfung: Sofern es möglich war, wurden Aktien der Gesellschaft erworben und HV-Eintrittskarten angefordert, um das HV-Portal tiefgreifender auf Schwachstellen zu untersuchen. Aufbauend auf den Ergebnissen aus der Bedrohungsanalyse, wurden dabei die folgenden sicherheitsrelevanten Bereiche untersucht:
    a) Authentifizierung
    b) Sessionmanagement
    c) Zugriffskontrolle (Access Control)
    d) CSRF

Die Ergebnisse aus Schritt 2 (potenzielle Angriffsfläche) und Schritt 3 (evtl. vorhandene Sicherheitslücken), flossen anschließend in die empirische Studie über das Sicherheitsniveau von virtuellen HVs ein (siehe Kapitel 4).

---

[6] HTTP Strict Transport Security (HSTS) ist ein Sicherheitsmechanismus für HTTPS-Verbindungen, der u. a. vor Downgrade-Angriffen und Session Hijacking schützen soll.

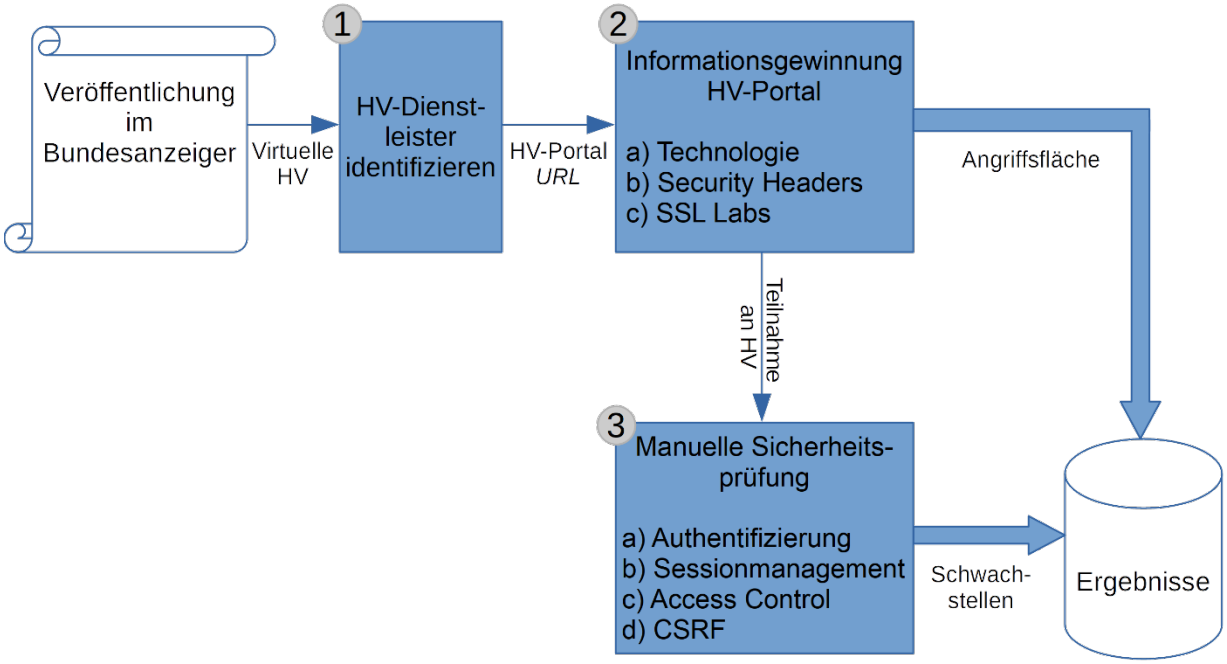


Abbildung 1: Die Methodik der Sicherheitsuntersuchung von virtuellen HVs

4. Ergebnisse

Für diese Studie wurden die virtuellen HVs der HV-Saison 2020 empirisch erfasst und nach der in Kapitel 3 vorgestellten Methodik systematisch untersucht. Hierfür nahm der Autor an 46 virtuellen HVs mit 71 unterschiedlichen Accounts teil. Der Auswertungszeitraum beginnt mit der ersten virtuellen HV (Bayer AG) am 28.04.2020 und endet am 31.12.2020. Insgesamt wurden in diesem Zeitraum 623 virtuelle HVs durchgeführt, wovon bei zwei HVs der HV-Dienstleister nicht identifiziert werden konnte. Von den verbleibenden 621 virtuellen HVs wurden 584 (94 %) von 15 unterschiedlichen HV-Dienstleistern durchgeführt. In 23 Fällen (3,7 %) wurde ein Videokonferenzsystem, wie z. B. Zoom, ohne ein spezielles HV-Portal eingesetzt. Hierbei fand die Abstimmung über die Tagesordnung typischerweise konventionell per Brief, Fax oder E-Mail statt. Bei 14 virtuellen HVs (2,3%) wurden eigene Softwarelösungen eingesetzt, welche für diesen Zweck von den AGs selbst oder durch Dritte entwickelt wurden. Die Verteilung der virtuellen HVs auf die einzelnen HV-Dienstleister kann Abbildung 2 entnommen werden. Die per Videokonferenzsystem durchgeführten HVs und die eigenentwickelten HV-Portallösungen werden im Folgenden nicht weiter betrachtet.

Bei der weiterführenden Untersuchung der einzelnen HV-Portale stellte es sich heraus, dass die HV-Dienstleister UBJ. GmbH, GFEI AG, ITTEB GmbH & Co. KG, BADER & HUBL GmbH, HV-Management GmbH, AAA HV Management GmbH, Art of Conference und HV AG alle dieselbe technische HV-Plattform, im Folgenden „BS HV-Portal“ genannt, einsetzen. Aufgrund dieser Tatsache reduziert sich die Anzahl der unterschiedlichen von HV-Dienstleistern eingesetzten HV-Portale auf acht Plattformen.

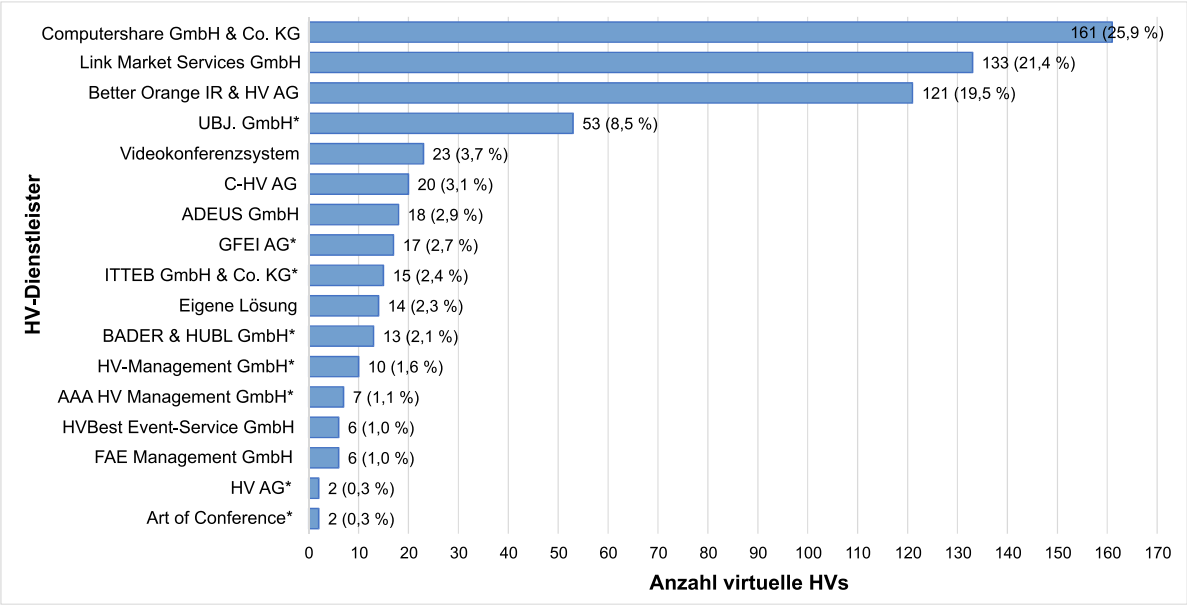


Abbildung 2: Anzahl der virtuellen HVs pro HV-Dienstleister inkl. Prozentangaben (*: HV-Dienstleister, welche dieselbe technische Plattform benutzen)

Die HV-Portale der drei größten HV-Dienstleister Computershare GmbH & Co. KG, Link Market Services GmbH und Better Orange IR & HV AG dominieren den Markt für virtuelle HVs mit insgesamt 66,8 % Marktanteil. Hinzu kommt das BS HV-Portal, welches von acht HV-Dienstleistern benutzt wird und in Summe auf gerundet 19,2 % Marktanteil kommt. Die restlichen 8 % teilen sich die HV-Portale der vier HV-Dienstleister C-HV AG (3,1 %), ADEUS Aktienregister-HV-Service GmbH (2,9 %), HVBest Event-Service GmbH (1,0 %) und die FAE Management GmbH (1,0 %) auf.

4.1. Informationsgewinnung

Bei den untersuchten HV-Portalen handelt es sich in sieben von acht Fällen um klassische Webanwendungen mit serverseitiger Anwendungslogik. Lediglich die Computershare GmbH & Co. KG setzt auf eine moderne Javascript-basierte Single Page Anwendung. Zur Realisierung werden die unterschiedlichsten Programmiersprachen (1x Javascript, 2x PHP, 2x ASP.NET und 3x Java), Frameworks (Angular JS, JavaServer Pages, Java Server Faces, Telerik Web UI und CodeIgniter) und Webserver (3x Microsoft IIS, 2x Nginx und 1x Apache) eingesetzt. In zwei Fällen konnte der Webserver der HV-Portalsoftware nicht identifiziert werden.

In fünf von acht HV-Portalen wurden insgesamt zehn veraltete Softwarebibliotheken mit bekannten Schwachstellen gefunden (Marktanteil: 52,4 %). Die verwundbaren Bibliotheken stammen aus den Jahren 2015-2019. Bei den Anbietern Link Market Services GmbH, BS HV-Portal und HVBest Event-Service GmbH wurden keine veralteten Softwarekomponenten mit bekannten Schwachstellen entdeckt.

Alle HV-Portale sind per HTTPS erreichbar und setzen das TLS-Protokoll ein. Vier von acht HV-Portale erreichten lediglich ein B-Rating (Marktanteil: 49,2 %). Die drei Anbieter Link Market Services GmbH, Better Orange IR & HV AG und FAE Management

GmbH (Marktanteil: 41,9 %) konnten ein A-Rating vorweisen. Lediglich das HV-Portal der ADEUS GmbH erreichte mit A+ das beste mögliche Rating (Marktanteil: 2,9 %).

Beim Security Header-Rating konnten die Computershare GmbH & Co. KG und die ADEUS GmbH mit einem befriedigenden C-Rating am besten abschneiden (Marktanteil: 28,8%). Die HVBest Event-Service GmbH erzielte ein D-Rating. Die restlichen HV-Portale, mit einem Marktanteil von 64,2 %, erreichten mit einem F-Rating die schlechteste mögliche Note.

In Tabelle 2 werden die beschriebenen Ergebnisse zusammenfassend dargestellt.

HV-Portal	Technologie	Verwundbare Software	SSL Labs Rating	Security Header Rating
Computershare GmbH & Co. KG	Microsoft IIS und Angular JS	Ja (Webserver und Javascript Bibliotheken)	B	C
Link Market Services GmbH	Webserver unbekannt, PHP	Nein	A	F
Better Orange IR & HV AG	Nginx und Java Server Faces	Ja (Webserver)	A	F
BS HV-Portal	Microsoft IIS, ASP.NET mit Telerik Web UI	Nein	B	F
C-HV AG	Microsoft IIS, ASP.NET mit Telerik Web UI	Ja (Javascript Bibliothek)	B	F
ADEUS GmbH	Webserver unbekannt, Java	Ja (Javascript Bibliothek)	A+	C
HVBest Event-Service GmbH	Nginx und JavaServer Pages	Nein	B	D
FAE Management GmbH	Apache, PHP und CodeIgniter	Ja (Javascript Bibliotheken, CodeIgniter)	A	F

Tabelle 2: Die Ergebnisse der Informationsgewinnung aus Schritt 2 der Methodik (Legende Rating: A=sehr gut, F=ungenügend)

4.2. Manuelle Sicherheitsprüfung

Im Rahmen der manuellen Blackbox-Sicherheitsprüfung wurde zunächst die Passwort-basierte Authentifizierung untersucht. Die Ergebnisse werden in Tabelle 3 dargestellt.

Alle HV-Dienstleister verwenden als Benutzername eine numerische Zahl, welche aufsteigend aus einem festen Nummernkreis vergeben wird (meist 4- oder 5-stellig mit führenden Nullen). Durch diese Art der Accountvergabe, sind die Benutzernamen sehr leicht zu erraten.

Basierend auf den 71 untersuchten HV-Accounts werden, je nach HV-Dienstleister, unterschiedliche Passwort-Richtlinien eingesetzt. Bis auf eine Ausnahme, vergeben alle

Dienstleister zufällig generierte Passwörter. Die Länge beträgt 5-10 Zeichen, wobei von den verfügbaren Zeichen Groß-, Kleinbuchstaben und Zahlen verwendet werden. Als Sonderzeichen kommt nur in einem Fall das „*“-Zeichen zum Einsatz. Das HV-Portal der HVBest Event-Service GmbH verwendet zur Authentifizierung kein zufälliges Passwort, sondern eine Kombination aus Anzahl der gehaltenen Aktien, PLZ und Wohnort des Aktionärs[7]. Die Passwort-Richtlinien der HV AG und der Art of Conference konnten nicht untersucht werden, da keine Zugangsdaten zur Verfügung standen (Marktanteil: 0,6 %).

Die HV-Portale von Better Orange IR & HV AG und HVBest Event-Service GmbH zeigen dem Anwender an, ob entweder Benutzername oder Passwort falsch eingegeben wurden (Marktanteil: 20,5 %). Dieses Verhalten liefert wertvolle Informationen über die Existenz von gültigen Accounts.

Die HV-Dienstleister Link Market Services GmbH und HVBest Event-Service GmbH sichern den Login zusätzlich über Captchas[8] ab, welche für einen erfolgreichen Login korrekt gelöst werden müssen (Marktanteil: 22,4 %).

Bei den HV-Portalen der Computershare GmbH & Co. KG, C-HV AG, ADEUS GmbH und FAE Management GmbH führte die mehrfache Eingabe von falschen Passwörtern zur Sperrung des jeweiligen Accounts (Marktanteil: 32,9 %). Während diese Sperrung bei der FAE Management GmbH nach ca. 30 Minuten wieder aufgehoben wird, können die Accounts bei den anderen drei Dienstleistern nur manuell durch den Support entsperrt werden. Bei allen anderen HV-Dienstleistern erfolgt durch die wiederholte Eingabe falscher Passwörter keine Sperrung von existierenden Accounts.

Im Bereich des Sessionmanagements waren die drei HV-Portale von Better Orange IR & HV AG, BS HV-Portal und C-HV AG für Session Fixiation-Angriffe anfällig (Marktanteil: 41,8 %). Bei der Better Orange IR & HV AG und der Computershare GmbH & Co. KG war die Logout-Funktionalität des HV-Portals wirkungslos (Marktanteil: 45,4 %). Als Konsequenz blieb die Benutzersession trotz Logout des Users gültig.

Die Untersuchungen im Bereich Broken Access Control offenbarten bei der Computershare GmbH & Co. KG (Marktanteil: 25,9 %) eine sehr kritische Schwachstelle, welche das Schutzziel der Vertraulichkeit betraf. Bei allen anderen HV-Dienstleistern wurden keine Schwachstellen im Bereich der Zugriffskontrolle gefunden.

Für CSRF-Angriffe war lediglich die FAE-Management GmbH anfällig. Hierdurch konnte z. B. das Abstimmungsverhalten des Opfers unbemerkt verändert werden.

Die beschriebenen Sachverhalte aus der Sicherheitsuntersuchung der Bereiche Sessionmanagement, Access Control und CSRF werden in Tabelle 4 zusammengefasst.

---

[7] Nach Rücksprache mit dem Dienstleister unterstützt dieser auch alternative Authentifizierungsverfahren und hat diese auch im Einsatz.

[8] Captchas sind kleine Aufgaben, welche in diesem Kontext dazu dienen, Brute Force-Angriffe zu erschweren. Sie sollen von Menschen (zumeist) effizient und von Maschinen nicht lösbar sein.

HV-Portal	Passwort-Richtlinie	Anzeige Passwort /Account falsch?	Captcha vorhanden?	Account sperrbar?
Computershare GmbH & Co. KG	6-stellig [A-Z, a-z, 0-9] oder [A-Z, a-z]	Nein	Nein	Ja
Link Market Services GmbH	6-stellig [0-9]	Nein	Ja	Nein
Better Orange IR & HV AG	8-stellig [A-Z, a-z, 0-9, *]	Ja	Nein	Nein
UBJ. GmbH	5-stellig [A-Z, a-z, 0-9] oder 8-stellig [0-9]	Nein	Nein	Nein
C-HV AG	8-stellig [A-Z, 0-9]	Nein	Nein	Ja
ADEUS GmbH	10-stellig [0-9]	Nein	Nein	Ja
GFEI AG	5- oder 6-stellig [A-Z, a-z, 0-9]	Nein	Nein	Nein
ITTEB GmbH & Co. KG	6-stellig [A-Z, a-z, 0-9]	Nein	Nein	Nein
BADER & HUBL GmbH	10-stellig [a-z, 0-9]	Nein	Nein	Nein
HV-Management GmbH	8-stellig [A-Z, a-z, 0-9]	Nein	Nein	Nein
AAA HV GmbH	5-stellig [0-9]	Nein	Nein	Nein
HVBest Event-Service GmbH	Anzahl Aktien, PLZ und Ort	Ja	Ja	Nein
FAE Management GmbH	8-stellig [0-9, a-f]	Nein	Nein	Ja (temporär für ca. 30 Min.)
HV AG	n/a	Nein	Nein	Nein
Art of Conference	n/a	Nein	Nein	Nein

Tabelle 3: Die Ergebnisse aus der Untersuchung der Passwort-basierten Authentifizierung

HV-Portal	Sessionmanagement		Broken Access Control	CSRF-Angriff möglich?
	Session Fixiation-Angriff möglich?	Session-Logout wirksam?		
Computershare GmbH & Co. KG	Nein	Nein	Ja	Nein
Link Market Services GmbH	Nein	Ja	Nein	Nein
Better Orange IR & HV AG	Ja	Nein	Nein	Nein
BS HV-Portal	Ja	Ja	Nein	Nein
C-HV AG	Ja	Ja	Nein	Nein
ADEUS GmbH	Nein	Ja	Nein	Nein
HVBest Event-Service GmbH	n/a	n/a	n/a	n/a
FAE Management GmbH	Nein	Ja	Nein	Ja

Tabelle 4: Die Ergebnisse der Sicherheitsuntersuchung der Bereiche Sessionmanagement, Access Control und CSRF

5. Diskussion

Nachfolgend werden die in Kapitel 4 dargestellten Ergebnisse diskutiert und bewertet.

5.1. Informationsgewinnung

Bei fünf von acht HV-Portalen, mit einem Marktanteil von 52,4 %, wurden verwundbare Softwarekomponenten verwendet. Hierbei ist anzumerken, dass durch den Einsatz von anfälligen Softwarebibliotheken nicht zwangsläufig die dort enthaltenen Schwachstellen ausgenutzt werden können. Es ist z. B. möglich, dass die verwundbare Funktionalität überhaupt nicht benutzt wird. Jedoch gehört es zu den anerkannten Security Best Practices, verwundbare Softwarekomponenten möglichst zeitnah zu patchen. Dies zeigt, dass es aus Sicherheitssicht im Patchmanagement einiger HV-Dienstleister noch Optimierungspotenzial gibt. Insbesondere, da die gefundenen Schwachstellen bereits 1-5 Jahre bekannt waren und Sicherheitsupdates existierten.

Die TLS-Konfiguration der HV-Portale offenbarte keine gravierenden Schwachstellen. Das B-Rating, welches der Hälfte der Anbieter vergeben wurde, beruht auf der serverseitigen Unterstützung der veralteten Protokollversionen TLS 1.0 und TLS 1.1. Jedoch werden beide Versionen bereits seit Mitte 2018 aus Sicherheitsgründen nicht mehr zum praktischen Einsatz empfohlen [9], [10]. Werden sie abgeschaltet, kann das Rating auf A und damit auf ein sehr gutes Sicherheitsniveau verbessert werden.

Im Bereich der Security Header wird leider sehr viel Potenzial für die Gesamtsicherheit der HV-Portale verschenkt, zumal es sich hierbei um eine Vielzahl von breit unterstützten und etablierten Security Best Practices handelt. Durch deren Einsatz können viele

bekannte Angriffe effektiv unterbunden werden. Diese aus Sicherheitssicht „Low Hanging Fruits“ können ohne großen Aufwand und innerhalb kürzester Zeit implementiert werden. Dies zeigen auch die Rückmeldungen und bereits umgesetzten Optimierungen der HV-Dienstleister.

Insgesamt zeigen die Ergebnisse aus der Informationsgewinnung, dass die potenzielle Angriffsfläche der HV-Portale noch deutlich verringert werden kann.

5.2. Manuelle Sicherheitsprüfung

Für die Aktionärs-Accounts erzeugen sieben von acht HV-Dienstleistern zufällig generierte Passwörter und verhindern somit schwache Kennwörter, wie z. B. „12345“ oder den Einsatz derselben Passwörter bei unterschiedlichen Diensten.

Nach den Empfehlungen des Bundesamtes für Sicherheit in der Informationstechnik, sollte ein sicheres Passwort, nach Stand der Technik, mindestens acht Zeichen lang sein und aus Groß- und Kleinbuchstaben, Ziffern und Sonderzeichen bestehen. Passwörter, welche nicht alle vier Zeichenarten verwenden, sollten deutlich länger sein (>12 Zeichen) oder mit einer Mehr-Faktor-Authentisierung abgesichert werden [11].

Bei den untersuchten HV-Portalen wurden diese Vorgaben von keinem Anbieter erfüllt. Die geringste Passwort-Komplexität wies die AAA Management HV GmbH auf (10^5 mögliche Passwort-Kombinationen). Die HVBest Event-Service GmbH verwendet keine Passwörter, sondern erwartet neben dem Benutzernamen die richtige Eingabe von Aktienanzahl, PLZ und Wohnort des Aktionärs. Diese Art der Authentifizierung ist als relativ schwach anzusehen, da die notwendigen Informationen leicht beschaffbar (PLZ und Wohnort eines Aktionärs) bzw. erratbar sind (Aktienanzahl).

In der Diskussion mit den HV-Dienstleistern wurde als Grund für die niedrigere Passwort-Komplexität der geringere Supportaufwand angeführt. Im Hinblick auf die Kritikalität der Zugangsdaten im Kontext von virtuellen HVs und das verfügbare Angreifer-Zeitfenster von bis zu 12 Tagen[9], sind komplexere Passwort-Richtlinien bzw. die Verwendung einer Mehrfaktor-Authentifizierung sehr zu empfehlen.

HVBest Event-Service GmbH und Link Market Services GmbH sichern den Login zusätzlich durch Captchas ab. Hierdurch soll ein Brute Force-Angriff erschwert werden. Jedoch sind die verwendeten Captchas sehr einfach aufgebaut und verwenden nur eine geringe Anzahl von möglichen Zeichen (siehe Abbildung 3). Wie in [12] gezeigt wurde, kann diese Art von Captchas automatisiert gelöst werden.

Abbildung 3: Beispiele für Captchas der Link Market Services GmbH (links) und der HVBest Event-Service GmbH (rechts)

[9] 12 Tage vor einer HV ist der Nachweisstichtag über den Anteilsbesitz ([1], Artikel 2, §1 Abs. 3). Spätestens ab diesem Zeitpunkt ist das HV-Portal aktiv zu schalten.

Um das systematische Erraten von Passwörtern zu verhindern, sperren vier HV-Dienstleister nach mehrmaliger Falscheingabe den zugehörigen Benutzeraccount. Dadurch eröffnet sich aber die Möglichkeit eines Brute Force-Angriffs auf die Verfügbarkeit des HV-Portals. Die FAE Management GmbH entschärft diese Bedrohung, indem die Accounts automatisch nach ca. 30 Minuten wieder entsperrt werden. Problematisch ist in diesem Zusammenhang, dass alle Dienstleister vorhersehbare Benutzernamen einsetzen. Dadurch ist es möglich, kurz vor einer HV, gezielt einzelne oder alle Aktionärs-Accounts zu sperren. Zudem können bei der Better Orange IR & HV AG und der HVBest Event-Service GmbH (Marktanteil: 20,5 %) gültige Accounts aufgrund der Fehlermeldung vom Angreifer identifiziert werden. Es wird daher als Security Best Practice geraten, bei fehlerhaften Loginversuchen immer generische Fehlermeldungen anzuzeigen (z. B. „Login fehlgeschlagen!“).

Die im Bereich Sessionmanagement nachgewiesenen Session Fixiation-Angriffe zeigten, dass bei drei HV-Portalen (45,8 % Marktanteil) der Angreifer die Benutzersession des Opfers vollständig übernehmen kann. Grundlegend hierfür war, dass dem HV-Teilnehmer, nach erfolgreicher Authentifizierung, keine neue Session ID vergeben wurde. Die bei zwei HV-Portalen (Marktanteil: 45,4 %) nicht wirksame Logout-Funktionalität eröffnete ebenfalls die Möglichkeit der vollständigen Übernahme der Session. Diese kann z. B. dadurch übernommen werden, dass der Angreifer nach dem ordnungsgemäßen Logout des Opfers Zugriff auf die im Browser weiterhin gespeicherten Sessiondaten bekommt.

Im Bereich fehlerhafte Zugriffskontrolle ist die im HV-Portal des größten HV-Dienstleisters (Marktanteil: 25,9 %) gefundene Schwachstelle als sehr kritisch einzustufen. Sie erlaubte es, die personenbezogenen Daten (Name, Adresse, Geburtsdatum, usw.) *aller* Aktionäre, inkl. Abstimmungsverhalten und deren Anteilsbesitz, von *allen* durchgeführten virtuellen HVs auszulesen. Hierfür war lediglich ein gültiger Account für eine beliebige HV notwendig.

Der bei einem HV-Portal (Marktanteil: 1,0 %) durchführbare CSRF-Angriff erlaubte es, durch Klick des Opfers auf einen Link das Abstimmungsverhalten zu ändern.

Tabelle 5 zeigt, basierend auf den praktisch nachgewiesenen Angriffen, die verletzten Schutzziele je HV-Portal. Dabei wurden die schwachen Passwort-Richtlinien, welche zu einem Verlust der Vertraulichkeit und Integrität führen können, nicht mit einbezogen. Hintergrund: Brute Force-Angriffe wurden bei den Untersuchungen nicht durchgeführt, um die Funktionalität der HV-Portale nicht zu gefährden. Jedoch haben alle kontaktierten HV-Dienstleister diese Angriffsmöglichkeit als sicherheitsrelevant eingestuft.

Insgesamt wurde bei sechs von acht HV-Portalen (Marktanteil: 71,6 %), mindestens ein Schutzziel auf Basis der Bedrohungsanalyse und der dort gezeigten Auswirkungen kompromittiert. Lediglich das HV-Portal der Link Market Services GmbH wies keine praktisch nachgewiesenen Angriffsmöglichkeiten auf. Das HV-Portal HVBest Event-Service GmbH konnte aufgrund fehlender Zugangsdaten nicht vollumfänglich getestet werden. Dennoch zeigt die Analyse der Angriffsfläche, dass auch bei diesen Anbietern durchaus Potenzial zur Erhöhung des Sicherheitsniveaus besteht.

HV-Portal	Marktanteil	Kompromittiertes Schutzziel		
		Vertraulichkeit	Integrität	Verfügbarkeit
Computershare GmbH & Co. KG	25,9 %	X		X
Link Market Services GmbH	21,4 %			
Better Orange IR & HV AG	19,5 %	X	X	
BS HV-Portal	19,2 %	X	X	
C-HV AG	3,1 %	X	X	X
ADEUS GmbH	2,9 %			X
HVBest Event-Service GmbH	1,0 %	n/a	n/a	n/a
FAE Management GmbH	1,0 %		X	

Tabelle 5: Darstellung der kompromittierten Schutzziele der untersuchten HV-Portale

6. Stand der Forschung

Die bisherigen Veröffentlichungen im Bereich virtuelle HVs beschränken sich weitgehend auf deren juristische, organisatorische und technische Risiken. In [13] untersucht C. Danwerth empirisch die rechtlichen Modalitäten und Gestaltungsvarianten von virtuellen HVs. Die ermittelten Marktanteile der drei größten Dienstleister für virtuelle HVs decken sich mit den Ergebnissen dieses Beitrags.

M. Scholze und M. Kaspar zeigen in [14], dass der Fokus bei virtuellen HVs auf einer rechtssicheren und störungsfreien Durchführung liegt. Aspekte der IT-Sicherheit spielen hierbei keine oder nur eine untergeordnete Rolle. Auf der anderen Seite beklagen Aktionärsvereinigungen berechtigterweise die juristisch begründeten Einschränkungen der Aktionärsrechte durch das neue virtuelle Format [15], [16].

7. Fazit und Ausblick

Virtuelle HVs haben sich innerhalb kürzester Zeit als ein systemrelevantes Kriseninstrument etabliert. Neben allen juristischen, organisatorischen und technischen Aspekten werden virtuelle HVs langfristig nur bei einem ausreichend hohen Sicherheitsniveau erfolgreich sein. Die Ergebnisse der ersten empirisch und systematisch durchgeführten Sicherheitsuntersuchung von virtuellen HVs zeichnet dabei ein verbesserungswürdiges Bild der Sicherheitslage. In sechs von acht HV-Portalen, mit einem Marktanteil von 71,6 %, wurden kritische Sicherheitslücken gefunden.

Im Hinblick auf die aktuelle und zukünftig sicher zunehmende Wichtigkeit von virtuellen HVs ist hier weiterer Handlungsbedarf zu sehen. Im Rahmen dieses Beitrags konnten die betroffenen HV-Dienstleister sensibilisiert werden. In einem ersten Schritt wurde gemeinsam mit ihnen das Sicherheitsniveau der HV-Portale deutlich erhöht. Hierbei ist positiv zu erwähnen, dass alle Beteiligten sehr professionell agierten und zeitnah die gemeldeten Schwachstellen behoben haben. Darüber hinaus wurden viele der genannten Security Best Practices (z. B. der Einsatz von Security Headern, komplexere Passwort-Richtlinien und aktives Patchmanagement) proaktiv umgesetzt.

Literaturhinweise

[1] Bundesministerium für Justiz und Verbraucherschutz, „Gesetz zur Abmilderung der Folgen der COVID-19-Pandemie im Zivil-, Insolvenz- und Strafverfahrensrecht“, https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/Bgbl_Corona-Pandemie.pdf, 27.3.2020.

[2] Bundesministerium der Justiz und für Verbraucherschutz, „Verlängerung der Regelungen zur virtuellen Hauptversammlung bis Ende 2021 tritt in Kraft“, https://www.bmjv.de/SharedDocs/Pressemitteilungen/DE/2020/102920_Virtuelle_Hauptversammlung.html, 23.10.2020.

[3] The OWASP Foundation, „OWASP Top 10 – 2017“, https://owasp.org/www-project-top-ten/2017/, 2017.

[4] M. Kolšek, „Session Fixation Vulnerability in Web-based Applications“, http://www.acrossecurity.com/papers/session_fixation.pdf, 2002-2007.

[5] W. Zeller, E. W. Felten, „Cross-Site Request Forgeries: Exploitation and Prevention“, https://www.cs.utexas.edu/users/shmat/courses/library/zeller.pdf, 2008.

[6] Wappalyzer, „Wappalyzer Browser Add-On“, https://www.wappalyzer.com/, 2008-2021.

[7] S. Helme, „Security Headers“, https://securityheaders.com/, 2016-2021.

[8] Qualys Inc., „SSL Labs: SSL Server Test“, https://www.ssllabs.com/ssltest/, 2009-2021.

[9] Bundesamt für Sicherheit in der Informationstechnik, „Technische Richtlinie TR-02102-2 Kryptographische Verfahren: Empfehlungen und Schlüssellängen - Teil 2 - Verwendung von Transport Layer Security (TLS)“, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR02102/BSI-TR-02102-2.pdf, 28.2.2020.

[10] K. Moriarty, S. Farrell, „Deprecating TLSv1.0 and TLSv1.1 draft-moriarty-tls-oldversions-diediedie-01“, IETF Internet Draft, https://tools.ietf.org/id/draft-moriarty-tls-oldversions-diediedie-01.html, 25.7.2018.

[11] BSI für Bürger, „Sichere Passwörter erstellen“, https://www.bsi-fuer-buerger.de/BSIFB/DE/Empfehlungen/Passwoerter/passwoerter_node.html, online abgerufen am 31.12.2020.

[12] J. Yan, A. Salah El Ahmad, „Captcha robustness: A security engineering perspective“, Computer 44.2: 54-60, 2010.

[13] C. Danwerth, „Modalitäten und Gestaltungsvarianten der virtuellen Hauptversammlung – Eine empirische Untersuchung aller im April 2020 veröffentlichten Einberufungen präsenzloser Versammlungen von börsennotierten Unternehmen“, Die Aktiengesellschaft (AG) 2020, Heft 11: 418-432, 2020.

[14] M.Scholze, M. Kaspar, „Virtuelle Hauptversammlungen: Rückschau und Ausblick“, HV Magazin 04/2020: 12-17, 2020.

[15] Schutzgemeinschaft der Kapitalanleger e.V., „SdK fordert keine Einschränkung von Aktionärsrechten bei virtuellen Hauptversammlungen“, https://sdk.org/veroeffentlichungen/pressemitteilungen/sdk-fordert-keine-einschraenkung-von-aktionaersrechten-bei-virtuellen-hauptversammlungen/, 6.4.2020.

[16] DSW - Deutsche Schutzvereinigung für Wertpapierbesitz e.V., „Online-HV nur inklusive Aktionärsrechte“, https://www.dsw-info.de/presse/pressemitteilungen-2020/online-hv-nur-inklusive-aktionaersrechte/, 20.3.2020.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Die sichere Digitalisierung von Wahlen am Beispiel des Modellprojektes Online-Sozialwahlen 2023

Jennifer Breuer[1], Michael Hoppe[1], Stephan Kohzer[1], Marina Voigtländer[1]

Kurzfassung:
Dieser Beitrag soll einen Überblick darüber geben, wie Wahlen, die bisher als Präsenz- und/oder Briefwahl stattfinden, digital umgesetzt werden können. Am Beispiel der Online-Sozialwahlen werden die in dem Modellprojekt erarbeiteten IT-sicherheitstechnischen Vorgaben abstrahiert, um diese auf andere Wahlen übertragen zu können. Sowohl die Umsetzung der Wahlrechtsgrundsätze (allgemein, unmittelbar, frei, gleich, geheim und öffentlich) als auch die Absicherung solch einer Wahl bedürfen einer sorgfältigen Vorarbeit im Rahmen einer Schutzbedarfsfeststellung. Darauf aufbauend müssen sowohl geeignete Maßnahmen im Rahmen eines zertifizierten ISMS identifiziert und umgesetzt, als auch ein geeignetes Mittel zur Identifizierung der Wahlberechtigten nach dem zuvor festgelegten Vertrauensniveau gewählt werden. Zudem muss sichergestellt werden, dass das Wahlgeheimnis gewahrt wird. Kryptografie und ein geeignetes Schlüsselmanagement sind unerlässlich, um eine sichere Online-Wahl zu gewährleisten. Die Wahlverantwortlichen müssen anhand einer Risikoanalyse festlegen, wie die Restrisiken minimiert werden. Im Ausblick wird die Möglichkeit zur Nutzung mobiler Identitäten als gleichermaßen sichere wie nutzerfreundliche Lösung genannt. Dazu müssen die IT-sicherheitstechnischen Anforderungen separat betrachtet werden.

Stichworte: BSI TR-03107, BSI TR-03162, eIDAS, IT-Grundschutz, Kryptografie, Online-Ausweisfunktion, Online-Wahlen, Personalausweis, Sozialwahlen, Wahlrechtsgrundsätze

1. Einführung

Nicht nur vor dem aktuellen Hintergrund der Corona-Pandemie hat die Digitalisierung einen neuen Schub erhalten. Für Präsenzwahlen wie die Kommunalwahlen in Nordrhein-Westfalen mussten besondere Vorkehrungen getroffen werden. Die Warteschlangen wurden länger, was dafür gesorgt haben kann, dass so mancher umgekehrt ist. Die Alternative Briefwahl muss fristgerecht beantragt und abgeschickt werden. Doch auch unabhängig davon gilt es den Weg zur Wahlurne zu verkürzen und somit die Wahlbeteiligung zu erhöhen. Wahlen gibt es auf vielen Ebenen: Politische Wahlen, Betriebsratswahlen, Wahlen in Vereinen. Ein anderes Beispiel sind die Sozialversicherungswahlen.

Im Rahmen eines Modellprojektes wird den Krankenkassen bei den Sozialversicherungswahlen im Jahr 2023 neben der herkömmlichen Stimmabgabe per Briefwahl fakultativ die Möglichkeit eröffnet, Online-Wahlen durchzuführen. Die Wahlberechtigten erhalten somit die Möglichkeit ihr Votum neben der Briefwahl auch elektronisch über das Internet abzugeben. Mit der Technischen Richtlinie TR-03162 „IT-sicherheitstechnische Anforderungen zur Durchführung einer Online-Wahl im Rahmen des Modellprojekts nach § 194a Fünftes Buch Sozialgesetzbuch (Online-Wahl)“ macht das BSI Vorgaben für die Informationssicherheit und schafft somit eine wesentliche Grundlage für die sichere Digitalisierung der Sozialversicherungswahlen 2023. Der vorliegende

[1] Bundesamt für Sicherheit in der Informationstechnik

Beitrag gibt einen Überblick, wie das Beispiel des Modellprojekts eine Grundlage für die Digitalisierung weiterer Wahlen sein kann.

2. Vorgehensweise zur Digitalisierung der Wahl

Aufgrund dessen, dass es sich bei den Online-Sozialwahlen um ein beispielloses Modellprojekt handelt, wurde zur Vorbereitung der Digitalisierung der Wahl zunächst die Machbarkeit geprüft. Für die Digitalisierung anderer Wahlen, müssen diese dafür erforderlichen Schritte ebenso vollzogen werden.

1. Prüfung der rechtlichen Rahmenbedingungen und Voraussetzungen zur Durchführung einer Online-Wahl. Sind diese noch nicht gegeben, müssen sie geschaffen werden.
   - Im Rahmen der Online-Sozialwahlen (OSW) sind die Voraussetzungen mit der Neufassung der §§ 194a bis 194d SGB V, der Online-Wahl-Verordnung sowie den geänderten Satzungen der Krankenkassen gegeben.
2. Analyse der Wahl und Anwendung der Wahlrechtsgrundsätze
   - Bei einer reinen Briefwahl spielen andere Wahlrechtsgrundsätze ggf. eine stärkere Rolle als z.B. bei einer Präsenzwahl (siehe dazu Kapitel 3). Eine weitere Überlegung sollte zudem sein, ob die Wahl rein digital stattfinden oder die Möglichkeit zur Online-Abstimmung als zusätzliche Alternative neben einer Brief- oder Präsenzwahl ermöglicht werden soll.
3. Wahlprozesse analysieren und Schutzbedarf festlegen
   - Im Rahmen des Modellprojektes OSW wurden Geschäftsprozesse und Anwendungen identifiziert sowie
   - eine Schutzbedarfsfeststellung auf Basis des IT-Grundschutzes erarbeitet. Soweit der Schutzbedarf nicht bereits in einer Verordnung oder einem Gesetz explizit vorgegeben ist, obliegt es dem Prozessverantwortlichen, diesen festzulegen.
4. Risikoanalyse durchführen
   - In einer Risikoanalyse muss festgelegt werden, wie die Risiken minimiert werden und mit den Restrisiken umgegangen wird.

3. Umsetzung der Wahlrechtsgrundsätze

Wie bereits im vorangegangenen Kapitel kurz erwähnt, spielen die Wahlrechtsgrundsätze eine wichtige Rolle für die Digitalisierung von Wahlen. Denn für Online-Wahlen gelten die Grundsätze allgemein, unmittelbar, frei, gleich, geheim und öffentlich im gleichen Maße wie für Präsenz- und Briefwahlen.

- Die Allgemeinheit einer Wahl besagt, dass grundsätzlich alle natürlichen Personen teilnehmen dürfen. Bei einer Bundestagswahl sind es beispielsweise alle Staatsbürger. Eingeschränkt werden kann dies z.B. durch ein Mindestalter.

- Eine unmittelbare bzw. direkte Wahl bedeutet, dass Wähler direkt die Wahlvorschläge wählen, ohne dass dazwischen Wahlmänner bzw. -frauen stehen, wie beispielsweise in den USA.
- Um dem Wahlrechtsgrundsatz der freien Wahl gerecht zu werden, muss die Wahlentscheidung durch den Wähler ohne Zwang oder sonstige Beeinflussung getroffen werden können. Dies ist bei einer Präsenzwahl so umgesetzt, dass immer nur eine Person allein in der Wahlkabine ihr Kreuz machen darf. Bei einer Online- oder Briefwahl ist die Umgebung, in der der Wähler sein Kreuz macht, nicht durch den Wahlverantwortlichen kontrollierbar.
- Wahlgleichheit bedeutet, dass jeder Wähler die gleiche Anzahl an Stimmen hat und jede Stimme die gleiche Gewichtung. Bei einer Online-Wahl muss dies durch eine sichere Identifizierung der Wähler und weitere Sicherheitsmaßnahmen sichergestellt werden.
- Die geheime Wahl erfordert, dass jeder Wähler anonym wählen kann und keine Zuordnung zwischen Wähler und abgegebener Stimme möglich ist (siehe dazu auch Abschnitt 5.2.).
- Der Grundsatz der Öffentlichkeit ermöglicht den Wählern, das Wahlergebnis nachvollziehen zu können. Bei einer Präsenzwahl hat der Wähler beispielsweise das Recht, sich während der Wahl und bei der Ermittlung des Wahlergebnisses im Wahlraum aufzuhalten. Dies ist bei einer Brief- und Online-Wahl so nicht umsetzbar.

Die Umsetzung der Wahl kann dabei nicht jedem Wahlrechtsgrundsatz auf dieselbe Weise gerecht werden. Die größte Herausforderung bei einer Online-Wahl ist, den beiden Wahlrechtsgrundsätzen Geheim und Öffentlich in gleichem Maße gerecht zu werden. Insbesondere seit dem Urteil des Bundesverfassungsgerichtes[2] ist die Möglichkeit des Wählers, nachvollziehen zu können, ob die Online-Stimme wie abgegeben in der (elektronischen) Wahlurne gespeichert wurde, nicht mehr außer Acht zu lassen.

Dies bedeutet, dass auf der einen Seite die Stimmabgabe so geschützt werden muss, dass eine hohe Sicherheit in Bezug auf die Wahrung des Wahlgeheimnisses erreicht wird. Auf der anderen Seite muss trotzdem insgesamt ein ausreichendes Maß an Transparenz über das Wahlgeschehen und eine weitgehende Nachvollziehbarkeit der wesentlichen Schritte der Wahlhandlung, der Ermittlung des Wahlergebnisses und des Wahlergebnisses selbst für die Öffentlichkeit gewährleistet werden.

Für die Digitalisierung einer Wahl spielt daher auch eine Rolle, wie die Wahlrechtsgrundsätze bisher gewichtet wurden. Ziel sollte dennoch sein, den Wahlrechtsgrundsätzen so gut wie möglich gerecht zu werden. Zur Nachvollziehbarkeit dieser Entscheidungen sollte dokumentiert werden, wenn ein Wahlrechtsgrundsatz beispielsweise aufgrund der Umsetzung eines anderen Wahlrechtsgrundsatzes und/oder zur Erhöhung der Manipulationssicherheit der Wahl in geringerem Maße umgesetzt werden kann.

[2] BVerfG, Urteil des Zweiten Senats vom 3. März 2009, 2 BvC 3/07, 2 BvC 4/07

4. Mögliche Angriffsvektoren

Zur Erfüllung der Wahlrechtsgrundsätze und zum Erhalt eines validen Wahlergebnisses spielt die Sicherheit vermutlich die wichtigste Rolle. Insbesondere Online-Wahlen beinhalten sensible Informationen und sind einer Reihe von Bedrohungen ausgesetzt. Jede Art von Manipulation der Wahlhandlung muss ausgeschlossen werden. Der Begriff „Manipulation“ beinhaltet dabei jede Form des unberechtigten Lesens, Änderns, Hinzufügens oder Löschens von Informationen sowie der Beeinflussung der Verfügbarkeit.

Bedrohungen im Rahmen von Online-Wahlen lassen sich in verschiedene Kategorien mit spezifischen technischen und/oder organisatorischen Gegenmaßnahmen gliedern:

- Angriffe durch Außentäter vs. Innentäter
- Angriffe auf IT-Systeme vs. Beeinflussung des Wahlberechtigten

Exemplarische Bedrohungen sind[3]:

- Angriffe auf notwendige Daten des Wahlverfahrens
- Angriffe auf das Online-Wahlsystem
- Beeinflussung von verantwortlichen Personen
- Beeinflussung des Wahlberechtigten
- Angriff auf das Client-Gerät des Wählers

Im Gegensatz zu Präsenz- und Briefwahlen können daher im schlimmsten Fall nicht nur einzelne Stimmen, sondern das komplette Wahlergebnis manipuliert werden. Auch wenn keine 100 prozentige Sicherheit gewährleistet werden kann, muss doch sichergestellt werden, dass das Risiko einer Manipulation und damit der Nichtigkeit der Wahl, auf ein Minimum reduziert wird. Dies muss in einer Risikoanalyse des jeweils Prozessverantwortlichen dokumentiert sein.

5. Allgemeine Absicherung der Wahl

Um die unter Kapitel 4 genannten Risiken zu minimieren, müssen Sicherheitsmechanismen eingebaut werden. Diese beinhalten sowohl technische als auch organisatorische Maßnahmen. Egal welche Wahl, ob zum Vereinsvorstand oder die Bundestagswahl, per Online-Wahl durchgeführt wird, muss der Informationsverbund, der genutzt wird, durch ein Informationssicherheitsmanagementsystem (ISMS) abgesichert werden. Eine Vorgehensweise nach IT-Grundschutz[4] ist hier zu empfehlen. Zur Umsetzung des IT-Grundschutzes müssen vorab sowohl eine Schutzbedarfsfeststellung als auch eine Strukturanalyse durchgeführt werden. Diese fallen bei jeder Wahl unterschiedlich aus, da die Rahmenbedingungen zur Online-Wahldurchführung anders sind und die Wahlverantwortlichen auch für die Übernahme der Restrisiken verantwortlich sind. So sind beispielsweise betriebsinterne Wahlen in einer kontrollierten Umgebung anders zu betrachten als Wahlen über das Internet.

[3] Siehe auch TR-03162 Kapitel 2.1.1.3
[4] BSI IT-Grundschutz, 2020

Die ISMS-Anwendung sollte durch eine Zertifizierung nachgewiesen werden. Dies ist nicht nur hilfreich, um eventuelle Lücken im ISMS zu finden und entsprechende Maßnahmen zu ergreifen, sondern auch, um das Vertrauen der Wähler zu gewinnen. Denn nur eine IT-sichere Wahl kann manipulationssicher durchgeführt werden und so ein valides Ergebnis liefern.

Doch der Schutz der Online-Wahl muss selbstverständlich über ein zertifiziertes ISMS hinausgehen. Die eingesetzte Wahlsoftware muss den zuvor festgelegten Schutzzielen und der Schutzbedarfsfeststellung gerecht werden. In den folgenden Abschnitten sind dazu exemplarisch drei wichtige Faktoren beschrieben.

5.1. Identifizierung der Wahlberechtigten

Ein Schlüsselelement zur Gewährleistung der IT-Sicherheit von Online-Wahlen ist die sichere Identifizierung und Authentisierung der Wahlberechtigten. Damit soll sichergestellt werden, dass nur identifizierte Nutzer mit gültiger Wahlberechtigung ihre Stimme abgeben können. Eine sichere Identifizierung der Wähler ist auch wichtig, um dem Wahlrechtsgrundsatz der freien Wahl gerecht zu werden. Identitätsdiebstahl würde dafür sorgen, dass jemand im Namen einer anderen Person wählt und somit Einfluss auf die abgegebene Stimme nimmt. Dazu kommt, dass niemand öfter als erlaubt wählt (Wahlrechtsgrundsatz Gleich). Um dies zu gewährleisten, ist es wichtig, ein geeignetes Mittel zur Identifizierung und Authentisierung der Wahlberechtigen einzusetzen. Unter Berücksichtigung des ermittelten Schutzbedarfs sollte ein Verfahren zur elektronischen Identifizierung eingesetzt werden, dessen Vertrauensniveau die Anforderung der Technischen Richtlinie TR-03107[5] bzw. der eIDAS-Verordnung[6] erfüllt. In Abhängigkeit vom Vertrauensniveau stehen dann unterschiedliche Verfahren zur Verfügung.

Eine sichere Möglichkeit, die auch bei der Online-Sozialwahl genutzt wird, ist der Einsatz des Personalausweises, des elektronischen Aufenthaltstitels oder der eID-Karte für Unionsbürger. Mithilfe der Online-Ausweisfunktion erfüllt dieser das eIDAS-Vertrauensniveau „hoch“ und ist bisher das einzige auf diesem Niveau notifizierte Identifizierungsmittel in Deutschland. Die integrierte 2-Faktor-Authentifizierung (Besitz des Personalausweises und Wissen der PIN) und die sichere Übertragung der Identitätsdaten (siehe Abbildung 1) sorgen dafür, dass der Wahlvorstand zusichern kann, dass nur sicher identifizierte Wahlberechtigte abgestimmt haben.

[5] BSI TR-03107-1, 2019

[6] eIDAS EU VO 910/2014, 2014

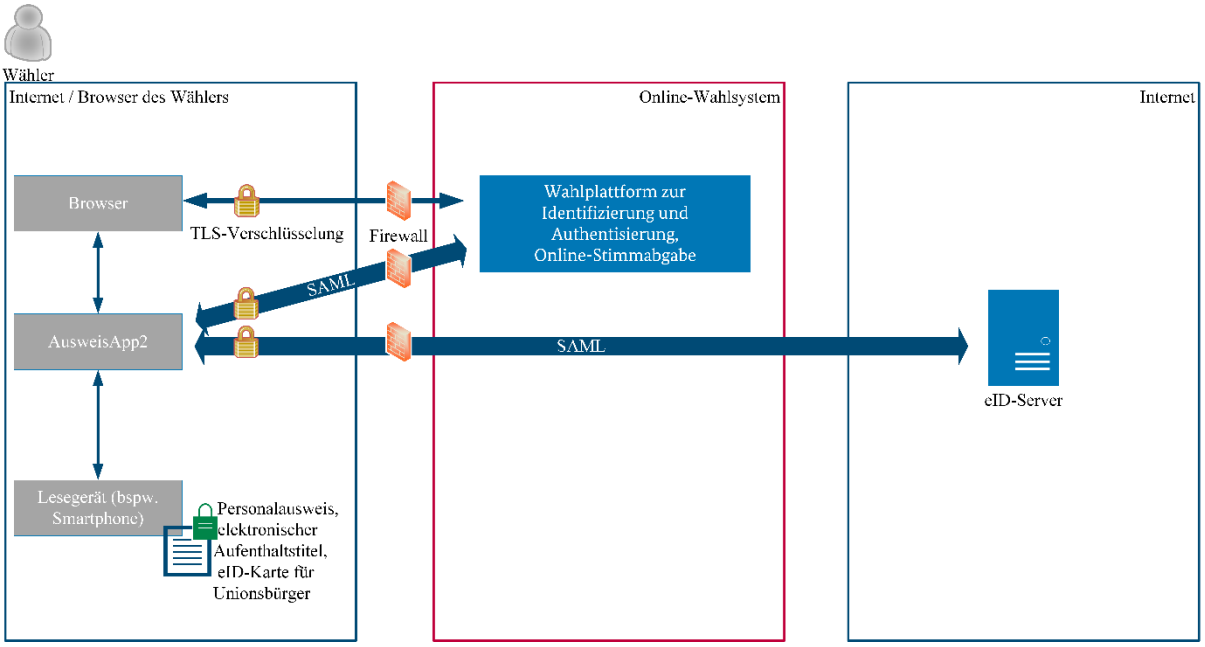


Abbildung 1: Nutzung der Online-Ausweisfunktion im Rahmen von Online-Wahlen

Identifiziert sich der Wähler mit der Online-Ausweisfunktion wird nach Eingabe der PIN durch den Nutzer durch den eID-Client (bspw. die AusweisApp2) ein Ende-zu-Ende verschlüsselter Kanal zwischen Ausweis-Chip und eID-Server hergestellt. Hierbei prüft der eID-Client im Zusammenspiel mit dem Ausweis-Chip das Berechtigungszertifikat des Diensteanbieters, um dessen Authentizität sowie dessen Berechtigung zum Auslesen der Identitätsdaten des Nutzers zu verifizieren. Seitens des eID-Server werden die Authentizität und Integrität des Ausweises, und zudem durch Abgleich mit der zentralen Sperrliste dessen Gültigkeit, geprüft. Anschließend werden die Identitätsdaten durch den eID-Server aus dem Ausweis ausgelesen. Von dort aus werden die Daten dann Ende-zu-Ende verschlüsselt an die Webanwendung bzw. das Online-Wahlsystem übertragen.

Die Online-Ausweisfunktion ist sowohl im Personalausweis, als auch im elektronischen Aufenthaltstitel und der eID-Karte für Unionsbürger integriert. Aufgrund dessen, dass diese Karten auch mit den aktuellen Smartphone-Modellen[7] unter Verwendung der NFC-Schnittstelle auslesbar sind, muss der Wähler so weder auf Sicherheit noch auf Nutzerfreundlichkeit verzichten.

5.2. Trennung der Wahlberechtigung vom Wählervotum

Um dem Wahlrechtsgrundsatz Geheim gerecht zu werden, ist es von immenser Wichtigkeit, die zuvor erwähnten Identifikationsdaten des Wählers von seiner Stimme zu trennen. Eine Zuordnung von Stimme zu Wähler muss ausgeschlossen sein. Vor dem Hintergrund der Nachvollziehbarkeit der Wahl ist dies eine weitere Herausforderung bei der Umsetzung der Wahlrechtsgrundsätze. Sobald die Wahlberechtigung geprüft wurde,

[7] Die Liste der kompatiblen Geräte finden Sie unter: https://www.ausweisapp.bund.de/mobile-geraete/

dürfen die personenbezogenen Daten nicht mehr mit dem weiteren Wahlablauf verknüpft werden. Dies gilt auch für etwaige Pseudonyme, wie einem Wahlkennzeichen. Für die Umsetzung des Wahlgrundsatzes Öffentlich müssen demnach Mechanismen integriert werden, die dem nicht widersprechen.

Erst nach der Trennung von den personenbezogenen Daten darf dem Wähler die Möglichkeit gegeben werden, die Online-Stimme abzugeben. Die folgende Abbildung zeigt ein Beispiel aus den Online-Sozialwahlen, wie dies aussehen kann. Dabei werden in den eckigen Klammern Optionen angezeigt, die entweder die Sicherheit erhöhen oder der Nachvollziehbarkeit des Wahlergebnisses dienen.[8]

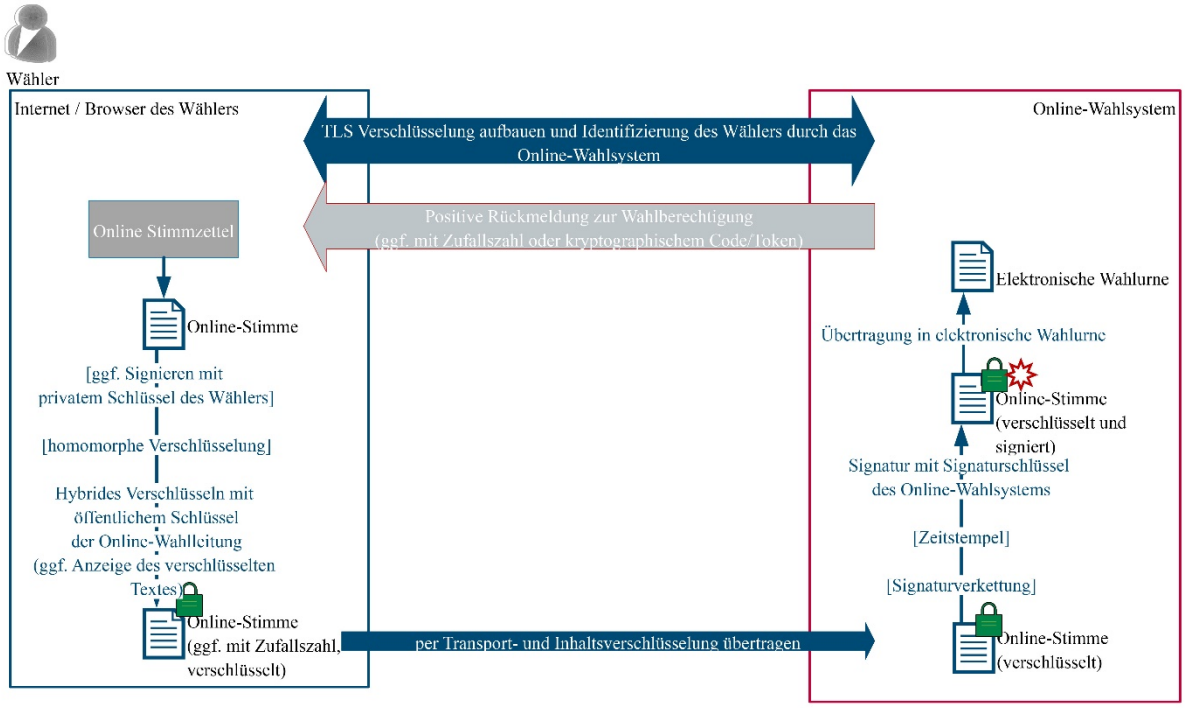


Abbildung 2: Möglichkeit der Umsetzung der Online-Stimmabgabe

Eine wichtige Rolle spielen bei Online-Wahlen auch die Übertragungswege. Sind diese nicht sicher, können sowohl die personenbezogenen Daten, als auch die Online-Stimme mitgelesen und so eine Verknüpfung hergestellt werden. Deshalb ist eine Transportverschlüsselung nach BSI TR-03116-4 bei allen Übertragungswegen Pflicht.[9]

5.3. Kryptografie und Schlüsselmanagement

Neben der zuvor genannten Transportverschlüsselung haben weitere Verschlüsselungsmechanismen, sowie Signaturen und Zeitstempel eine besondere Bedeutung zur Absicherung von Online-Wahlen. Mit deren Hilfe soll eine unbemerkte Manipulation der Wahl ausgeschlossen werden. Der Einsatz dieser Mittel ist unabhängig von der Art der Online-Wahl und muss immer nach Stand der Technik gewählt werden. Abbildung 3

[8] Siehe auch TR-03162 Kapitel 4.5
[9] BSI TR-03116-4, 2020

veranschaulicht beispielhaft, welches Schlüsselmaterial bei der Ermittlung des Wahlergebnisses genutzt werden kann. Ähnlich wird dies so bei den Online-Sozialwahlen umgesetzt.

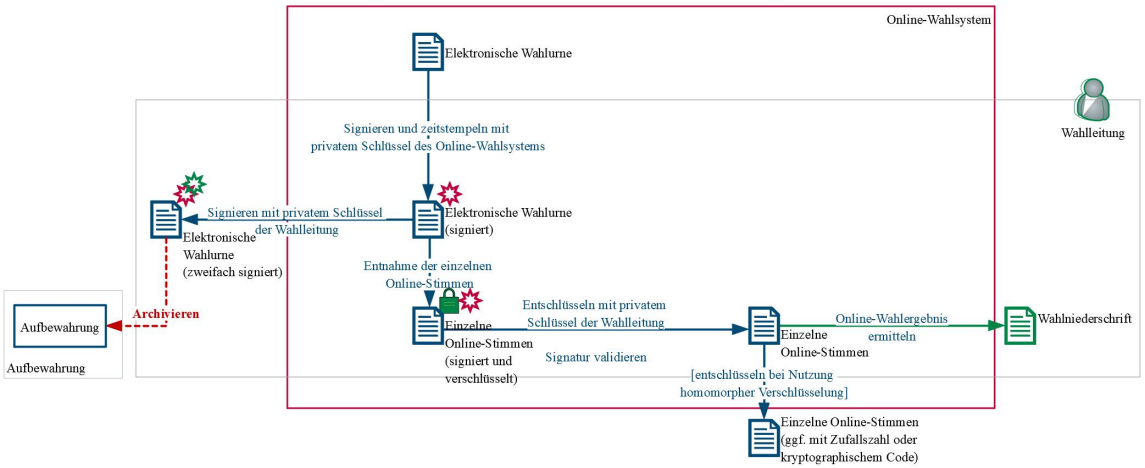


Abbildung 3: Beispielhafte Anwendung von Kryptografie im Online-Wahlverfahren

Die elektronische Wahlurne wird nach Abschluss der Wahl von dem Online-Wahlsystem signiert und mit einem Zeitstempel versehen. Vor der Entschlüsselung wird die elektronische Wahlurne dann noch einmal mit dem privaten Signaturschlüssel der Wahlleitung signiert. Diese zweifach signierte Version der Wahlurne wird außerhalb des Online-Wahlsystems aufbewahrt. Innerhalb der elektronischen Wahlurne befinden sich die signierten und verschlüsselten Online-Stimmen. Signaturen müssen immer validiert werden. Dann werden die Stimmen für die Ermittlung des Wahlergebnisses entschlüsselt. Optional ist hier die Möglichkeit genannt, dass die Stimmen homomorph verschlüsselt wurden. Der Vorteil einer homomorphen Verschlüsselung ist, dass die Stimmen nicht entschlüsselt werden müssen, um das Wahlergebnis zu ermitteln. Die homomorphe Verschlüsselung muss nur dann aufgehoben werden, wenn die einzelnen Online-Stimmen unverschlüsselt vorliegen müssen (bspw. für die Nachvollziehbarkeit durch den Wähler wie in Abbildung 2 angegeben).

Dieses Beispiel stellt dabei nur einen Teil der Wahlhandlung dar. Bereits in der Wahlvorbereitung muss unterschiedliches Schlüsselmaterial genutzt werden.[10] Die Nutzung derselben Schlüssel für unterschiedliche Vorgänge mindert die Sicherheit und sollte daher vermieden werden. Das macht deutlich, dass das Thema Kryptografie im Bereich Online-Wahlen komplex ist.

Ein weiterer wichtiger Aspekt der Nutzung von kryptografischen Schlüsseln ist das Schlüsselmanagement. Es muss sowohl technisch als auch organisatorisch sichergestellt werden, dass niemand unberechtigten Zugang zu den privaten Schlüsseln, ob zum Entschlüsseln, Zeitstempeln oder Signieren, hat.

[10] Siehe auch BSI TR-03162 Kapitel 2.5

6. Zusammenfassung und Ausblick

Zusammenfassend stellen sowohl die Umsetzung der Wahlrechtsgrundsätze als auch die Absicherung der Wahl eine besondere Herausforderung dar. Dazu ist eine gute Vorarbeit und Vorbereitung der Wahlprozesse unersetzlich. Denn je nach Rahmenbedingungen zur Durchführung der Online-Wahl, Schutzbedarfsfeststellung und Risikoanalyse sind verschiedene Maßnahmen zu treffen. Jedoch sollten die unter Kapitel 5 genannten Maßnahmen immer umgesetzt werden. Nur eine sichere Wahl schafft Vertrauen beim Wähler. Aus diesem Grund ist Transparenz wichtig. Der Wähler sollte nachvollziehen können, wie das System funktioniert und gesichert ist.

Die Technische Richtlinie TR-03162 ergänzt die Online-Wahl-Verordnung[11] um die IT-sicherheitstechnischen Anforderungen, die über die allgemeinen Vorgaben der Wahlordnung für die Sozialversicherung (SVWO) hinaus für die Durchführung des Online-Wahlverfahrens erforderlich sind. Sie enthält Vorgaben für den Betrieb und die Nutzung von Anwendungen und IT-Systemen, die bei der Durchführung des Modellprojekts Online-Sozialwahl zum Einsatz kommen. Betrachtet werden dabei unter anderem die Einrichtung eines Informationssicherheitsmanagementsystems (ISMS) sowie aktuelle kryptografische Methoden zur Verschlüsselung, zur elektronischen Signatur und zum elektronischen Zeitstempel.

Online-Wahlen könnten zukünftig zudem durch den Einsatz mobiler Identitäten noch nutzerfreundlicher gestaltet werden. Mobile Identitäten sind als aktuelle Entwicklung auch Teil des 9-Punkte-Plans der Bundesregierung für ein digitales Deutschland[12]. Dabei werden die unverwechselbaren Attribute einer Identität aus einer sicheren Quelle wie dem Personalausweis abgeleitet, auf das mobile Gerät übertragen und an das Secure Element eines Mobilgerätes gebunden. Aktuelle Smartphones enthalten standardmäßig hardwarebasierte Sicherheitsanker (z.B. embedded Secure Elements oder eSIM), sodass höchste Sicherheitsstandards technisch umgesetzt werden können. In Kombination mit Apps auf dem Smartphone bietet diese Lösung eine attraktive Herangehensweise zur Durchführung von Online-Wahlen. Nutzerfreundlichkeit gepaart mit zuverlässiger Sicherheit kann zudem auch insbesondere bei jungen Wählern die Wahlbeteiligung erhöhen. Anders als in Abbildung 1 dargestellt, werden dann weder Ausweisdokument noch Lesegerät benötigt. Die sicherheitstechnischen Anforderungen zur Implementierung und zum Einsatz von Apps sowie die Anforderungen an mobile Identitäten sollten für dieses Einsatzszenario gesondert untersucht und bei Bedarf in einer Technischen Richtlinie aufgestellt werden.

[11] Verordnung über die technischen und organisatorischen Vorgaben für die Durchführung einer Online-Wahl im Rahmen des Modellprojekts nach § 194a des Fünften Buches Sozialgesetzbuch

[12] Richter, Dr. Markus, CIO Bund, 9-Punkte-Plan für ein digitales Deutschland, 2020

Literaturhinweise

[1] BSI IT-Grundschutz. 2020. Informationssicherheit und IT-Grundschutz. BSI-Standards 200-1, 200-2, 200-3. 2020.

[2] BSI, TR-03107 Elektronische Identitäten und Vertrauensdienste im E-Government, 7. Mai 2019, Version 1.1.1

[3] BSI, TR-03116 Kryptographische Vorgaben für Projekte der Bundesregierung. 10. Januar 2020

[4] BSI, TR-03162 IT-sicherheitstechnische Anforderungen zur Durchführung einer Online-Wahl im Rahmen des Modellprojektes nach § 194a Fünftes Buch Sozialgesetzbuch (Online-Wahl), 29. Dezember 2020, Version 1.0

[5] BVerfG, Urteil des Zweiten Senats vom 3. März 2009, 2 BvC 3/07, 2 BvC 4/07

[6] Richter, Dr. Markus, CIO Bund, 9-Punkte-Plan für ein digitales Deutschland, Juli 2020

[7] Verordnung (EU) Nr. 910/2014 über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt und zur Aufhebung der Richtlinie 1999/93/EG, 23. Juli 2014. https://eur-lex.europa.eu/ (eIDAS-Verordnung)

[8] Verordnung über die technischen und organisatorischen Vorgaben für die Durchführung einer Online-Wahl im Rahmen des Modellprojekts nach § 194a des Fünften Buches Sozialgesetzbuch (Online-Wahl-Verordnung), 30. September 2020

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Auf den Lime gegangen: Vor- und Nachteile der neuen Mikromobilität

Jan-Niclas Hilgert[1], Martin Lambertz[1]

Kurzfassung:
Elektrische Roller sind inzwischen ein etabliertes Fortbewegungsmittel in zahlreichen Städten. Wie bei jeder neu eingeführten Technologie sind auch hier verschiedene sicherheitsrelevante Aspekte zu betrachten. In diesem Beitrag geht es um die von den Anbietern bereitgestellten Apps zur Nutzung der Roller und der dort generierten Daten aus den Blickwinkeln der IT-Forensik, des Datenschutzes bzw. der schützenswerten Daten und der IT-Sicherheit. Zusätzlich werden auch die bei den Anbietern abrufbaren Daten miteinbezogen.
Im Rahmen der Untersuchungen wurde festgestellt, dass die generierten Daten grundsätzlich für die Forensik sehr relevant, jedoch aufgrund der Möglichkeit zur Fälschung kritisch zu betrachten sind. Ebenso kritisch zu betrachten sind die vielen auch Unbefugten zugänglichen Informationen, die sowohl das Tracken von Rollern als auch die Ermittlung der gesamten Rollerflotte einschließlich der Positionen zulassen. Letztlich wurden Möglichkeiten gefunden, mit denen Roller auch ohne Nutzung der offiziellen App und auch ohne Bezahlung geliehen werden konnten.

Stichworte: BLE (Bluetooth Low Energy), Datenschutz, Digitale Forensik, Elektroroller, eScooter, Mikromobilität, Roller, Scooter

1. Einleitung

Im Sommer 2019 begannen Firmen wie Lime, TIER oder VOI auch in Deutschland, mit ihren elektrischen Rollern – auch eScooter genannt – die Stadtbilder zu erweitern. Gedacht für „die letzten Meter“, sollen die bereitgestellten Gefährte den Menschen eine neue Form der Fortbewegung bieten. Als Mikromobilität bezeichnet das ivm RheinMain und das Fraunhofer IML dieses Konzept der „*Fortbewegung unter Nutzung von elektrisch angetriebenen Kleinst- und Leichtfahrzeugen*“[2]. Um eine möglichst breite Masse anzusprechen und den Komfort der Benutzung zu erhöhen, lassen sich diese Fahrzeuge auf einfachem Weg über eine Smartphone-App anmieten und an beliebigen Orten innerhalb fest definierter Zonen abstellen. Fast zwei Jahre später lässt sich der Erfolg dieses neuen Mobilitäts-Konzepts nicht nur anhand der Anzahl der anzutreffenden Roller auf Deutschlands Straßen messen. Das Berliner Start-Up TIER ist seit Sommer 2020 nach eigenen Angaben in jeder Stadt profitabel und nähert sich Ende 2020 einem geschätzten Wert von einer Milliarde US-Dollar.

eScooter haben ihren Platz im Bewusstsein der Menschen und damit ihre Daseinsberechtigung gefunden. Zum Glück für die darauf angewiesenen Pendler, zum Ärger für die Menschen, welche die neue Art der Fortbewegung in ihren Gärten oder vor der Einfahrt wiederfinden. Doch auch in einigen anderen Bereichen, gibt es solche Vor- und Nachteile, die diese neue Form der Mobilität mit sich bringt. Die Benutzung der Roller mittels App beispielsweise bietet nicht nur einen hohen Komfort für den Nutzer, sondern

[1] Fraunhofer FKIE, Zanderstraße 5, 53177 Bonn
[2] https://www.ivm-rheinmain.de/was-ist-mikromobilitaet/

erzeugt auch zwangsläufig eine Menge digitaler Daten auf den Endgeräten der Benutzer sowie den Servern der Anbieter. Zusätzlich bietet ein neues, vernetztes Produkt wie ein eScooter auch jenseits seiner erzeugten Datenmengen weitere Möglichkeiten für Sicherheitsforscher oder Kriminelle.

In dieser Arbeit wird näher auf die zuvor erwähnten Punkte eingegangen, indem die neue Mikromobilität aus drei Blickwinkeln betrachtet wird:

1. **Forensik:** Die digitale Forensik beschäftigt sich mit der Analyse digitaler Spuren zur Aufklärung von Verbrechen. Sie ist daran interessiert, welche forensisch relevanten Arten von Daten die Benutzung einer Mikromobilitäts-App erzeugt und wie ein Ermittler an diese Daten gelangen kann.
2. **Schützenswerte Daten**: Bewegungsdaten gehören mitunter zu den empfindlichsten Daten einer Person. Aus diesem Grund sollten die entsprechenden Daten geschützt und nicht von unberechtigten Personen einsehbar sein. Dies gilt zusätzlich auch für schützenswerte Daten der Anbieter.
3. **IT-Sicherheit:** Wie bei der Einführung eigentlich aller digitaler Produkte üblich, ist es notwendig, die neue Mikromobilität auch aus Sicht der IT-Sicherheit zu betrachten.

2. Blickwinkel der Forensik

Für den IT-Forensiker bietet die Mikromobilität und die damit verbundene Nutzung einer Smartphone-App eine neue Quelle forensisch interessanter Daten. Im Vordergrund stehen hierbei zweifellos die Bewegungsdaten eines Nutzers. Dennoch lassen sich auch andere relevante Informationen an verschiedenen Stellen finden. Die Speicherung dieser Daten ist nicht nur auf den lokalen Speicher des Smartphones begrenzt, sondern kann auch auf den Servern der Anbieter geschehen. Zusätzlich zu den neu verfügbaren Daten ist es notwendig, verschiedene forensische Überlegungen im Zusammenhang mit diesen Artefakten zu diskutieren und für folgende Analysen zu beachten.

2.1. Forensische Überlegungen

- **Nutzer-Daten:** Diese Daten umfassen alle Informationen, welche direkt mit dem Nutzer der Smartphone-App in Verbindung stehen. Hierbei kann es sich beispielsweise um Namen, Telefonnummern oder E-Mail-Adressen handeln. Wichtig ist es zu erwähnen, dass diese Daten in der Regel nicht überprüft werden und somit ihre Korrektheit nicht immer gegeben ist.
- **Zahlungs-Daten**: Da eine gültige Zahlungsmethode in fast allen Fällen zwingend erforderlich ist, um einen der Mikromobilitäts-Services zu nutzen, sind diese Daten in der Regel vorhanden. Bei Zahlungsdaten kann es sich je nach Anbieter beispielsweise um Kreditkartennummern oder PayPal-Daten handeln. Zusätzlich sind diese Informationen in den meisten Fällen verifiziert.
- **Fahrtverlauf:** Ein besonders kritisches Artefakt stellen die Daten über die getätigten Fahrten eines Nutzers dar. Zu ihnen gehören zeitliche Informationen wie

Start- und Endzeit einer Fahrt sowie auch geografische Informationen wie Koordinaten der gefahrenen Strecke. Diese können in verschiedener Häufigkeit gespeichert werden, zum Beispiel nur Start- und Endpunkt einer Fahrt oder auch eine komplette Übersicht der gefahrenen Strecke. Zusätzliche Informationen können die Länge einer Fahrt oder die Durchschnittsgeschwindigkeit sein.

- **Roller-Ausleihe:** Für eine Ausleihe lassen sich vier Schritte definieren.
  1. Im ersten Schritt verschafft der Nutzer sich eine **Übersicht** über verfügbare Roller in seiner Nähe. Werden diese Informationen auf dem Gerät zwischengespeichert, lassen sich hier gegebenenfalls Informationen über Aufenthaltsorte und Nutzungszeitpunkte der Smartphone-App extrahieren.
  2. Im nächsten Schritt wird der Roller **gemietet** und **entsperrt**. Bei manchen Anbietern ist es notwendig, einen QR-Code oder Ähnliches auf dem Roller zu scannen, um zu gewährleisten, dass sich der Nutzer auch in dessen Nähe aufhält. Andere Anbieter wiederum fordern nur die Eingabe einer ID, wodurch sich Fahrten von beliebigen Orten ohne die physische Präsenz des Nutzers starten lassen.
  3. Während der **Fahrt** des Users werden unter Umständen Informationen über die Position des Rollers oder des Smartphones oder die Geschwindigkeit lokal oder auf externen Servern gespeichert.
  4. Das **Beenden** der Fahrt geschieht in der Regel mittels der zugehörigen App. Dennoch kann unter Umständen eine physische Präsenz erforderlich sein, beispielsweise zum Schließen eines gegebenenfalls verfügbaren Schlosses.
- **Vertrauenswürdigkeit:** Die Vertrauenswürdigkeit der extrahierten Daten spielt eine wichtige Rolle während forensischer Untersuchungen. Hier gilt es zu bewerten, welchem Ursprung gefundene Daten wie beispielsweise gefahrene Strecken entstammen und ob sie auf einfache Art und Weise manipuliert werden können.

2.2. Daten-Zugriff

Für den Zugriff auf die im vorherigen Abschnitt beschriebenen Informationen bieten sich im Rahmen der IT-Forensik zwei verschiedene Arten der Datenquellen: lokale sowie entfernte, also auf externen Servern, gespeicherte Daten. Lokale Daten lassen sich beispielsweise über das Öffnen der Applikation abrufen und einsehen. Diese Art der Datensicherung kann jedoch unerwünscht sein, da sie zum einen nicht zwangsläufig alle verfügbaren Daten beinhalten muss und zum anderen auch eine Veränderung der Beweismittel zur Folge haben kann. Eine weitere Methode zur Extraktion lokaler Daten ist der direkte Zugriff auf die Datenbanken einer Smartphone-Applikation ohne Benutzung der Anwendung selbst. Dieser erfordert jedoch unter Umständen privilegierten Zugriff auf das gesamte Gerät, beispielsweise in Form eines Jailbreaks. Auch dies kann während einer Untersuchung unerwünscht oder auch einfach technisch nicht möglich sein. Aus

diesen Gründen haben die entfernten Daten und ihre Extraktion ohne Zuhilfenahme der entsprechenden App eine hohe Bedeutung innerhalb der Forensik.

Für die weitere Analyse ist es notwendig zu verstehen, wie auf diese Daten zugegriffen werden kann bzw. an welcher Stelle sie zu finden sind, welchen Informationsgehalt sie bieten und wie vertrauenswürdig sie sind. Im Folgenden werden diese Fragen anhand gewonnener Forschungsergebnisse am Beispiel des Anbieters Lime und der Benutzung eines iOS-Devices näher erläutert. Zusätzlich wurde ein Framework entwickelt, was die Extraktion, Analyse und Darstellung der Daten verschiedener Anbieter ermöglicht.

2.3. Analyse: Lime

Für unsere Betrachtung wurde das vorliegende iPhone durch einen Jailbreak für die weitere Analyse vorbereitet. Anschließend wurden die Daten der Lime-Applikation extrahiert. Für den Zugriff auf nutzerspezifische Daten über die API ist ein Bearer-Token zur Authentifizierung erforderlich. Auf dem untersuchten Gerät selbst lässt sich das entsprechende Nutzer-Token nach dem Jailbreak aus der `keychain-2.db`, dem iOS-Schlüsselbund, extrahieren. Für den Fall, dass kein privilegierter Zugriff auf das Gerät bzw. den Schlüsselbund möglich ist, kann ein Token auch neu generiert werden. Hierdurch verlieren alte Token ihre Gültigkeit.

Lime unterstützt unter anderem die Registrierung per Telefonnummer sowie E-Mail-Adresse. Durch den Zugriff auf die Kurznachrichten bzw. das E-Mail-Konto eines Nutzers ist es somit möglich, ein neues Zugangs-Token über die Lime-API zu generieren. Hiermit kann anschließend auf die API zugegriffen werden.

2.3.1 Nutzer-Daten

Unter iOS nutzt Lime die `LimeBike.sqlite`-Datenbank zur Speicherung der nutzerrelevanten Daten. Diese lassen sich innerhalb der Datenbank in der `ZPSUSER`-Tabelle finden und beinhalten Informationen wie den Benutzernamen (Default: „Lime Rider“), die Telefonnummer, die E-Mail-Adresse und das User-Token.

Über den entsprechenden API-Endpunkt lassen sich dieselben Informationen anfragen.

2.3.2. Zahlungs-Daten

Lime unterstützt zur Zahlung PayPal, Apple Pay sowie Kreditkarten. Hinterlegte Methoden lassen sich in der zuvor erwähnten Datenbank in der Tabelle `ZPSPAYMENTMETHOD` finden, wobei jede Zahlungsmethode einem Eintrag entspricht. Für Kreditkarten beinhaltet ein solcher Eintrag die letzten vier Ziffern sowie das Jahr und den Monat des Ablaufdatums, wohingegen für PayPal nur der Typ der Zahlungsmethode gespeichert wird. In den durchgeführten Tests konnten ausschließlich in der lokalen Datenbank auch bereits gelöschte Zahlungsmethoden gefunden werden.

Über die API lassen sich wieder dieselben Informationen abrufen, lediglich für PayPal enthalten die API-Daten auch die zugeordnete E-Mail-Adresse.

2.3.3. Fahrtdaten

Lokal lassen sich Informationen über die getätigten Fahrten eines Nutzers in der `ZPSTRIPS`-Tabelle innerhalb der Lime-Datenbank finden. Diese sind jedoch sehr limitiert und beinhalten nur den Start- und Endzeitpunkt und einen Identifier der entsprechenden Fahrt.

Weitaus detailliertere Informationen können über die API abgerufen werden. Diese beinhalten neben zeitlichen Daten wie Start- und Endzeitpunkt sowie der Gesamtlänge in Minuten auch geografische Daten. Diese werden jedoch nicht direkt als Koordinaten, sondern in Form einer Google Polyline sowie eines PNG-Bildes des gesamten Fahrverlaufs, wie in der folgenden Abbildung dargestellt, abgespeichert. Auch lokal lassen sich diese PNG-Bilder auf dem Gerät wiederfinden.

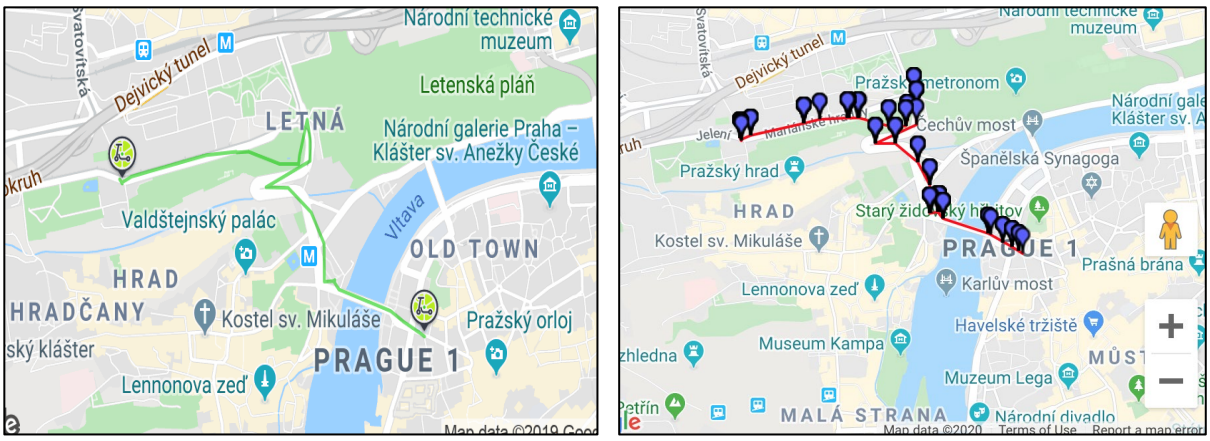


Abb. 1: Fahrtverlauf durch Prag als gerendertes PNG-Bild (links) sowie dekodierte Polyline (rechts)

Da sich anhand eines solchen Verlaufs jedoch der Start- und Endpunkt der Fahrt nicht identifizieren lässt, ist es wünschenswert die erzeugte Google Polyline zu dekodieren[3]. Da diese aus mehreren Koordinaten besteht, lässt sich auf diese Art und Weise der genaue Verlauf sowie anhand ihrer Reihenfolge Start und Ende einer Fahrt bestimmen. Ein Beispiel einer dekodierten Polyline findet sich ebenfalls in der vorherigen Abbildung (hier ohne Darstellung der Reihenfolge).

Zu den weiteren Informationen gehören die Kosten der getätigten Fahrt und die letzten drei Zeichen des „Nummernschilds“ des zugehörigen Rollers. Dieses ist ein sechsstelliger, alphanumerischer Wert und muss nicht mit dem tatsächlichen, physischen Nummernschild des Rollers übereinstimmen.

2.3.4. Roller-Ausleihe

Um eine aktuelle Übersicht der verfügbaren Roller zu erhalten, ist es zwangsläufig notwendig die API zu verwenden. Hier werden für jeden Roller in der Nähe des angefragten

[3] https://developers.google.com/maps/documentation/utilities/polylineutility?hl=de

Standorts die aktuelle Position, die Reichweite sowie die letzten drei Zeichen des Nummernschilds übertragen. Innerhalb der lokalen Lime-Datenbank lässt sich eine Auflistung der bisher gemieteten Roller finden.

Für das Ausleihen eines Rollers und das Starten einer Fahrt ist entweder das Scannen eines QR-Codes auf dem Lenker oder das Nummernschild des Rollers erforderlich. Da beide Informationen nicht oder nicht vollständig in der API-Antwort enthalten sind, ist es auch nicht möglich eine Fahrt ohne physische Präsenz zu starten. Da sich jedoch QR-Code und Nummernschild in der Regel über einen längeren Zeitraum nicht ändern, ist es möglich diese zu speichern und zu einem späteren Zeitpunkt wiederzuverwenden.

Während der Fahrt werden entsprechende Informationen wie die Position des Rollers an die Lime-Server gesendet und lassen sich auch über die API abrufen. Näheres hierzu findet sich am Beispiel von Tier im folgenden Kapitel zum Thema Privatsphäre und Datenschutz.

Das Beenden einer Fahrt erfordert keine Interaktion mit dem Gerät und lässt sich somit ohne physische Präsenz von jedem Ort bewerkstelligen. Nach der Fahrt hat der Benutzer die optionale Möglichkeit ein Foto des geparkten Rollers hochzuladen. Auch wenn der Pfad zu dem gespeicherten Bild in den extrahierten Daten nicht gefunden werden konnte, sind diese Informationen gegebenenfalls auf dem Server des Anbieters vorhanden.

2.3.5. Vertrauenswürdigkeit

Fahrtdaten stellen eine der wichtigsten Informationsquellen für den Forensiker dar. Gerade deshalb ist es essenziell ihre Vertrauenswürdigkeit genauer zu untersuchen. Anders formulieren lässt sich die Kernfrage als „Wie wahrscheinlich ist es, dass die gespeicherten Daten über die Fahrt eines Nutzers korrekt und nicht gefälscht sind?“.

Wie bereits im vorausgehenden Abschnitt erwähnt, sendet der Lime-Roller während der Fahrt kontinuierlich seine GPS-Position an die Lime-Server. Während der durchgeführten Tests fiel auf, dass auch das Smartphone des Nutzers in einigen Fällen Koordinaten an die Server sendete. Diese Koordinaten werden genutzt, um die zuvor erwähnte Polyline sowie das erzeugte PNG-Bild des Fahrtverlaufs zu generieren. Durch das falsche Senden dieser Datenpakete ist es nun möglich, den dargestellten Verlauf einer Fahrt beliebig zu ändern. Dies funktioniert auch für den Start- und Endpunkt, wodurch ein Roller aus Auckland auch für eine gefälschte Fahrt in Berlin genutzt werden kann. Der gesamte Fahrtverlauf enthält somit keine Informationen mehr über die eigentlich gefahrene Strecke. Abbildung 2 zeigt ein Beispiel für eine Weltreise für 1,20€.

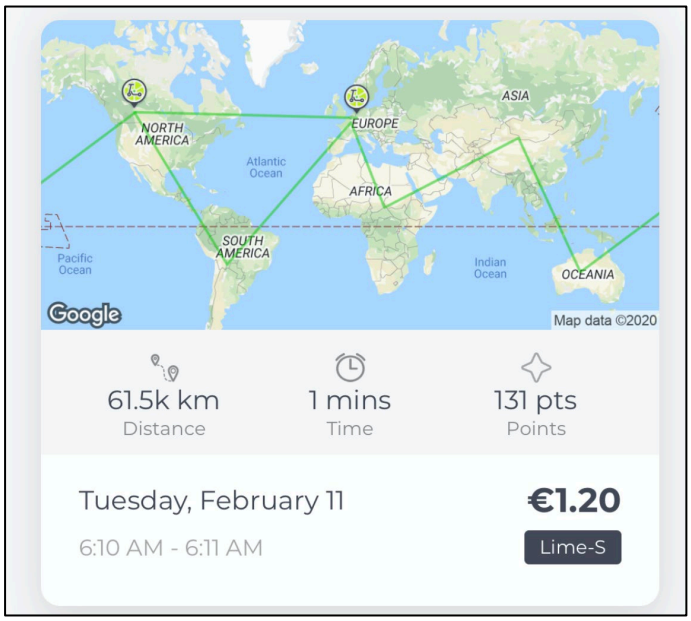


Abb. 2: Roller-Fahrt mit Lime um die ganze Welt

3. Blickwinkel der Privatsphäre und der schützenswerten Daten

Für den IT-Forensiker nehmen Bewegungsdaten eines Nutzers einen hohen Stellenwert ein, da diese einen Einblick in das Leben einer Person in Form von besuchten Orten sowie Gewohnheiten geben können. Aus eben diesem Grund ist es von höchster Wichtigkeit diese kritischen Daten vor unbefugten Zugriffen zu schützen. Doch nicht nur im Blick auf Privatpersonen fallen schützenswerte Daten an. Auch auf Seite der Anbieter ist es unter Umständen nicht wünschenswert, Informationen über die gesamte Flotte oder mögliche Wartungsroutinen preiszugeben. Dieser Blickwinkel befasst sich mit dem Thema des Datenschutzes am Beispiel von Tier.

Über die Nutzung der Tier-API besteht die Möglichkeit, detaillierte Informationen über Roller ihrer Flotte, identifiziert über eine sechsstellige Zahl, anzufragen. Diese Informationen beinhalten zum Beispiel den Batteriestand, den aktuellen Status aber auch die GPS-Koordinaten. Führt man diese Abfragen nun wiederholt für alle Zahlen bis 999999 durch, erlangt man innerhalb weniger Stunden einen Datensatz der aktuellen Tier-Roller-Flotte. Die folgende Tabelle zeigt die Anzahl ausgewählter `state`-Felder aus zwei ermittelten Datensätzen aus Frühjahr und Winter 2020.

`state`	Februar 2020	Dezember 2020
Active	23977	56137
Decomissioned	13827	17295
Out of order	3934	10197
Lost	3554	5740
Stolen	2781	2687
Damaged	124	1214

Auch wenn diese Arbeit keine Interpretation der gesammelten Werte gibt, so lässt sich an einigen Stellen ein deutlicher Trend erkennen. Es ist jedoch wichtig an dieser Stelle zu betonen, dass die definierten „states“ aus einer solchen Abfrage sowie ihr Setzen komplett in der Hand des Anbieters liegen. Dies gilt auch für alle weiteren Werte.

3.1. Tracking

Der Zugriff auf beendete Fahrten eines Nutzers über die API eines Anbieters wurde im vorherigen Kapitel bereits thematisiert. An dieser Stelle geht es um die Koordinaten, welche von einem Roller selbst kontinuierlich an die Server der Anbieter gesendet werden und sich über die Tier-API abfragen lassen. Diese Koordinaten werden fortwährend – auch während einer Fahrt – mit den aktuellen Werten des GPS-Moduls des Rollers aktualisiert. Der Zugriff auf diese Informationen ist nicht eingeschränkt und lässt sich jederzeit und von jedem durchführen. Kombiniert führen diese beiden Tatsachen zu der Möglichkeit, die Fahrt einer fremden Person in Echtzeit zu verfolgen. Lediglich die Roller-ID wird an dieser Stelle benötigt. Mit einem Datensatz, der ein vorbeifahrendes Nummernschild dem entsprechenden Roller-Code zuordnen kann, lassen sich so angetroffene Tier-Nutzer problemlos verfolgen. Die nachfolgende Abbildung zeigt ein Beispiel einer solchen aufgezeichneten Fahrt des Rollers.

Tier ist zudem kein Einzelfall, was diese Möglichkeit angeht. Auch bei Lime können die korrekten GPS-Koordinaten eines Rollers während der Fahrt abgerufen werden, selbst wenn andere gefälschte Daten, wie in Abschnitt 2.3.5 beschrieben, gesendet werden. Auch bei Voi war es möglich, Fahrten zu verfolgen.

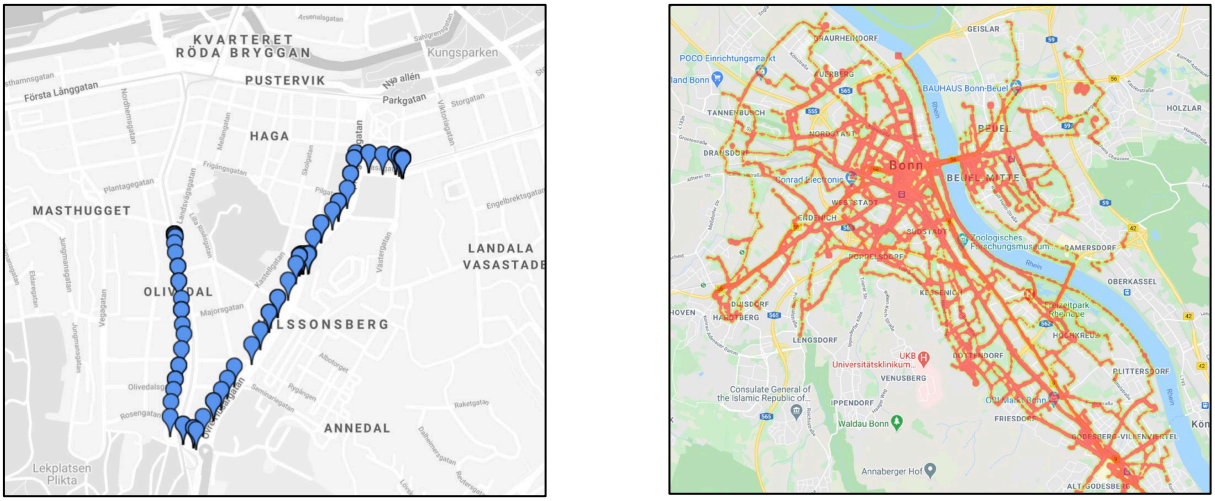


Abb. 3: Verlauf einer Tier-Fahrt (links) sowie eine Heatmap aller Positionen fahrender Tier-Roller über 12 Stunden (rechts)

3.2. Bewegungsprofil

Wendet man die zuvor beschriebene Tracking-Möglichkeit nicht nur auf einzelne Roller, sondern eine ganze Roller-Flotte innerhalb einer Stadt an, so lässt sich problemlos ein Bewegungsprofil der gesamten Stadt erstellen. Die auf diese Weise gesammelten Daten können auf unterschiedliche Arten verwendet werden. Ein Beispiel für ihre Verwendung ist in Abbildung 3 dargestellt. Innerhalb eines Zeitraums von zwölf Stunden

wurden die Koordinaten aller Roller während einer Fahrt gesammelt und zur Veranschaulichung mittels Heatmap dargestellt. Hiermit lässt sich beispielsweise einsehen, welche Straßen und Gegenden von fahrenden Rollern besonders frequentiert sind. Solche Informationen können natürlich auch von Mitbewerbern genutzt werden. Weitere Beispiele für gewonnene Daten sind zum Beispiel die Anzahl der Fahrten, ihre durchschnittliche Dauer oder die allgemeine Auslastung der Flotte.

4. Blickwinkel der IT-Sicherheit

Zuletzt lohnt es sich, die Mikromobilität aufgrund ihrer Anbindung über und der Abhängigkeit vom Internet sowie der Nutzung neuer IoT-Geräte auch aus Sicht der IT-Sicherheit zu betrachten. Exemplarisch werden im Folgenden zwei Beispiele aus diesem Bereich diskutiert.

4.1. API-Zugriff

Sowohl für die IT-Forensik als auch die Privatsphäre und den Datenschutz spielt die API eines Mikromobilitäts-Anbieters eine kritische Rolle. Auf der einen Seite ist es hilfreich, wenn sie dem berechtigten Nutzer oder Analysten relevante Daten wie Fahrtverläufe detailliert zur Verfügung stellt. Gleichzeitig ist es aber wichtig, dass eben jene Daten vor fremden Zugriffen geschützt sind, so dass eine Verletzung der Privatsphäre ausgeschlossen und Datenschutz gewährleistet ist. In dieser Arbeit wurden verschiedene Punkte identifiziert, die im Rahmen einer verbesserten IT-Sicherheit adressiert werden sollten. Beispielhaft seien hier aufgeführt:

- **Rate-Limitierung für API-Anfragen:** Wie zuvor beschrieben, war es möglich, alle Scooter des Anbieters Tier zu sammeln und die entsprechenden Roller-Informationen zu speichern. Dies gelang innerhalb kürzester Zeit durch wiederholte API-Anfragen. An dieser Stelle wäre eine erste geeignete Schutzmaßnahme, die Anzahl der möglichen API-Zugriffe durch Rate-Limitierung zu begrenzen. Mit Hilfe dieses Konzepts, welches beispielsweise von Lime eingesetzt wird, ist es nicht mehr ohne Weiteres möglich, große Datenmengen von einem Anbieter zu extrahieren.
- **Datensparsamkeit:** Neben einer Limitierung der Anfragen ist es auch wünschenswert, den allgemeinen Zugriff auf kritische Daten einzuschränken. Hierbei handelt es sich beispielsweise um Echtzeit-GPS-Daten, welche das Verfolgen eines Rollers während der Fahrt ermöglichen. Diese Datensparsamkeit kann entweder durch Zugriffsbeschränkungen der API-Endpunkte oder aber durch einfaches Entfernen der Attribute erreicht werden.

  Ein weiteres Beispiel der Datensparsamkeit betrifft die Informationen eines Rollers, welche zum Starten einer Fahrt benötigt werden. Werden diese nicht vollständig übermittelt, so lässt sich die Fahrt nur mit physischer Präsenz oder Kenntnis der Gesamtinformation starten. Der Anbieter VOI zeigt dem Nutzer der App nur einen Teil des Roller-Codes an, welcher zur Anmietung benötigt wird. Jedoch lieferte die entsprechende API die vollständigen Codes zurück, welche lediglich Client-seitig ausgeblendet wurden.

- **Kontoeröffnung:** Es sollte sichergestellt werden, dass die Möglichkeit, massenhaft Konten zu eröffnen erschwert wird oder nicht gegeben ist. Durch den Einsatz mehrerer Wegwerf-E-Mail-Adressen ließen sich bei dem Anbieter Spin problemlos und in kürzester Zeit eine Vielzahl von Accounts erstellen. Diese können beispielsweise genutzt werden um Gutscheine für Freifahrten vielfach zu nutzen oder auch um Freifahrten durch Werben neuer Nutzern zu erhalten.

4.2. Bluetooth Low Energy

Durch den Einsatz neuartiger IoT-Module entstehen weitere Ansatzpunkte für eine genauere Betrachtung im Rahmen der IT-Sicherheit. Eine weitverbreitete Technik der letzten Jahre, welche auch in eScootern Verwendung findet, ist Bluetooth Low Energy, kurz BLE. Wie der Name bereits andeutet, unterscheidet sich BLE von dem herkömmlichen Bluetooth-Standard in seinem Energieverbrauch. Ein weiterer Unterschied besteht in der geringen Datenmenge, welche üblicherweise über BLE ausgetauscht wird.

In eScootern lässt sich BLE nicht nur zur Abfrage von Werten wie Reichweite oder Batterieladung nutzen, sondern unter Umständen auch zum Sperren und Entsperren des Rollers. Diese Art der Zugriffskontrolle ist durchaus sinnvoll, da viele Rollermodelle auch auf dem freien Markt verfügbar sind und somit von Privatpersonen genutzt werden, denen nicht zwangsläufig eine Cloud-Infrastruktur zum Starten ihrer Fahrt zur Verfügung steht. Außerdem lassen sich über diese Schnittstelle auch unabhängig von einer API die Roller eines Anbieters entsperren, beispielsweise für Wartungsarbeiten. Dieser Umstand macht die BLE-Funktionalität eines eScooters zu einem besonders interessanten Ziel für Angreifer.

4.2.1. Fallbeispiel: Spin

Im Rahmen dieser Arbeit wurden auch BLE-Schnittstellen von Rollern verbreiteter Anbieter untersucht. Am Beispiel von Spin soll hier nun kurz beschrieben werden, wie Angreifer unberechtigten Zugriff über diesen Angriffsvektor erhalten können.

In der Region Köln/Bonn sowie weiteren Standorten setzt Spin das Modell Segway Ninebot Max ein, welches neben einer Anbindung an das Mobilfunknetz auch über eine BLE-Schnittstelle verfügt. Mit Segway Commercial (ehemals Segway Discovery) bietet Segway eine „ready-made scooter rental management solution"[4]. Online lassen sich Informationen über die entsprechende Gateway-API des Dienstes finden, mit deren Hilfe Kunden ihre Flotte verwalten können. Eine weitere Möglichkeit zur Kontrolle der Roller bietet BLE, wie der dargestellte Auszug eines API-Dokuments von Segway zeigt. Hieraus wird auch ersichtlich, dass sich mittels BLE-Kontrolle und vorhandenem „BLE secret key" ein Roller sperren sowie entsperren lässt.

Über die BLE-API der Segway-Roller existieren keine frei zugänglichen Informationen. Auch durch eine Analyse der dekompilierten Spin-Applikation wurden keine brauchbaren Anhaltspunkte einer BLE-Kommunikation gefunden. Dies war zu erwarten, da die

[4] https://b2b.segway.com/about/

Steuerung des Spin-Rollers über die Smartphone-App und die Spin-Infrastruktur abläuft und nicht direkt stattfindet.

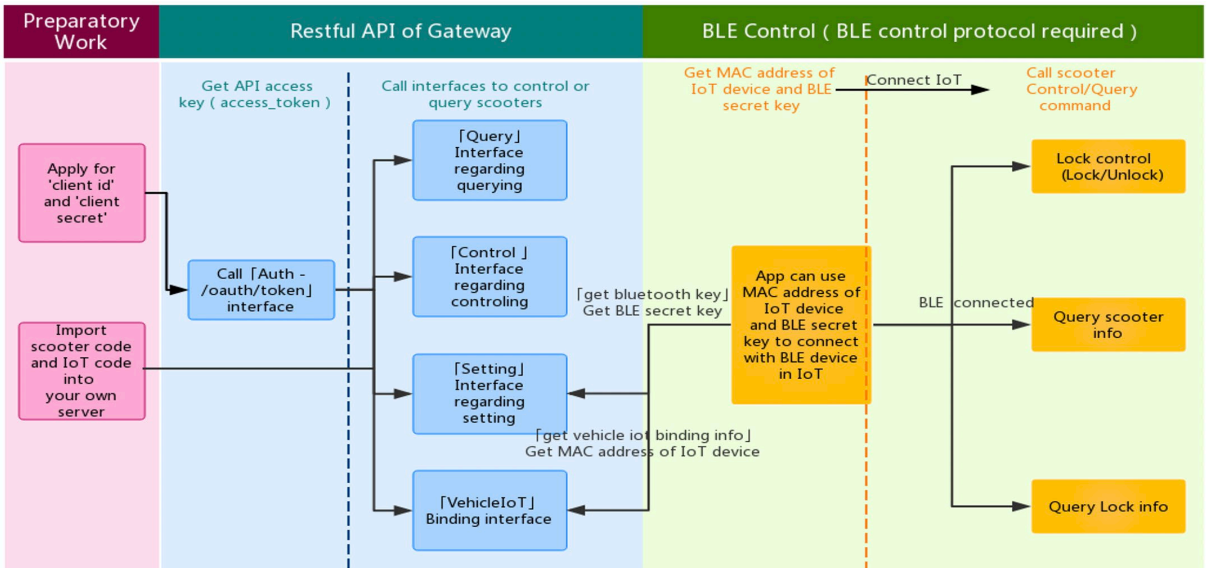


Abb. 4: Überblick über die Gateway-API und BLE-Kontrolle der Segway-Roller[5]

Für den weiteren Verlauf wurde das Reverse-Engineering auf andere Apps ausgeweitet. Da Segway Commercial von verschiedenen Kunden genutzt wird, lassen sich mehrere Segway-Apps, beispielsweise im Google Play Store, finden. Manche dieser Anwendungen bieten die Möglichkeit, die Sperre der Roller über BLE zu steuern. Mit Hilfe der somit gewonnenen Erkenntnisse konnte die Logik der BLE-Kommunikation nachvollzogen, reimplementiert und schließlich an den Spin-Rollern getestet werden. Für die entsprechenden Befehle zum Entsperren und Sperren der Roller wird der zuvor erwähnte Bluetooth-Key des Rollers benötigt. Dieser ließ sich über die frei zugängliche Spin-API problemlos für jeden Roller abrufen (Thema: Datensparsamkeit).

Dies bietet einem Angreifer die Möglichkeit, einen Spin-Roller zu entsperren und wieder zu sperren ohne Daten über die Spin-Infrastruktur austauschen zu müssen, wodurch Fahrten ohne Zahlung der eigentlich anfallenden Mietgebühr durchgeführt werden können. Zusätzlich gelang es, einen Spin-Roller auch zu sperren während er von einem legitimen Nutzer gemietet war. Dies kann unter Umständen zu weitaus drastischeren Folgen als den wirtschaftlichen Konsequenzen durch Freifahrten führen.

5. Zusammenfassung und Ausblick

Die vorliegende Arbeit hat anhand verschiedener Aspekte und Beispiele deutlich gemacht, dass die Einführung und Verbreitung von elektrischen Rollern, wie von anderen neu eingeführten Technologien bekannt, Vor- und Nachteile mit sich bringt. Es wurde gezeigt, dass die Benutzung der Roller sowohl auf den Smartphones der Nutzer als auch bei den Dienstanbietern eine Reihe von Daten erzeugt, die für die Ermittlungsbehörden wertvolle Informationen zu den Bewegungen involvierter Personen liefern.

[5] https://us-api.segway.pt/doc/extra/Api%20Additional%20remarks.docx

Das Vorhandensein der Daten ist gleichzeitig aber auch ein aufgedeckter Nachteil der aktuell im Einsatz befindlichen Systeme. Hier werden oft sehr viele schützenswerte Informationen, auch für Unbefugte zugänglich, bereitgestellt. Dadurch lassen sich beispielsweise Fahrtstrecken in Echtzeit nachvollziehen oder Betriebsinterna, wie Informationen über die gesamte Flotte, sammeln.

Daneben sind noch weitere aus Sicht der IT-Sicherheit problematische Aspekte aufgedeckt worden. Neben den Ursachen für die Möglichkeit, die genannten Daten in beträchtlicher Menge sammeln zu können, wurde ebenfalls gezeigt, wie ein unbedachter Umgang mit schützenswerten Daten zu direkten wirtschaftlichen Schäden bei den Anbietern führen kann, indem ihre Roller ohne Bezahlung gemietet werden können. Zusätzlich birgt die Möglichkeit, Spin-Fahrten legitimer Nutzer in der Nähe zu beenden, weitere Risiken

Da sich die APIs der vorgestellten Anbieter auch schon während dieser Arbeit laufend verändert haben, ist nicht gewährleistet, dass alle vorgestellten Methoden noch funktionieren. Auf der anderen Seite können Änderungen aber unter Umständen auch zu neuen Artefakten, ungeschützten Daten oder Sicherheitslücken führen. Dies gilt es in der Zukunft zu überprüfen.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Benutzerfreundliche Schutzmechanismen gegen USB-basierte Angriffe unter Linux

Timo Pohl[1], Arnold Sykosch[1]

Kurzfassung:
Die meisten Computer verfügen heutzutage über mindestens einen USB-Anschluss. Diese Schnittstelle ist durch ihre Flexibilität die am weitesten verbreitete Anschlussform für externe Geräte. Ihre Nutzung birgt jedoch auch Gefahren. Eine Lösung für dieses Problem hat in moderne Desktop Computer noch keinen Einzug gefunden. Bisherige Ansätze zum Schutz vor USB-basierten Angriffen setzen entweder tiefgehendes Wissen über USB voraus, schützen nur vor einem bestimmten Angriff oder schützen nur vor Angriffen durch Geräte, welche von dritten Personen angeschlossen wurden. Somit bieten sie entweder ein sehr niedriges Schutzniveau oder sind für viele Personengruppen nicht benutzbar. In dieser Arbeit wird ein Konzept vorgestellt, welches den Kontext, in dem ein USB-Gerät angeschlossen wird, sowie die Verifikation der erwarteten Fähigkeiten des USB-Geräts durch den Benutzer zur Einschätzung der potenziellen Maliziösität des USB-Geräts nutzt. Ein vorrangiges Designziel ist hierbei die Benutzerfreundlichkeit. Um diese zu bestimmen, wird ein Experiment durchgeführt, für welches ein Prototyp des Konzepts für die Linux-Desktopumgebung „GNOME Shell“ entwickelt wurde. Es hat sich gezeigt, dass die Benutzer im Durchschnitt mit dem Prototypen zufrieden waren, aber trotzdem eine durchschnittliche Fehlerrate von 25 % bestand. Auf zwei Drittel der Teilnehmer des Experiments konnte ein erfolgreicher Angriff mit einem USB-Gerät durchgeführt werden.

Stichworte: Bad USB, Desktop Linux, Open Source, USB-basierte Angriffe

1. Einleitung

Die meisten Computer besitzen heutzutage einen Anschluss in Form des *Universal Serial Bus* (USB). Der USB dient als Schnittstelle zu einer Vielzahl von Geräten, welche mit dem Computer kommunizieren.

Trotz der in jüngerer Vergangenheit aufgekommenen Zahl von unterschiedlichen Angriffen über den USB wird USB-Geräten im Allgemeinen vertraut. Menschen denken sich meist nichts Böses, wenn sie fremde USB-Geräte an ihre Computer anschließen, und auch Computer geben USB-Geräten sehr weitgehende Kontrolle, beispielsweise über Eingaben oder Netzwerkkommunikation. Besonders anschaulich ist dies bei Eingabegeräten wie Mäusen oder Tastaturen. Mit diesen Eingabegeräten steuern die meisten Menschen ihre Computer. Für den Computer sind sie die direkte Schnittstelle zum Benutzer und ihre Eingaben gelten als Repräsentation für die Intention des Benutzers. Es wird allerdings nie validiert, ob ein USB-Gerät nun wirklich nur Befehle abgibt, welche die Intention des Benutzers widerspiegeln oder diese vom USB-Gerät selbstständig abgegeben werden. Diese fehlende Validierung wird von maliziösen USB-Geräten ausgenutzt, um den Computer dazu zu bringen, schädliche Befehle auszuführen.

Zwar existieren bereits Ansätze, solche Angriffe zu unterbinden, diese schützen aber oft nur gegen spezifische Angriffe oder setzen umfangreiches Wissen über den USB voraus,

[1] Rheinische Friedrich-Wilhelms-Universität Bonn

und sind daher für die meisten Personen nicht nutzbar. Ein benutzerfreundlicher Schutz gegen eine möglichst große Bandbreite dieser Angriffe ist wünschenswert.

Im vergangenen Jahr haben Mitglieder der GNOME Foundation einen Ansatz vorgestellt, welcher durch Unterscheidung von Szenarien, ob ein legitimer Benutzer anwesend ist oder nicht, einen benutzerfreundlichen Schutz bietet. Dieser Ansatz bietet allerdings nur Schutz gegen Angriffe, die erfolgen, während der Benutzer nicht am Computer anwesend ist.

Dieser Ansatz wird im Rahmen dieser Arbeit erweitert. Zum einen wird die Ableitung des Kontextes raffiniert und zum anderen soll auch in Anwesenheit des Benutzers Schutz gegen Angriffe geboten werden. Da der Benutzer in diesem Kontext anwesend ist, wird zur Verhinderung von Angriffen auf zusätzliche Verifikation durch Interaktion mit dem Benutzer gesetzt.

Um herauszufinden, ob mit diesem Ansatz ein benutzerfreundlicher Schutz vor USB-basierten Angriffen realisiert werden kann, wird zunächst ein Konzept basierend auf dem beschriebenen Ansatz entwickelt. Anschließend wird eine Software, der *USB Protector*, implementiert, welche dieses Konzept umsetzt. In einem Experiment wird die Benutzerfreundlichkeit dieser Implementierung überprüft.

Die Desktopumgebung *GNOME Shell* der GNOME Foundation bietet sich für die Implementierung an, da die GNOME Foundation Interesse an der Umsetzung eines benutzerfreundlichen Schutzes gegen USB-basierte Angriffe gezeigt hat. Außerdem ist die *GNOME Shell* eine weit verbreitete Desktopumgebung für Linux. Daher wird hier eine Chance gesehen, dass die Ergebnisse dieser Arbeit eine Auswirkung auf ein weit verbreitetes System haben.

2. Grundlagen

Zum Verständnis der Arbeit werden in diesem Kapitel die Technologien und Konzepte erläutert, welche im weiteren Verlauf der Arbeit betrachtet werden. Zunächst werden die relevanten Aspekte des USB erläutert und anschließend die aktuell bekannten USB-basierten Angriffe beschrieben.

2.1. Grundlagen USB

Auf logischer Ebene sind USB-Geräte in verschiedene Schichten aufgeteilt, welche durch Deskriptoren, Datenstrukturen mit Informationen über das USB-Gerät, beschrieben werden. Ein USB-Gerät hat genau einen Geräte-Deskriptor, welcher Informationen über das USB-Gerät enthält [14]. Die relevanten Felder des Geräte-Deskriptors sind in Tabelle 1 dargestellt.

Unter dem Geräte-Deskriptor liegen ein oder mehrere Konfigurations-Deskriptoren, welche die Konfigurationen des Geräts repräsentieren [14]. Ein Gerät kann in verschiedenen Betriebsmodi verschiedene Konfigurationen nutzen, in denen beispielsweise unterschiedlich viel Strom vom USB-Gerät verbraucht wird. Die genaue Belegung der Felder ist für diese Arbeit nicht relevant.

Feld Name	Beschreibung
bDeviceClass	Geräte-Klasse
bDeviceSubClass	Geräte-Subklasse
bDeviceProtocol	Geräte-Protokoll
idVendor	Eindeutige ID des Herstellers
idProduct	In Kombination mit *idVendor* eindeutige ID des Produkts
bcdDevice	Version des USB-Geräts

Tabelle 1: Relevante Einträge des Geräte-Deskriptors. Die Geräte-Klasse, Geräte-Subklasse, und das Geräte-Protokoll sind Zahlenwerte, welche vom *USB Implementers Forum*, einer Organisation, welche von den Entwicklerfirmen des USB gegründet wurde, bestimmten Fähigkeiten zugeordnet sind [3].

Jeder Konfigurations-Deskriptor besitzt ein bis mehrere Interface-Deskriptoren. Diese repräsentieren jeweils eine Funktion des USB-Geräts. Funktionen eines USB-Geräts sind beispielsweise „Maus“, „Tastatur“ oder „Massenspeicher“. Die relevanten Einträge des Interface-Deskriptors sind in Tabelle 2 dargestellt.

Feld Name	Beschreibung
bInterfaceClass	Interface-Klasse
bInterfaceSubClass	Interface-Subklasse
bInterfaceProtocol	Interface-Protokoll

Tabelle 2: Für diese Arbeit relevante Einträge im Interface-Deskriptor. Die Interface-Klasse, Interface-Subklasse und das Interface-Protokoll sind Zahlenwerte, welche vom *USB Implementers Forum* bestimmten Fähigkeiten zugeordnet sind [3]. Diese Zuordnung ist gleich der Zuordnung der jeweils analogen Werte im Geräte-Deskriptor.

Wird ein USB-Gerät an einen Linux Computer angeschlossen, betrachtet der Kernel die Felder *idVendor, idProduct, bcdDevice, bDeviceClass, bDeviceSubClass, bDeviceProtocol, bInterfaceClass, bInterfaceSubClass* sowie *bInterfaceProtocol* der Deskriptoren und sucht passende Treiber, welche dann an die Interfaces des USB-Geräts gebunden werden [12]. Ein Treiber ist ein Programm, welches dazu dient, die Signale eines Geräts zu verarbeiten, und beispielsweise bei Bewegung einer Maus den Mauszeiger bewegen zu lassen. Das *Binden* eines Treibers an ein Interface bedeutet, dass die Signale dieses Interfaces von dem entsprechenden Treiber verarbeitet werden [2]. Insbesondere können Treiber generisch für eine bestimmte Interface- oder Geräte-Klasse bestimmt sein oder speziell für ein bestimmtes Produkt eines Herstellers.

2.2. USB-Angriffsvektoren

Aufgrund der hohen Verbreitung bietet der USB ein attraktives Angriffsziel [7]. Daher existieren bereits verschiedene Angriffe, welche den USB verwenden [10], [29]. Zur Definition des Angreifermodells in Abschnitt 4.1 werden diese Angriffe im Folgenden in vier Kategorien unterteilt.

Die erste Kategorie sind physische Angriffe. USB-Geräte, welche physische Angriffe durchführen, brauchen nur Kontakt zum USB-Steckplatz. Sie müssen nicht den USB-Standard implementieren und sind aus der Perspektive des Betriebssystems nicht zwingend sichtbar. Ein physischer Angriff kann beispielsweise das Abgeben einer hohen Spannung sein, welche dazu führen kann, dass die Elektronik des Computers zerstört wird [10].

Darüber hinaus existieren Angriffe, welche auf den Bootvorgang des Computers abzielen. Hier wird versucht von einem maliziösen Gerät zu booten und so zum Beispiel ein Root-Kit zu starten, welches unterhalb des Betriebssystems läuft [10]. Auch Angriffe dieser Kategorie sind aus Perspektive des Betriebssystems nicht zwingend sichtbar.

Die dritte Kategorie sind Angriffe, bei denen Schwachstellen in den im Betriebssystem installierten USB-Treibern ausgenutzt werden, um zum Beispiel Root-Berechtigungen zu erlangen [6], [7]. Dies ist möglich, da USB-Treiber unter Linux mit Kernel-Privilegien laufen [1].

Zuletzt gibt es Angriffe durch USB-Geräte, welche vorgeben die Intention des Benutzers auszuführen. Hierbei führen die maliziösen USB-Geräte ausschließlich legitime Aktionen aus, zu welchen sie vom jeweiligen Treiber gewollt befähigt werden. Dies kann beispielsweise das Simulieren von Tasteneingaben sein, um Code auszuführen [10].

3. Verwandte Arbeiten

Verwandte Arbeiten existieren sowohl unter den akademischen Arbeiten als auch nicht-akademischen Softwarelösungen, welche vor USB-basierten Angriffen schützen sollen.

Vor allem gegen Angriffe von USB-Geräten, welche beim Anschluss Tasteneingaben tätigen, existieren verschiedene Ansätze. In 2018 haben Neuner et al. ein Konzept entwickelt, welches anhand von Zeitabständen zwischen Tasteneingaben (im folgenden „Zwischenankunftszeiten“) einen Angriff erkennt. Die Grundannahme ist, dass ein USB-Gerät, welches einen solchen Angriff ausführt, die Tasteneingaben signifikant schneller tätigt als ein Mensch. Sie schlagen ein System vor, das diese Zwischenankunftszeiten eines Geräts kontinuierlich misst. Wird die Zwischenankunftszeit von 20 ms bei mehr als drei aufeinander folgenden Tasteneingaben unterschritten, wird das Gerät als maliziös eingestuft und blockiert. [9]

Basierend auf diesen Erkenntnissen wurde später die Software „ukip“ veröffentlicht, welche einen Dienst implementiert, der Tasteneingaben von USB-Geräten überwacht und bei Unterschreitung einer konfigurierbaren Zwischenankunftszeit für eine konfigurierbare Anzahl aufeinanderfolgender Tasteneingaben das Gerät blockiert. [8]

Ein weiterer Ansatz wird von verschiedenen Antivirensoftwareanbietern wie G DATA [15] oder Kaspersky [5] angeboten. Beim Anschluss eines USB-Geräts, welches der Computer als Tastatur erkennt, wird eine Benachrichtigung generiert, auf welcher bestätigt werden muss, dass eine Tastatur angeschlossen wurde. Mithilfe der Maus muss diese Aktion durch Eingabe einer angezeigten Zahl auf einer eingeblendeten Bildschirmtastatur bestätigt werden. [5]

Es gibt auch Ansätze, die gegen allgemeine USB-basierte Angriffe entwickelt wurden. Die Software „USBGuard“ erlaubt es, Regeln zu definieren, nach denen ein USB-Gerät beim Anschluss automatisch zugelassen oder abgelehnt werden soll [16]. Diese Regeln betrachten dabei unter anderem die in Kapitel 2.1 besprochenen Eigenschaften, aber auch andere Kriterien wie beispielsweise den benutzten USB-Anschluss [16]. Einen ähnlichen Ansatz verfolgt auch die Software „usbauth“, mit dem Unterschied, dass diese nicht nur USB-Geräte, sondern auch USB-Interfaces einzeln verwalten kann [6].

Diese beiden Ansätze haben gemein, dass sie technisches Wissen voraussetzen und effektive Regeln zur Blockierung von maliziösen USB-Geräten vom Benutzer verlangen. Im Gegensatz dazu funktioniert der von Müller et al. vorgeschlagene Ansatz ohne Benutzerinteraktion oder Konfiguration. Es wird angenommen, dass der Benutzer nicht anwesend ist, wenn der Sperrbildschirm aktiv ist. Ist der Sperrbildschirm aktiv, werden grundsätzlich keine neuen USB-Geräte mehr zugelassen. Die einzige Ausnahme sind Tastaturen, da diese essenziell zur Authentifizierung des Benutzers sind. Dieser Ansatz bietet nur Schutz, während der Sperrbildschirm aktiv ist. [7]

4. Der USB Protector

In diesem Kapitel wird das Konzept des USB Protectors vorgestellt. Zunächst wird hierzu das Angreifermodell erläutert, gegen welches der USB Protector schützt, und anschließend werden die dazu eingesetzten Mechanismen erklärt.

4.1. Angreifermodell

Nicht alle Arten der in Kapitel 2 vorgestellten USB-Angriffe können durch eine im Betriebssystem laufende Software verhindert werden. Sowohl physische Angriffe als auch Angriffe auf den Bootvorgang sind aus der Perspektive des Betriebssystems nicht zwingend sichtbar und können daher im Allgemeinen nicht durch Software verhindert werden. Um gegen physische Angriffe zu schützen, ist eine physische Lösung notwendig. Gegen Angriffe auf den Bootvorgang gibt es andere Schutzmaßnahmen, wie beispielsweise die Vergabe eines BIOS/UEFI-Passworts. Im Rahmen dieser Arbeit werden nur Angriffe auf Gerätetreiber und Angriffe durch Simulation von Benutzerverhalten, während das Betriebssystem bereits läuft, betrachtet.

Neben den Angriffstypen werden auch noch die Kontexte betrachtet, in denen der Angriff stattfindet. Der Kontext beschreibt hierbei den Status, in dem sich der Computer zum Zeitpunkt des Anschlusses eines USB-Geräts befindet. Ist der Computer ausgeschaltet oder gesperrt, kann nicht davon ausgegangen werden, dass der legitime Benutzer aktuell am Computer anwesend ist. Ist der Computer eingeschaltet und entsperrt, so wird davon ausgegangen, dass der legitime Benutzer am Computer anwesend ist.

4.2. Konzept

Der Grundgedanke besteht darin, in das automatische Binden von Treibern an USB-Geräte einzugreifen, und vor dem Binden der Treiber die möglichen Auswirkungen dieser Aktion zu evaluieren. Diese Evaluation geschieht zunächst auf Basis des aktuellen Kontexts und gegebenenfalls durch weitere Verifikation durch den Benutzer. Das Binden

eines Treibers an ein USB-Gerät wird im Folgenden als „zulassen“ bezeichnet, das Unterlassen des Bindens eines Treibers an ein USB-Gerät sowie das Entfernen eines bereits gebundenen Treibers als „blockieren“.

USB-Geräte, die angeschlossen werden, während der Computer ausgeschaltet oder gesperrt ist, werden blockiert, da nicht davon auszugehen ist, dass der legitime Benutzer den Computer bedient. Das Potenzial für einen Angriff in diesem Kontext ist hoch, aber die Wahrscheinlichkeit, dass der legitime Benutzer selbst in diesem Kontext ein USB-Gerät anschließt, gering. Um zu bestimmen, ob ein USB-Gerät angeschlossen wurde, während der Computer ausgeschaltet war, wird durchgehend eine Liste aller zugelassenen USB-Geräte geführt. Erkennt der USB Protector beim Start ein USB-Gerät, welches nicht auf dieser Liste steht, wird angenommen, dass dieses angeschlossen wurde, während der Computer ausgeschaltet war.

Ist der Computer eingeschaltet und entsperrt, wird davon ausgegangen, dass der legitime Benutzer sich am Computer befindet. In diesem Fall muss die Möglichkeit bestehen, neue USB-Geräte zuzulassen. Allerdings besteht hier ebenfalls die Gefahr, dass der Benutzer selbst ein maliziöses USB-Gerät anschließt, beispielsweise wenn das USB-Gerät aus einer nicht vertrauenswürdigen Quelle bezogen wurde. Daher wird hier in einem zweiten Schritt die potenzielle Maliziösität des USB-Geräts durch Benutzerinteraktion verifiziert.

```
Bus 002 Device 008: ID 05ac:0220 Apple, Inc.
                         Aluminum Keyboard (ANSI)
Device Descriptor:
  [...]
  Configuration Descriptor:
    [...]
    Interface Descriptor:
      bInterfaceClass         8 Mass Storage
      [...]
    Interface Descriptor:
      bInterfaceClass         3 Human Interface
                                Device
      bInterfaceProtocol      1 Keyboard
      [...]
```

Abbildung 1: Die unterschiedlichen Darstellungen des USB Rubber Ducky. Links ist die Darstellung für einen Menschen, rechts für einen Computer.

Es wird angenommen, dass Angriffe durch USB-Geräte, welche der legitime Benutzer selbst an seinen Computer anschließt, insbesondere durch solche USB-Geräte ausgeführt werden, von welchen der Benutzer andere Fähigkeiten erwartet, als das USB-Gerät dem Computer gegenüber angibt. In Abbildung 1 wird dies am Beispiel des „USB Rubber Ducky“ dargestellt, welcher sich einem Menschen gegenüber als USB-Stick (Massenspeicher) darstellt, aber dem Computer gegenüber unter anderem als Tastatur.

Daher wird in diesem Fall eine Benachrichtigung generiert, in welcher die Fähigkeiten, die das USB-Gerät dem Computer gegenüber angibt, angezeigt werden. Die Fähigkeiten werden aus den in Abschnitt 2.1 besprochenen Informationen über die Geräte- und Interface-Klasse abgeleitet. Diese kann der Benutzer nun mit seinen Erwartungen vergleichen und, falls die Erwartungen mit den Fähigkeiten des USB-Geräts übereinstimmen,

durch Knopfdruck an der Benachrichtigung das Zulassen eines USB-Geräts veranlassen. Abbildung 2 zeigt eine solche Benachrichtigung.

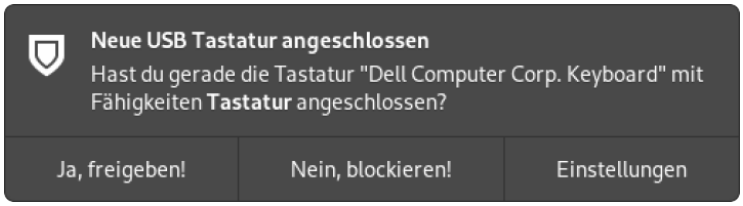


Abbildung 2: Dialog zur Abfrage der Benutzererwartung bei Anschluss des USB-Geräts.

Darüber hinaus werden einige Ausnahmefälle behandelt, welche das Arbeiten mit dem USB Protector vereinfachen. Tastaturen werden mit einem eingeschränkten Tastensatz automatisch zugelassen, welcher so gewählt ist, dass ein Benutzer sich zwar authentifizieren kann, aber keine Programme allein durch Tasteneingaben gestartet werden können. Weiterhin werden Mäuse automatisch zugelassen, da mindestens eine Maus notwendig ist, um mit den Benachrichtigungen zu interagieren und weitere USB-Geräte zuzulassen. Zuletzt werden auch USB-Hubs automatisch zugelassen, da Mäuse und Tastaturen auch häufig an einen USB-Hub angeschlossen werden. Wäre der entsprechende USB-Hub nicht zugelassen, so würde der Computer nicht erkennen, dass ein USB-Gerät an diesen USB-Hub angeschlossen wurde.

Weiterhin gibt es eine *Allowlist*, auf welcher der Benutzer USB-Geräte eintragen kann, welche beim Anschluss immer automatisch zugelassen werden sollen. Dies ist insbesondere für USB-Geräte gedacht, welche häufig ein- und ausgesteckt werden. Für die Verwaltung der Einstellungen des USB Protectors steht dem Benutzer eine grafische Schnittstelle zur Verfügung.

5. Benutzerfreundlichkeitsstudie

Zur Überprüfung, ob die Implementierung des USB Protectors benutzerfreundlich ist, wird ein Experiment durchgeführt. Dieses Kapitel beschreibt zunächst die Auswahl der Metriken sowie den Aufbau und Ablauf des Experiments. Anschließend werden die Ergebnisse des Experiments besprochen.

5.1. Auswahl der Metriken

Zur Messung der Benutzerfreundlichkeit werden die in der DIN EN ISO 9241-11 [4] zur Messung der Gebrauchstauglichkeit standardisierten Metriken *Effektivität*, *Effizienz* und *Zufriedenstellung* verwendet. Zur Messung der Effektivität wird die Fehlerrate betrachtet, mit der ein Teilnehmer eine gegebene Aufgabe durchführt. Die Effizienz wird anhand der Anzahl der Unterbrechungen, welche durch die Benutzung der USB-Geräte bei der Bewältigung einer Aufgabe entstehen, gemessen. Um die Zufriedenstellung der Teilnehmer zu messen, wird die *System Usability Scale* (SUS) gewählt [10].

5.2. Rekrutierung der Teilnehmer

Zur Teilnehmerrekrutierung wurde im Bekanntenkreis nach Interessierten gefragt. Insgesamt haben sich neun Personen bereit erklärt am Experiment teilzunehmen.

5.3. Aufbau des Experiments

Für das Experiment wird sowohl ein Computer als auch eine Auswahl an USB-Geräten benötigt, welche die Teilnehmer nutzen, um mit dem Computer zu interagieren. Die eingesetzten USB-Geräte und ihre Funktion im Experiment sind in Tabelle 1 aufgelistet. Außerdem wird ein Fragebogen mit der SUS benötigt.

Auf dem Computer ist Linux mit der Desktopumgebung „GNOME Shell“ in der Version 3.36.2 installiert. Außerdem ist der USB Protector installiert und eingeschaltet. Weiterhin ist der E-Mail-Client „Mozilla Thunderbird“ mit einem eingerichteten E-Mail-Konto installiert, in dessen Postfach sich eine E-Mail mit einem Anhang befindet.

USB-Gerät	Funktion / Demonstration
Rubber Ducky	Maliziöses USB-Gerät
Massenspeicher	Alternative zum Rubber Ducky
Tastatur	Keyboard-Mode
Maus	Automatisches Zulassen
WLAN-Adapter	Automatisches Blockieren
Hub	Erhöhung der Anzahl verfügbarer USB-Anschlüsse

Tabelle 3: USB-Geräte, welche den Teilnehmern bei der Durchführung des Experiments gestellt werden.

Zu Beginn des Experiments ist der Computer entsperrt und alle USB-Geräte sind nicht eingesteckt. Den Teilnehmern stehen alle USB-Geräte mit Ausnahme des USB-Massenspeichers zur Verfügung.

5.4. Ablauf des Experiments

Zu Beginn des Experiments werden die Teilnehmer um eine Selbsteinschätzung ihrer technischen Kompetenzen auf einer Skala von 1-10 gebeten. Die Teilnehmer werden außerdem gebeten, während der Durchführung des Experiments alle Aktionen laut auszusprechen zu begründen. Dann werden den Teilnehmern die relevanten Funktionen der GNOME Shell erklärt, um den Einfluss von Schwierigkeiten in der Benutzung der GNOME Shell auf die Bewertung der Nutzbarkeit des USB Protectors zu verringern. Da für das Experiment ein Laptop verwendet wird, wird außerdem erwähnt, dass die in dem Laptop eingebaute Tastatur sowie das eingebaute Touchpad nicht zur Bearbeitung der Aufgabe genutzt werden dürfen. Dies hat den Zweck, den Teilnehmer dazu zu zwingen, die USB-Geräte zu nutzen.

Nun wird dem Teilnehmer die Aufgabe erklärt. Zunächst soll er eine E-Mail von dem eingerichteten E-Mail-Konto abrufen und den Anhang auf einem USB-Stick speichern. Anschließend soll der Computer neu gestartet werden und der Teilnehmer soll sich anmelden. Das Passwort zur Anmeldung wird ihm dafür zur Verfügung gestellt. Nach der Anmeldung wird dem Teilnehmer gesagt, es gäbe eine Funktion, mit der erreicht werden

kann, dass ein USB-Gerät in Zukunft immer automatisch zugelassen wird. Der Teilnehmer wird aufgefordert, nach der Schnittstelle für diese Funktion zu suchen, um sie für ein beliebiges USB-Gerät zu benutzen.

Falls der Teilnehmer während des Experiments die Maliziösität des gegebenen USB Rubber Duckys erkennt, wird dieser durch den USB-Massenspeicher ausgetauscht. Als Letztes wird der Teilnehmer gebeten, den Fragebogen mit der SUS auszufüllen. Danach ist das Experiment beendet.

5.5. Ergebnisse

Zunächst wird die Bewertung der Benutzerfreundlichkeit in den drei Teilaspekten Effektivität, Effizienz und Zufriedenstellung einzeln betrachtet. Anschließend werden weitere Auffälligkeiten der Erhebung besprochen.

5.5.1. Fehlerrate

Während der ersten Aufgabe, dem Speichern des Anhangs einer E-Mail auf dem USB-Massenspeicher, interagiert ein Teilnehmer ausschließlich über die Benachrichtigungen mit dem USB Protector, daher werden während der ersten Aufgabe nur diese Interaktionen betrachtet. Das Zulassen eines maliziösen Geräts sowie das Blockieren eines nicht maliziösen Geräts werden als Fehler gewertet.

In der zweiten Aufgabe, dem Eintragen eines USB-Geräts auf der Allowlist, muss ein Teilnehmer zunächst die GUI starten, das gewünschte USB-Gerät identifizieren und anschließend die entsprechende Schaltfläche zum Eintragen dieses Geräts auf die Allowlist betätigen. Als Fehler gilt sowohl, falls der Teilnehmer nicht in der Lage ist, selbstständig die GUI zu öffnen, als auch der Fall in dem der Teilnehmer nicht in der Lage ist, bei bereits geöffneter GUI das gewünschte USB-Gerät auf die Allowlist einzutragen.

Abbildung 4 stellt die Fehlerraten der Teilnehmer dar. Im Durchschnitt ergab sich eine Fehlerrate von 0,26. Im Median haben die Teilnehmer acht zu betrachtende Aktionen durchgeführt und dabei zwei Fehler gemacht. Teilnehmer 7 hat alle Aufgaben ohne Fehler durchgeführt. Die maximale Fehlerrate liegt mit 0,5 bei Teilnehmer 3. Die Standardabweichung der Fehlerrate liegt bei 0,14.

Das Zulassen des maliziösen USB Rubber Ducky wird stärker gewichtet als das Blockieren eines nicht maliziösen USB-Geräts, da Teilnehmer, welche den USB Rubber Ducky blockieren, durch den Einsatz des nicht maliziösen USB-Massenspeichers insgesamt mehr betrachtete Aktionen durchführen, was bei gleicher Fehleranzahl zu einer geringeren Fehlerrate führt.

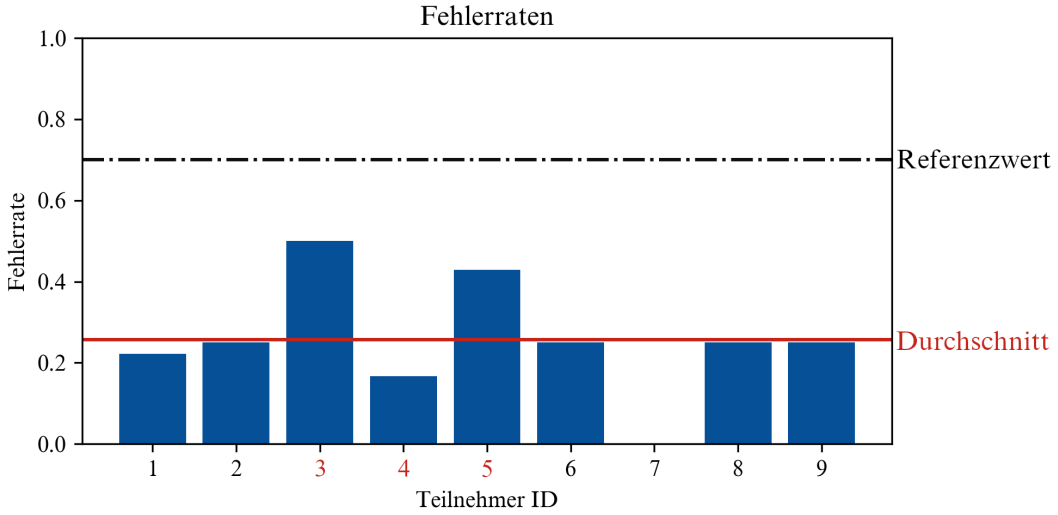


Abbildung 4: Fehlerraten der Teilnehmer. Der Referenzwert anderer Studien liegt bei 0,7 [17], die durchschnittliche Fehlerrate des Experiments bei 0,26. Auf die rot markierten Teilnehmer 3, 4 und 5 konnte erfolgreich ein Angriff durchgeführt werden.

Die relativ kleine Schwankung der Fehlerrate weist darauf hin, dass die Anzahl der Fehler eher auf den USB Protector an sich oder seine Integration, als auf unterschiedliche Eigenschaften der Teilnehmer zurückzuführen ist.

Eine Beschwerde aller Teilnehmer war, dass die Benachrichtigungen verschwinden, bevor die Teilnehmer sie lesen können. Diese verschwindenden Benachrichtigungen haben teilweise dazu geführt, dass USB-Geräte nicht wie gewollt zugelassen oder blockiert werden können. Dieses Verhalten kann ein Grund für die hohe Fehlerrate sein.

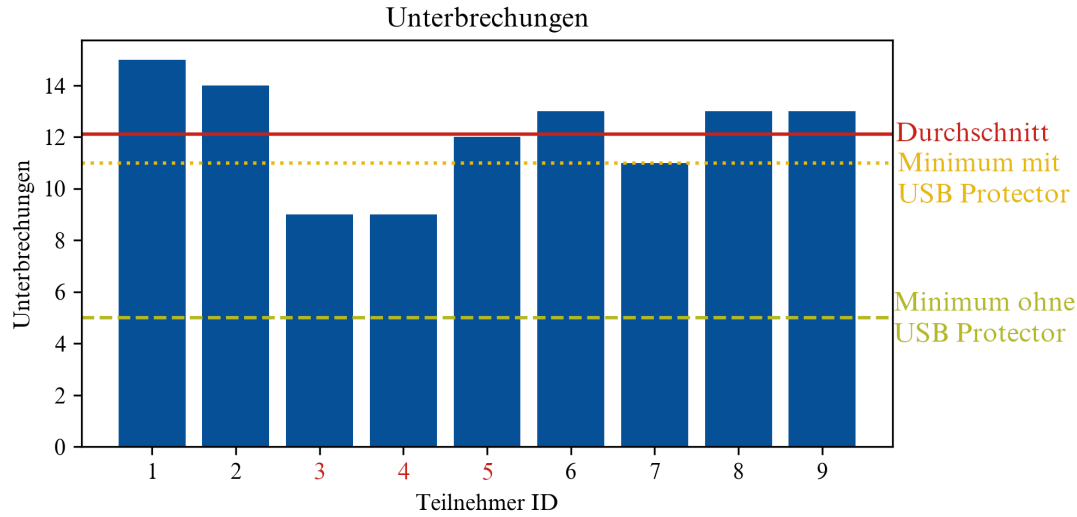


Abbildung 5: Unterbrechungen der Teilnehmer. Der Durchschnitt an Unterbrechungen liegt bei 12,1. Das Minimum an Unterbrechungen bei fehlerfreier Durchführung mit USB Protector liegt bei 11, ohne USB Protector bei 5. Auf die rot markierten Teilnehmer 3, 4 und 5 konnte erfolgreich ein Angriff durchgeführt werden.

5.5.2. Unterbrechungen

Als Unterbrechung wird sowohl das Erscheinen einer durch den USB Protector generierten Benachrichtigung als auch das Einstecken eines USB-Geräts gewertet, da dies

die Aktionen sind, welche durch die Nutzung des USB Protectors potenziell häufiger stattfinden als ohne dessen Nutzung.

Mit Einsatz des USB Protectors wird ein Teilnehmer bei fehlerfreier Durchführung der Aufgabe elfmal unterbrochen. Ohne den Einsatz des USB Protectors gäbe es nur fünf Unterbrechungen, durch die Anschlüsse der fünf USB-Geräte. Dies ist dem Konzept geschuldet, die vom Nutzer erwartete und tatsächliche Funktionalität des Gerätes zu vergleichen.

Abbildung 5 zeigt die Unterbrechungen der Teilnehmer. Im Durchschnitt wurde ein Teilnehmer 12,1 mal unterbrochen. Teilnehmer 1 hatte mit 15 Unterbrechungen die meisten, Teilnehmer 3 und 4 mit jeweils neun Unterbrechungen die wenigsten Unterbrechungen.

5.5.3. Zufriedenstellung

Zuletzt wird die Zufriedenstellung anhand der SUS Punktzahlen gemessen. Die Verteilung der Punktzahlen wird in Abbildung 6 dargestellt. Die durchschnittliche Punktzahl der SUS liegt mit 70,6 knapp über dem Referenzwert von 68 [13]. Teilnehmer 4 und 6 haben mit jeweils 92,5 Punkten die höchste Punktzahl vergeben, Teilnehmer 7 mit 30 Punkten die niedrigste. Die Standardabweichung der Punkte beträgt 21,2.

Insgesamt ist die Zufriedenstellung gut, schwankt aber relativ stark. Dies deutet darauf hin, dass die Teilnehmer möglicherweise verschiedene Kriterien für die Zufriedenstellung haben oder aufgrund unterschiedlicher Aktionen verschiedene Eindrücke des USB Protectors bekommen haben. Eine signifikante Korrelation mit den Unterbrechungen oder Fehlerraten konnte nicht festgestellt werden.

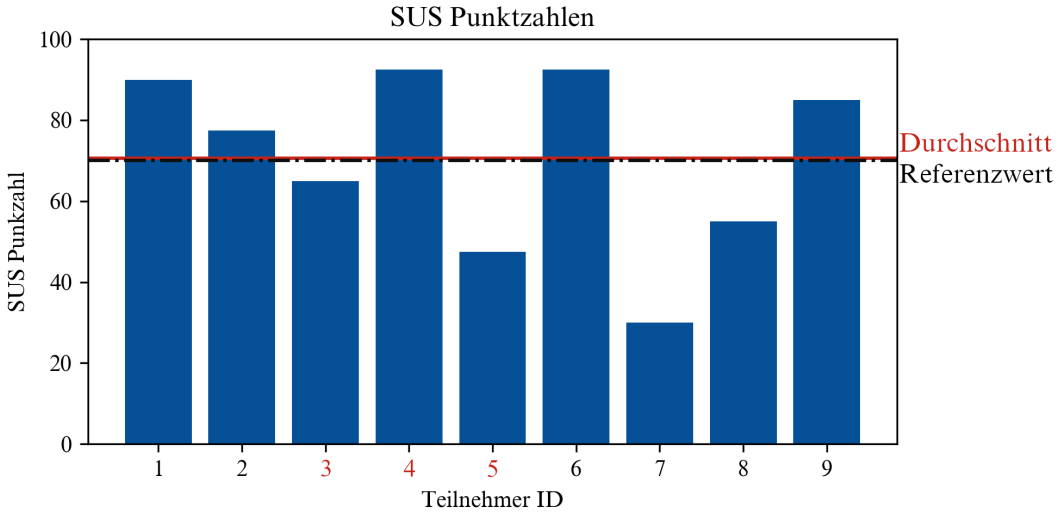


Abbildung 6: SUS Punktzahlen der Teilnehmer. Der Durchschnittswert liegt bei 70,6. Der Referenzwert aus anderen Untersuchungen liegt bei 68 [13].

5.5.4. Weitere Auffälligkeiten

Aus den Beobachtungen der Teilnehmer konnten noch weitere Erkenntnisse gewonnen werden. Zum einen ist aufgefallen, dass vier der Teilnehmer einige Benachrichtigungen nicht vollständig gelesen haben, und das USB-Gerät sofort zuließen. Bei einem solchen

Verhalten bietet der USB Protector keinen Schutz vor Angriffen von Geräten, welche der Benutzer selbst anschließt. Weiterhin ist aufgefallen, dass alle neun Teilnehmer mit Benachrichtigungen für Geräte interagierten, welche bereits automatisch zugelassen wurden.

Die Selbsteinschätzung der technischen Kompetenz war unterschiedlich, mit einer durchschnittlichen Einschätzung von 6,4 und einer Standardabweichung von 2,3. Es konnte keine Korrelation mit einem der anderen Werte festgestellt werden.

6. Diskussion und Ausblick

Das Konzept des USB Protectors basiert auf einigen Annahmen, welche bei der Verallgemeinerung der Ergebnisse beachtet werden müssen. Nicht alle USB-Geräte implementieren generische USB-Klassen, aus denen die Fähigkeiten des Geräts abgeleitet werden könnten. Für USB-Geräte, welche herstellerspezifische Treiber laden, können keine Fähigkeiten ermittelt werden, die der Benutzer mit seiner Erwartung vergleichen kann.

Weiterhin ist zu beachten, dass die Testumgebung das Verhalten der Teilnehmer beeinflusst haben könnte. Es ist denkbar, dass für einige Teilnehmer das erfolgreiche Abschließen der gestellten Aufgabe eine höhere Priorität als der Schutz vor einem maliziösen USB-Gerät hat. Insbesondere bei den Teilnehmern, welche einige Benachrichtigungen nicht gelesen haben oder den USB Rubber Ducky zuließen, ist denkbar, dass sie diese Entscheidung aufgrund der Testsituation getroffen haben, aber an ihrem privaten Computer vorsichtiger gehandelt hätten.

Es konnte gezeigt werden, dass ein benutzerfreundlicher Schutz gegen USB-basierte Angriffe weitgehend möglich ist. Jedoch konnte auch Optimierungspotenzial aufgezeigt werden. Im Experiment konnte auf ein Drittel der Versuchsteilnehmer ein erfolgreicher Angriff durchgeführt werden. Dieser kam durch Zulassen des USB Rubber Ducky, entweder ohne Lesen der Benachrichtigung oder durch eine falsche Interpretation dieser zustande. Es ist davon auszugehen, dass dieser Effekt durch die kurze Sichtbarkeit der Benachrichtigungen noch verstärkt wurde.

Ein Ansatz, welcher auf Verifikation durch den Benutzer baut, sollte also sicherstellen, dass der Benutzer sich ausreichend mit der Verifikation auseinandersetzt und auch die Gelegenheit dazu hat. Außerdem scheint man dem Benutzer noch klarer machen zu müssen, dass das Zulassen eines maliziösen USB-Geräts für den Benutzer nicht erwünschte Folgen haben kann.

Die kurze Zeitspanne, in der mit der Benachrichtigung interagiert werden konnte und die damit wahrscheinlich verbundene hohe Fehlerrate wird sich zukünftig verbessern. In ihrer jetzigen Form eignen sich die Benachrichtigungen wie die GNOME Shell sie darstellt nicht für die Kommunikation in sicherheitsrelevanten Entscheidungsprozessen. Das GNOME Projekt strebt jedoch eine Neugestaltung dieses Bereichs der Schnittstelle an [11].

Der USB Protector steht unter der GNU GPLv3 öffentlich zur Verfügung [18].

Literaturhinweise

[1] Butt, S. u.a.: „Protecting commodity operating system kernels from vulnerable device drivers“. In: „2009 Annual Computer Security Applications Conference“. IEEE. 2009, S.301-310.

[2] The Kernel Development Community: „Driver Binding“, https://www.kernel.org/doc/html/latest/driver-api/driver-model/binding.html [Online, abgerufen am 11.09.2020].

[3] „Defined Class Codes“, https://www.usb.org/defined-class-codes [Online, abgerufen am 03.08.2020], USB Implementers Forum, Inc.

[4] DIN-EN-ISO 9241-11 „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten Teil 11: Anforderungen an die Gebrauchstauglichkeit — Leitsätze“, Deutsches Institut für Normung, März 1998.

[5] „Keyboard authorization“. https://support.kaspersky.com/KESWin/11/en-us/97196.htm. [Online, abgerufen am 12.08.2020]

[6] Koch, S.: „Sicherheitsaspekte beim Anschluss von USB-Geräten“. Masterarbeit. 2017.

[7] Mueller, T., Zimmer, E. und de Nittis, L.: „Using Context and Provenance to defend against USB-borne attacks“. In: „Proceedings of the 14th International Conference on Availability, Reliability and Security“. 2019, S. 1-9.

[8] Neuner, S.: „USB Keystroke Injection Protection“. https://github.com/google/ukip. [Online, abgerufen am 12.08.2020]. 2020.

[9] Neuner, S. u. a.: „Usblock: Blocking usb-based keypress injection attacks“. In: „IFIP Annual conference on Data and Applications Security and Privacy“. Springer. 2018, S.278-295

[10] Nissim, N., Yahalom, R. und Elovici, Y.: „USB-based attacks“. In: „Computers & Security“ 70. 2017, S. 675-688.

[11] Pícolo, M.: „GSoC Ending“. https://marianapicolo.com/blog/2020/08/29/gsoc-ending/. [Online, abgerufen am 15.09.2020]. 2020

[12] Rubini, A. und Corbet, J.: „Linux device drivers“, O‘Reilly Media, Inc., 2001.

[13] Sauro, J.: „A practical guide to the system usability scale: Background, benchmarks & best practices“, Measuring Usability LLC, 2011.

[14] „Universal Serial Bus Specification“, Rev 2.0, USB Implementers Forum, Inc., 2000.

[15] „USB Keyboard Guard | Zuverlässiger Schutz vor manipulierten USB-Sticks | G DATA“. https://www.gdata.de/de-usb-keyboard-guard. [Online, abgerufen am 12.08.2020].

[16] „USBGuard“. https://usbguard.github.io/. [Online, abgerufen am 13.08.2020].

[17] Sauro, J.: „A Practical Guide to Measuring Usability: 72 Answers to the Most Common Questions about Quantifying the Usability of Websites and Software“, CreateSpace Independent Publishing Platform, 2010.

[18] https://gitlab.gnome.org/timopohl/usb-protector.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Automatisierte Compliance-Prüfung in Software-Artefakten

Christian Banse[1], Florian Wendland M.Sc.[1], Konrad Weiss[1]

Kurzfassung:
Die stetig wachsende Zahl von Regularien und Standards im Bereich der IT-Sicherheit sowie die weiterhin steigende Komplexität von Software macht eines deutlich: Die Prüfung entsprechender Standards muss in Zukunft automatisiert geschehen. Während dies in einigen Teilbereichen, wie dem Cloud Computing, bereits heute in Teilen möglich ist, bedarf es in anderen Bereichen noch weiterer Forschung. Eine grundlegende Herausforderung ist hierbei unter anderem, dass regulatorische Werke in der Regel in natürlicher Sprache geschrieben sind. Es ist also zunächst eine formale Modellierung dieser Sicherheitsanforderungen nötig. Im Rahmen dieses Artikels werden verschiedene existierende Ansätze dieser Modellierung aufgezeigt. Darüber hinaus diskutiert der Artikel, wie, basierend auf einer formalen Modellierung, Sicherheitsanforderungen technisch in Software abgeprüft werden können. Schließlich werden im Ausblick noch offene wissenschaftliche Fragestellungen diskutiert. Insbesondere ist die Integration in einen mehrstufigen automatisierten Prozess von quelltextabhängigen Anforderungen zu Metriken bis zu Analyseverfahren eine noch zu lösende Herausforderung.

Stichworte: Automatisierte Software-Prüfung, Compliance-Prüfung, Formale Modellierung

1. Motivation und Problemstellung

Die Anzahl der relevanten Regularien und Standards im Bereich Sicherheit und Datenschutz, wie der EU-DSGVO oder der BSI C5, ist in den letzten Jahren enorm angestiegen. Gleichzeitig befindet sich die Komplexität und der Umfang von Quellcode sowie die Anzahl von externen Abhängigkeiten in einem Software-Produkt ebenfalls in einem stetigen Anstieg. Dies hat zur Folge, dass Hersteller und insbesondere Entwickler von Software unter einem enormen Druck stehen, einerseits qualitativ hochwertige Software zu entwickeln, die aber andererseits auch den bestehenden Sicherheitsstandards entspricht. Werden diese Fragestellungen nicht rechtzeitig adressiert und verbleiben Sicherheitsmängel in den Programmen, drohen ernstzunehmende Konsequenzen von Imageschäden bis hin zu erheblichen Strafen bei Compliance-Verstößen.

Im schlimmsten Fall werden diese Lücken erst zur Laufzeit bei einer Auditierung entdeckt. Falls eine Zertifizierung einer Dienstleistung angestrebt wird, ist es aber ratsam, viele Anforderungen bereits während der Entwicklung ordnungsgemäß zu erfüllen, um Zeit und Kosten bei einer Auditierung zu sparen. Ein zunehmend beobachtbarer Trend ist daher, die Sicherheit im Software-Entwicklungs- und Lieferprozess sehr früh zu betrachten („shift left“). Zudem wird im Zuge der DevSec–Ops-Methodologie immer mehr Verantwortung beim Betrieb und der Sicherheit einer Anwendung auf den Entwickler ausgelagert, insbesondere bei Cloud-basierten Applikationen. Der Entwickler muss nun nicht nur technisch, in Bezug auf Programmiersprachen, Bibliotheken und Design Patterns, auf der Höhe der Zeit bleiben. Es wird vom ihm zusätzlich gefordert, ein Experte für Sicherheits- und Compliance-Fragen zu sein. Dies bedeutet in der Konsequenz, dass

[1] Fraunhofer AISEC, Lichtenbergstraße 11, 85748 Garching b. München

zunehmend Werkzeuge und Methoden verfügbar gemacht werden müssen, um Entwickler von Software bei diesem Wandel zu entlasten.

Zwar haben sich viele Projekte, von Open-Source-Tools bis hin zu kommerziellen Angeboten, zum Ziel gesetzt, Programmierfehler, Schwachstellen oder bereits bekannte Verwundbarkeiten in Programmen oder Bibliotheken zu finden. Diese gefundenen Schwachstellen spiegeln aber in der Regel nur einen Bruchteil von dem wider, was als Sicherheitsanforderung definiert sein kann. Darüber hinaus haben nur wenige Werkzeuge tatsächlich die Überprüfung der *Einhaltung* von Sicherheitsanforderungen (und nicht nur den Verstoß einiger weniger Anforderungen) im Fokus. Dies ist vor allem dadurch bedingt, dass Kriterienkataloge wie der BSI C5[2] oder Technische Richtlinien des BSI Anforderungen in Prosaform beschreiben, die von Softwareentwicklern mittels passender Implementierungen umgesetzt werden müssen. Während beispielsweise die *TR-02102-1: „Kryptographische Verfahren“* Empfehlungen und Hilfestellungen über Schlüssellängen und geeignete Algorithmen gibt, obliegt es dem Entwickler, für konkrete Anwendungen die geeigneten Funktionsaufrufe und Parameter der verwendeten kryptografischen Softwarebibliothek auszuwählen. Umgekehrt benötigen Werkzeuge Hinweise, dass eben diese ausgewählten Methoden und Konfigurationseinstellungen geeignet sind, die Richtlinien, z. B. an Kryptografie, zu erfüllen.

2. Modellierung von Anforderungen

Es braucht also zunächst einmal Formate, Protokolle oder auch domänenspezifische Sprachen, um Anforderungen formal und maschinenlesbar zu definieren. Erste Ansätze finden sich bereits 2005 bei Giblin et. al. [6] unter dem Begriff „model-Driven Compliance”. Die Autoren schlagen weiterhin ein Framework namens *Regulations Expressed As Logical Models (REALM)* [5] vor, welches regulatorische Standards wie beispielsweise den Sarbanes-Oxley-Act in einer formalen Weise modellieren kann. Aus diesem Modell können schließlich Regeln abgeleitet werden, die zur Überprüfung von Software-Programmen dienen können. Ein REALM-Modell besteht hierbei einerseits aus einem UML-basiertem Konzept-Modell und andererseits aus einem Regelwerk, basierend auf Ausdrücken der temporalen Logik.

Speziell für den Bereich der IT-Sicherheit hat die NIST mit der Open Security Controls Assessment Language (OSCAL) einen umfangreichen Standard geschaffen, um Kontrollkataloge, wie NIST 800-53 oder auch BSI C5 maschinenlesbar zu beschreiben. Der Standard bzw. das Modell der Sprache wurde als Open-Source-Lizenz veröffentlicht[3] und enthält als Beispiel bereits sehr große Teile der US-Standards FedRAMP und NIST 800-53 in maschinenlesbarer Form. Als technisches Format kann XML, JSON und YAML zum Einsatz kommen.

OSCAL ist grundsätzlich in mehrere Ebenen aufgeteilt und enthält zunächst Sprachelemente für die eigentliche Beschreibung von Security Controls und deren Zuordnung zu Rahmenwerken wie Zertifizierungs-Katalogen. Es ist auch zum Aufbau und der Pflege

[2] https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/CloudComputing/Anforderungskatalog/2020/C5_2020.pdf
[3] https://github.com/usnistgov/OSCAL

von Unternehmens-spezifischen Anforderungen geeignet und erlaubt die Kombination und den Querverweis mehrerer Quellen. Dies ist ein übliches Vorgehen in Unternehmen: Zum Aufbau eines unternehmensweiten Security-Frameworks werden in der Regel geeignete Controls aus etablierten Standards wie ISO 27001 oder BSI C5 ausgewählt. Diese werden dann um eigene, unternehmensspezifische Anforderungen ergänzt.

Darüber hinaus enthält der Standard auch beschreibende Elemente für Tools, wie Code-Scanner oder andere Security-Compliance-Werkzeuge, die eine (automatische) Bewertung der Sicherheit von Diensten oder Artefakten vornehmen. Wie genau diese Tools eine Zuordnung ihrer Analyseergebnisse (z. B. „SQL Injection gefunden“) zu einem Security Control vornehmen, welcher am Ende immer noch eine textuelle Beschreibung (z. B. „Benutzer-Eingaben müssen validiert werden“) ist, wird jedoch nicht näher spezifiziert. Es fehlt an einem „Bindeglied“, z. B. in Form von Metriken und an Methoden, um komplexere Analysen automatisiert von Compliance-Anforderungen abzuleiten.

Verschiedene andere domänenspezifische Sprachen versuchen, diese Brücke Analyseseitig zu schlagen. Beispielsweise ist es mit der Sprache *CrySL* [3], welche Teil des Eclipse *CogniCrypt* [4] ist, möglich, auf technische Art und Weise Anforderungen an die Benutzung von Java-Objekten im Kontext der Verschlüsselung zu stellen. Dies umfasst beispielsweise Anforderungen an Programmparameter, z. B. zur Definition von Verschlüsselungsalgorithmen oder Schlüssellängen. Ähnliche Funktionalitäten bietet die domänenspezifische Sprache *MARK*[4] als Teil des Open-Source-Projektes *Codyze*[5].

In dem vom BSI geförderten Projekt *Codyze,* entwickelte das Fraunhofer AISEC ein Werkzeug zur automatisierten Prüfung von kryptografischen Anforderungen in Quelltext. *MARK* erlaubt es Entwicklern von Bibliotheken, sogenannte *Entitites*, welche semantische Eigenschaften von abstrakten Objekten repräsentieren, zu modellieren. Auf diese Art und Weise bietet MARK beispielsweise die Möglichkeiten, die Metrik *Schlüssellänge* eines generischen Objektes *Verschlüsselung* zu modellieren. Diese generischen Objekte lassen sich schließlich auf konkrete Klassen, z. B. *javax.crypto.Cipher* in der Java Cryptography API übertragen. Zusätzlich lassen sich mit *Rules* Anforderungen an diese Entitäten stellen. Im Rahmen des Codyze-Projektes wurden so beispielsweise die Anforderungen der TR-02102[6] des BSI in MARK modelliert und auf die kryptografischen Softwarebibliotheken Botan[7] und Bouncy Castle[8] angewendet. Codyze zeigt damit Entwicklern an, wenn sie diese Bibliotheken in einer Art und Weise verwenden, die nicht den Anforderungen der TR-02102 entspricht. Mit diesen Hinweisen können Entwickler bereits in der Implementierungsphase die fehlerhafte Verwendung kryptografischer Verfahren ausschließen.

Dieser Ansatz hat somit sein Potenzial unter Beweis gestellt. Ein maximaler Nutzen für die Entwickler von Anwendungen ergibt sich jedoch erst, wenn eine ausreichende

4 https://www.codyze.io/docs/mark-authors/

5 https://codyze.io/

6 https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr02102/tr02102_node

7 https://botan.randombit.net/

8 https://www.bouncycastle.org/java.html

Menge an technischen Regeln in Sprachen wie MARK oder CrySL vorhanden sind. Die Hoffnung besteht, dass Entwickler von Bibliotheken, wie beispielsweise Bouncy Castle oder OpenSSL, diese Konzepte aufgreifen und zusätzlich zur Dokumentation auch spezifische Verwendungsregeln mitliefern. Die automatisierte Prüfung dieser Regeln erlaubt es somit, eine inkorrekte Verwendung von API-Objekten zu vermeiden, die von den Bibliotheksentwicklern nicht geplant, aber durch die verwendete Programmiersprache möglich ist. Sicherheitsschwächen können dadurch entstehen, dass Kryptografie-Objekte nicht korrekt initialisiert oder die unterstützten Funktionen auf diesen in der falschen Reihenfolge ausgeführt werden.

3. Automatisierte Prüfung

Zur automatisierten Prüfung von Sicherheitsanforderungen ist es zuerst notwendig, die Brücke zwischen abstrakten Anforderungen und technisch bereits umsetzbaren Analyseverfahren zu schlagen. Das Übersetzen auf einer semantischen Ebene ist dabei die Kern-Herausforderung.

3.1. Von Anforderungen zu Metriken

Unabhängig von der Modellierungssprache müssen sich Unternehmen zunächst Gedanken machen, mit welchen Metriken konkret Sicherheitsanforderungen gemessen werden können. Nur so ist eine automatisierte und kontinuierliche Prüfung von Anforderungen möglich. Formal stellen Metriken eine Messvorschrift dar. Das Messergebnis dient idealerweise für den Nachweis, dass gewisse technische oder organisatorische Maßnahmen (TOMs) getroffen wurden. Diese Maßnahmen stellen die konkrete Implementierung eines abstrakten Security Controls dar. Innerhalb der Community gibt es eine anhaltende Diskussion, ob der Nachweis über die getroffenen Maßnahmen genügt, oder ob auch eine Messung der Effektivität dieser Maßnahmen entscheidend ist.

Die Komplexität einer Metrik kann beliebig einfach oder schwierig gestaltet sein. Beispiele für relativ einfache Metriken wären die Sicherheitsparameter einer TLS-Verbindung, wie Verschlüsselungsalgorithmus und die Stärke der Verschlüsselung. Beide lassen sich mit einfachen Mitteln aus dem TLS-Handshake berechnen. Deutlich komplexer sind jedoch Metriken zum Nachweis der sogenannten organisatorischen Maßnahmen, z. B. Einhaltung der Least-Privilege-Konzepte bei einer Rollenvergabe. Es existieren beispielsweise mehrere wissenschaftliche Arbeiten [7], [8], die sich dem Least-Privilege-Problem im Bereich Cloud Computing widmen. Die Messung solcher Metriken ist deutlich komplexer und umfasst in der Regel mehrere Teilschritte sowie eine Konfiguration durch den Benutzer.

Während derzeit noch Unternehmen sehr stark in der individuellen Verantwortung sind, selbst geeignete Metriken auszuwählen, wurde dieser Trend mittlerweile von der Wissenschaft und von Zertifizierungsstellen aufgegriffen. So hat sich das im November 2020 gestartete EU-Projekt MEDINA[9] zum Ziel gesetzt, einen Katalog von verschiedenen Metriken zur automatisierten Prüfung von Anwendungen in der Cloud zu erarbeiten.

[9] http://medina-project.eu

Unternehmen können so in Zukunft auf ein vordefiniertes Repository zurückgreifen. Verstärkt sichtbar wird dies auch in den Diskussionen rund um die derzeit geplante Harmonisierung des Zertifizierungsmarktes in Europa durch den EU Cyber Security Act. Insbesondere für Anwendungen, die einem hohen Assurance-Level entsprechen müssen, sind Methoden der kontinuierlichen Zertifizierung, welche eine automatisierte Prüfung von Metriken voraussetzt, in der Diskussion.

3.2. Technische Umsetzung

Aus der Anforderung, möglichst früh im Entwicklungsprozess mit der Überprüfung von Sicherheitsanforderungen zu beginnen, ergibt sich die Notwendigkeit, diese nicht auf den ausführbaren Programmen, die in binärer Form ausgeliefert werden, sondern auf dem Quelltext zu tun. Für eine Analyse eines bereits kompilierten Programmes fehlen so früh im Entwicklungsprozess oftmals benötigte Abhängigkeiten oder Toolchains.

Zur Prüfung von Sicherheitsanforderungen in Quelltext kann auf eine Vielzahl von verschiedenen Werkzeugen zurückgegriffen werden. Grundsätzlich lässt sich hierbei eine Klassifikation zwischen statischer und dynamischer Softwareanalyse vornehmen. Statische Analyse beruht rein auf der Analyse des vorliegenden Quelltextes oder des Kompilats, ohne die konkrete Ausführung. Dynamische Analyse hingegen betrachtet die Auswirkungen der konkreten Ausführung zur Laufzeit. Es existieren weitere Mischformen, wie beispielsweise Symbolic Execution oder Abstract Interpretation, in denen ein Programm abstrakt ausgeführt wird, also z. B. anstelle von konkreten Inputwerten Symbole bzw. die Menge aller möglichen Inputwerte angenommen wird.

Über die Jahre hat sich eine Vielzahl von Werkzeugen für die statische Codeanalyse etabliert. Kommerzielle Werkzeuge wie SonarSource, SonarQube[10] und Checkmarx SAST[11] sowie cppcheck[12] und FindBugs/SpotBugs[13] bieten Funktionalitäten, die von der Einhaltung von Coding Standards, über das Auffinden häufiger Programmierfehler bis zu Metriken zur Codequalität reichen. Insbesondere kommerzielle Angebote fallen durch einen großen Berichtsumfang auf und bieten für häufige Programmierfehler eine ausführliche Erklärung sowie entsprechende Lösungsansätze, wie diese zu beheben wären. Ein Nachteil bei diesen Werkzeugen ist jedoch, dass die zu überprüfenden Anforderungen im Werkzeug fest hinterlegt sind. Eine Erweiterung oder Anpassung der Anforderungen erfordert die Entwicklung von Plugins. Diese starke Kopplung von Codeanalyse an die konkrete Implementierung der Analyse behindert die einfache Erweiterung oder Anpassung von Anforderungen.

Neuere Werkzeuge für die Codeanalyse, wie beispielsweise die bereits erwähnten Projekte CogniCrypt und Codyze, wirken dieser Einschränkung entgegen, indem sie die Spezifikation von Anforderungen von der eigentlichen Codeanalyse trennen. Diese Trennung wird über domänenspezifische Sprachen ermöglicht. Diese Sprachen erlauben

[10] https://www.sonarqube.org
[11] https://www.checkmarx.com/products/static-application-security-testing
[12] https://github.com/danmar/cppcheck
[13] https://github.com/spotbugs/spotbugs

es, Anforderungen deskriptiv und deklarativ zu formulieren. Das Erweitern oder Anpassen von Anforderungen erfordert nur das Schreiben neuer Regeln in der entsprechenden DSL. Dazu verwenden sowohl CrySL (CogniCrypt) als auch MARK (Codyze) logische Ausdrücke. Es ist nicht mehr notwendig, eine Analyse in einer Programmiersprache und mit der API des Analysewerkzeuges zu schreiben. Eine Anforderung muss sich lediglich als logischer, deklarativer Ausdruck in der DSL darstellen lassen. Daraus werden die notwendigen Analyseschritte und die dazugehörigen Bedingungen erzeugt.

Für die Durchführung der statischen Analyse hat sich das Konzept der sogenannten Code Property Graphs (CPG) [1] als vielversprechendes Darstellungsformat von Quellcode gezeigt. Ein CPG versteht sich hierbei als Super-Graph über verschiedene, in der Programmanalyse verwendete Teilgraphen, wie Abstract Syntax Tree (AST), Control Flow Graphs (CFG) oder Data Flow Graphs (DFG). Es existieren Open-Source-Projekte verschiedener Organisationen[14], um beispielsweise aus Quelltextteilen (C/C++ oder Java) Graphen-Strukturen zu erzeugen. Dies erlaubt, ähnlich einer Datenbank, „Anfragen“ an den Quelltext zu stellen (z. B. *selektiere die Parameter aller Funktionsaufrufe, die ein Crypto-Objekt erzeugen*). Codyze nutzt dies, um in MARK modellierte Sicherheitsanforderungen als Abfragen in den Graphen zu übersetzen. Weighted Pushdown Systems (WPDS) [2] sind weitere Konzepte, die es erlauben, Anforderungen an die Benutzung von Programmaufrufen zu prüfen. Sie erlauben einerseits eine Datenflussanalyse und sind in der Kombination mit Nondeterministic finite automaton (NFA) geeignet, die Reihenfolge von Aufrufen (Typestate Analyse) zu analysieren.

Die Integration und Interaktion mit den Werkzeugen gestalten sich sehr unterschiedlich. In der Regel bieten alle Tools eine Integration in CI/CD-Prozesse an. Um allerdings Sicherheitsprobleme bereits möglichst früh zu erkennen, bietet sich eine Integration in Entwicklungsumgebung (IDEs) an. Das von Microsoft initiierte Language Server Protocol[15] (LSP) bietet dazu ein IDE-unabhängiges Protokoll an.

4. Ausblick

Während domänenspezifische Sprachen wie MARK die Grundlage gelegt haben, spezifische Anforderungen technisch prüfen zu können, bedarf es weiterer Kooperationen zwischen Industrie und Wissenschaft, um weitere Metriken wie Zugangskontrollen oder sichere Datenbankzugriffe in Zukunft automatisiert prüfen zu können. Das vom Fraunhofer AISEC entwickelte Tool Codyze hat diese neue Herangehensweise an Codeanalyse für kryptografische Softwarebibliotheken illustriert. Das Werkzeug kann die richtige Verwendung von kryptografischen Bibliotheken wie der Java Cryptography API erzwingen, hat aber bereits die technischen Grundlagen geschaffen, weitere Compliance-Regeln abdecken zu können. Es ist zu erwarten, dass in Zukunft kommerzielle Tools, aber auch Open-Source-Projekte, eine höhere Abdeckung von technischen Prüfungen anbieten werden und dass diese dann auch mittels Sprachen wie OSCAL mit Zertifizierungsrahmenwerken korreliert werden können. Dies erfordert jedoch auch,

[14] https://github.com/ShiftLeftSecurity/codepropertygraph, https://github.com/Fraunhofer-AISEC/cpg
[15] https://microsoft.github.io/language-server-protocol/

dass Entwickler von sicherheitskritischen Bibliotheken neben der menschenlesbaren Dokumentation spezifische Verwendungsregeln in maschinenlesbaren Formaten und Modellen zur Verfügung stellen.

Für eine effektivere Nutzung sind einige nuancierte Probleme technisch zu lösen. In Compliance-Anforderungen kann eine Regel das Einhalten des State-of-the-Art fordern, ohne dass dieser explizit genannt wird. Zur Modellierung von Sicherheitsanforderungen ist es notwendig solche „moving targets“ zu betrachten und auf technischer Ebene ein automatisches Anpassen der Metriken in Abhängigkeit der Änderungen durchzuführen. Ebenfalls müssen in diesem Zuge zeitlich begrenzte Regelungen modelliert werden. So sind Anforderungen, welche die Verwendung von spezifischen kryptografischen Algorithmen vorschreiben, zeitlich begrenzt, da die Sicherheit dieser nicht auf unbestimmte Zeit erwartet wird.

Eine weitere Herausforderung ist die Kombination von Ansätzen der statischen Code-Analyse mit weiteren Analysen, z. B. aus dem Bereich der Cloud-Sicherheit. So ist es zwar mit Quelltext-Analyse möglich, die Sicherheit der Anwendung selbst zu überprüfen, meistens beinhaltet dies jedoch keine Information über die Umgebung, in der die Anwendung läuft. Gerade in heutigen „cloud native“-Anwendungen wird jedoch ein Großteil der Sicherheit durch die Konfiguration der Umgebung beeinflusst. Es ist daher notwendig, bereits bestehende Verfahren [9], [10] und Tools der Cloud-Zertifizierung, wie z. B. Clouditor[16], mit Methoden der statischen Analyse zu verbinden, um eine Aussage über die Sicherheit der Cloud-Ressourcen und der darauf laufenden Anwendungen zu erreichen.

Auf Nutzerseite ist durch den “shift left”-Ansatz eine hohe Performance notwendig. Wenn sich Teile des Quelltextes bei laufender Implementierung ändern, ist für Unterstützung in eine IDE notwendig, dass sich diese Änderungen inkrementell in die Datenmodelle und Analyseverfahren integrieren lassen, um keine hohen Laufzeiteinbußen hinnehmen zu müssen. Inkrementelle Verfahren existieren bereits für singuläre Analyseverfahren. Die vertikale Integration in einen mehrstufigen automatisierten Prozess von quelltextabhängigen Anforderungen, zu Metriken, bis Analyseverfahren ist aber eine noch zu lösende Herausforderung.

[16] https://github.com/clouditor/clouditor

Literaturhinweise

[1] Yamaguchi, F; Golde, N.; Arp, A; Rieck, K: Modeling and Discovering Vulnerabilities with Code Property Graphs. IEEE Symposium on Security and Privacy 2014

[2] Reps, T; Schwoon, S.; Jha, S.; Melski, D.: Weighted pushdown systems and their application to interprocedural dataflow analysis. Sci. Comput. Program (2005)

[3] Krüger, S.; Späth, J.; Ali, K.; Bodden, E.; Mezini, M. (2018): CrySL: An Extensible Approach to Validating the Correct Usage of Cryptographic APIs. In: 32nd European Conference on Object-Oriented Programming

[4] Krüger, S.; Nadi, S.; Reif, M.; Ali, K.; Mezini, M.; Bodden, E. et al. (2017): CogniCrypt: Supporting Developers in Using Cryptography. In: 32Nd IEEE/ACM International Conference on Automated Software Engineering

[5] Christopher Giblin, Alice Y. Liu, Samuel Müller, Birgit Pfitzmann, and Xin Zhou. Regulations expressed as logical models (REALM). In Legal Knowledge and Information Systems - JURIX 2005: The Eighteenth Annual Conference on Legal Knowledge and Information Systems, Brussels, Belgium, 8-10 December 2005, volume 134 of Frontiers in Artificial Intelligence and Applications, pages 37–48. IOS Press, 2005

[6] Christopher Giblin, Samuel Müller, and Birgit Pfitzmann. From regulatory policies to event monitoring rules: Towards model-driven compliance automation.

[7] Matthew W. Sanders and Chuan Yue. Automated least privileges in cloud-based web services. In Qun Li and Songqing Chen, editors, Proceedings of the fifth ACM/IEEE Workshop on Hot Topics in Web Systems and Technologies, HotWeb 2017, San Jose / Silicon Valley, CA, USA, October 12 -14, 2017, pages 3:1–3:6. ACM, 2017

[8] Matthew W. Sanders and Chuan Yue. Mining least privilege attribute based access control policies. In David Balenson, editor, Proceedings of the35th Annual Computer Security Applications Conference, ACSAC 2019, San Juan, PR, USA, December 09-13, 2019, pages 404–416. ACM, 2019.

[9] Philipp Stephanow, Christian Banse: Evaluating the performance of continuous test-based cloud service certification. CCGrid 2017: 1117-1126

[10] Philipp Stephanow, Mohammad Moein, Christian Banse: Continuous Location Validation of Cloud Service Components. CloudCom 2017: 255-262

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Zur Security Awareness bei Nutzern von Smartphones

Philipp Goldberg[1]

Kurzfassung:
Immer mehr Menschen nutzen ein Smartphone. Mittlerweile ist die Anzahl von Nutzern auf gut 3,2 Milliarden angestiegen. Das Thema Sicherheit spielt bei einer derartigen Verbreitung eine wichtige Rolle. Doch wie gut sind die Nutzer über mögliche Risiken und Schutzmaßnahmen informiert? Wie hat sich dies im Laufe der Jahre entwickelt und mit welchen Methodiken wurde dies gemessen? Welche Anstrengungen in der Forschung haben dazu beigetragen, die Security Awareness zu verbessern?
Zur Beantwortung dieser Fragen wird die vorhandene Literatur mit Fokus auf die Security Awareness von Nutzern von Smartphones durchsucht. Hierzu wird ein strukturiertes Literaturreview für die Jahre 2011 bis 2020 durchgeführt. Es werden insgesamt 46 relevante Arbeiten analysiert. Hierbei werden die gängigen Methodiken und Erkenntnisse synthetisiert. Um die zukünftige Forschung zu unterstützen, erfolgt die Auflistung gängiger Limitationen und Hinweise zur weiteren Ausrichtung der Forschungsaktivitäten.

Stichworte: Awareness, Literaturreview, Security Awareness, Sicherheitsbewusstsein, Smartphones, Strukturiertes Literaturreview

1. Einleitung und Motivation

Die Zahl der Smartphone-Nutzer steigt Jahr für Jahr an, auf nunmehr gut 3,2 Milliarden [1]. Durch die zunehmende Verbreitung der Geräte wird es immer wichtiger, dass die Nutzer dabei nicht unnötigen Gefahren ausgesetzt werden. Für den sicheren Umgang mit IT-Ressourcen ist die Security Awareness der Nutzer von Bedeutung. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) sieht in seinem IT-Grundschutz-Kompendium [2] vor, dass Mitarbeiter/Nutzer die Sicherheitsziele kennen, dass sie bereit sind diese umzusetzen und fähig sind, in sicherheitskritischen Situationen angemessen zu reagieren.

In seinem Lagebericht zur IT-Sicherheit in Deutschland für das Jahr 2014 merkt das BSI an, dass Smartphones mit ihren Betriebssystemen gut gesichert sind. Eine weitere Komponente für die IT-Sicherheit seien aber auch die Nutzer. Hier wird angemerkt, dass beispielsweise die Berechtigungsabfragen der Apps von den Nutzern in den allermeisten Fällen unreflektiert bestätigt werden. [3]

Hat sich diese Situation geändert? Sind die Nutzer nun besser für den sicheren Umgang mit ihren Smartphones gerüstet? Um diese Fragen zu beantworten und auch zu evaluieren, worauf der Fokus und die Erkenntnisse aus der Forschung liegen, wird in dieser Arbeit ein strukturiertes Literaturreview (SLR) durchgeführt.

[1] Fachbereich Informatik, Hochschule Darmstadt, mail@pgoldberg.de

2. Zielsetzung und Forschungsfragen

Ziel dieser Arbeit soll es sein, einen Überblick darüber zu erhalten, was sich in der Forschung im Bereich der Security Awareness mit Fokus auf Smartphones und deren Nutzer getan hat. Hierfür soll die in den letzten Jahren betriebene Forschung analysiert werden. Neben der Beantwortung der Forschungsfragen soll ein Mehrwert für die zukünftige Forschung generiert werden. So erfolgt die Nennung typischer Limitationen der untersuchten Arbeiten und die Erstellung einer Forschungsagenda.

In der Fragestellung sind 2 Themenschwerpunkte enthalten. Zum einen das Thema Security Awareness, zum anderen auch die Nutzer von Smartphones. Hier stellen sich folgende Fragen:

- Welche Ansätze gibt es, um die Security Awareness zu erhöhen bzw. zu verbessern?
- Mit welchen Methodiken wurde die Security Awareness gemessen?
- Wie wird die Gültigkeit der Aussagen belegt? Welche typischen Limitierungen gibt es?
- Wie hat sich die Security Awareness über die Jahre verändert?

3. Begriffsdefinitionen

In diesem Kapitel werden einige für diese Arbeit wichtige Begriffe definiert.

3.1. Security Awareness

Die gängigen Begriffsdefinitionen zu Security Awareness (deutsch: Sicherheitsbewusstsein) des US-Amerikanischen National Institute of Standards and Technology (NIST) [4] und des BSI [2] nennen ähnliche Punkte. Diese lassen sich auf das Zusammenspiel der Aspekte *Können*, *Wollen* und *Wissen* reduzieren, wie beispielsweise Helisch [5] in seinem Buch erkannt hat.

3.2. Smartphone

Eine allgemeingültige Definition, was ein Smartphone genau ist und was es ausmacht, scheint es aktuell nicht zu geben [6]. Vielmehr gibt es mehrere Definitionen, welche mehr oder minder ähnliche Elemente haben. Hierbei handelt es sich um die Erläuterungen des Duden [7], des Gabler Wirtschaftslexikon [8] und der International Telecommunication Union (ITU) [6].

Nimmt man die Kernpunkte zusammen, so sind die Bedienung über einen Touchscreen, ein modernes/komplexes Betriebssystem, dem eine Vielzahl an Sensoren (beispielsweise GPS) zur Verfügung stehen und die Möglichkeit der Funktionserweiterung mittels Apps, welche durch den Nutzer installiert werden können, Mindestvoraussetzungen für ein Smartphone.

3.3. Strukturiertes Literaturreview

Das Ziel eines SLR ist es, bereits vorhandenes Wissen in Form von relevanten Publikationen zu finden. Es soll durch die Synthese zentraler Aussagen zusammengefasst werden. Weiterhin werden die bisher unternommenen Untersuchungen und Fragestellungen

genannt. Neben der Erstellung einer Forschungsagenda werden die aufgetretenen Limitierungen und methodischen Probleme genannt. Somit können diese bei der weiteren Forschung vermieden werden.

4. Vorgehensweise

In dieser Arbeit wird das Rahmenwerk von Jan vom Brocke et al. [9] genutzt. In ihrer Arbeit gehen sie der Frage nach, ob Literatur Reviews im Bereich der Informatik nachvollziehbar und systematisch durchgeführt werden. Das Ergebnis der Arbeit ist eine Anleitung zur Durchführung eines solchen SLR. Dies war einer der Beweggründe, diese Vorgehensweise zu wählen. Diese teilt sich in 5 aufeinanderfolgende Teilschritte auf, siehe Abbildung 1. Für den generellen Aufbau und die Vorgehensweise wurden auch bereits vorhandene Reviews als Orientierung genutzt, etwa [10].

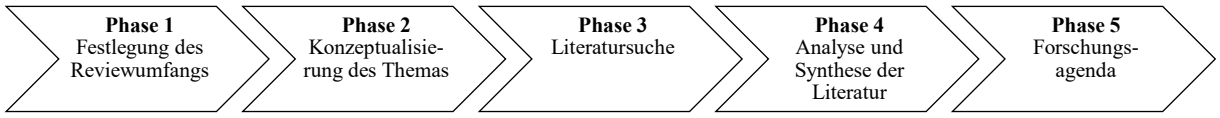


Abbildung 1: Phasen des SLR nach Vom Brocke et al. Quelle: in Anlehnung an [9]

Phase 1: In dieser Phase werden grundlegende Entscheidungen für die Art und Weise des SLR nach der Taxanomie von Cooper [11] festgelegt und in Fettschrift hervorgehoben, siehe Tabelle 1.

Merkmal	*Ausprägung*			
Fokus	**Forschungsergebnisse**	**Forschungsmethoden**	Theorien	Anwendungen
Ziel	**Integration**	Kritisieren	**Herausforderungen**	
Perspektive	**Neutrale Wiedergabe**		Standpunkt vertreten	
Abdeckung	**vollständig**	vollständig, selektive Zitation	repräsentativ	zentral bzw. grundlegend
Organisation	historisch	**konzeptionell**	**methodisch**	
Zielgruppe	**Fachleute**	**Wissenschaft**	**Praktiker**	Allgemeinheit

Tabelle 1: Taxanomie dieser Arbeit in Anlehnung an [11]

Phase 2: Hier wird eine Konzeptkarte erstellt. Diese enthält die Kernpunkte dieser Arbeit: *Security Awareness*, *Smartphone* und *User* (Nutzer). Zu allen Begriffen werden beispielsweise Synonyme und weitere verwandte Wörter gesucht und gruppiert.

Phase 3: Suche von Arbeiten und Prüfung auf Relevanz. Die Literatursuche lässt sich wiederum in 4 Subphasen einteilen:

- **Relevante Journale identifizieren**: Computers & Security, Springer Lecture Notes in Computer Science, …
- **Datenbanken suchen**, welche mindestens die oben genannten Journale beinhalten: ACM DL, IEEE DL und dblp.

- **Schlüsselwörter (Keywords) zum Durchsuchen** der Datenbanken nutzen. Diese stammen aus Phase 2.
- **Rückwärts- und Vorwärtssuche**

Phase 4: Nun werden die für relevant befundenen Arbeiten analysiert und synthetisiert. Jede Publikation wird (mehrfach) gelesen, um wichtige Informationen zu extrahieren. Neben dem Forschungsziel der Arbeit wird beispielsweise notiert, wie die Informationen erhoben wurden, wer die Probanden waren und welche Schlüsse gezogen wurden. Um ähnliche Arbeiten miteinander vergleichen zu können, müssen diese gruppiert werden. Dies geschieht mit einer Konzeptmatrix nach Webster und Watson [12].

Phase 5: Schlussendlich erfolgt die Erstellung einer Forschungsagenda. Sie gibt Hinweise darauf, welche Themengebiete weiter Forschung benötigen und auf welche Punkte (methodisch) geachtet werden sollte. Weiterhin werden die eingangs erwähnten Forschungsfragen beantwortet.

5. Durchführung der Literatursuche

Aus den Begriffen der Konzeptkarte (Phase 2) wird eine Suchanfrage formuliert. Sie ist iterativ entstanden, um eine Balance zwischen einer zu breiten respektive zu eingeschränkten Suche zu finden. Weiterhin wurde darauf geachtet, dass die Ergebnismenge für eine Person handhabbar bleibt. Eine Vielzahl der Publikationen der Informatik ist auf Englisch verfasst. Aus diesem Grund wird auch die Suchanfrage auf Englisch formuliert. Diese enthält Variationen der Kernpunkte, welche mit UND verknüpft werden. Etwaige Synonyme und verwandte Begriffe sind mit ODER verbunden:

(smartphone ODER android ODER ios ODER iphone ODER mobile phone) UND (user ODER participant ODER student ODER employee ODER individual) UND (security awareness ODER user awareness ODER personal information ODER information security)

Dieser Suchtext wird für die Suchfunktion jeder Datenbank angepasst. Die primäre Suche wurde am 15.06.2020 durchgeführt und führte zu 506 Ergebnissen. Durch die inhaltliche Relevanzprüfung, die Anwendung der Ein- und Ausschlusskriterien aus Tabelle 2 und die Anwendung weiterer Qualitätskriterien wurden insgesamt 46 Publikationen für die Folgeschritte identifiziert.

Merkmal	*Einschluss*	*Ausschluss*
Veröffentlichung	01.01.2011 bis 31.05.2020	Anderer Zeitraum
Publikationstyp	Vollständiges Paper	Extended Abstracts, Poster, …
Zugriff	Öffentlich zugänglich oder von der Hochschule lizenziert	Kostenpflichtige Artikel
Sprache	Englisch	Andere Sprachen
Fragestellung	Forschungsfragen und ähnliche	Andere Fragestellungen
Nutzereinbindung	Nutzer werden eingebunden	Keine Nutzereinbindung

Tabelle 2: Ein- und Ausschlusskriterien für das SLR (Auszug)

Einige der Ein- und Ausschlusskriterien werden nachfolgend begründet.

Veröffentlichung: Smartphones gibt es bereits seit der Einführung des Apple iPhone im Jahr 2007 [13], was aber nicht bedeutet, dass es seit diesem Zeitpunkt relevante Publikationen geben muss. Das behördliche und somit auch öffentliche Interesse hat im Jahr 2011 stark zugenommen, siehe [14]–[16]. Somit kann ab diesem Zeitpunkt von der Zunahme relevanter Forschungsbeiträge ausgegangen werden. Da diese Arbeit im Juni 2020 begonnen hat, werden nur zuvor veröffentlichte Publikationen betrachtet.

Nutzereinbindung: Im Titel dieser Arbeit werden die Nutzer von Smartphones explizit genannt. Aus diesem Grund sollten auch sämtliche Publikationen den Nutzer einbinden. Dies kann auf vielfältige Weise geschehen. Durch die Einbeziehung des Nutzers können die vorgestellten Verbesserungen oder Aussagen praktisch validiert werden.

Um ähnliche Arbeiten im nächsten Kapitel miteinander vergleichen zu können, müssen diese zuerst gruppiert werden. Parallel zur vorangegangenen Lektüre aller Publikationen wurde eine Konzeptmatrix nach Webster und Watson [12] erstellt. Die Konzepte werden in die beiden Kategorien *Ziel* und *Untersuchungsgegenstand* unterteilt. Als Ziel bzw. Intention einer Publikation konnten die Punkte *Unterstützung des Nutzers*, *Aufklärung und Schulung des Nutzers* sowie *Messen und Analysieren* identifiziert werden. Der Untersuchungsgegenstand bzw. Fokus kann auf den Bereichen *Berechtigungen von Apps, Risikobewertung von Apps*, *Privatsphäre*, *Security Awareness*, *App-Beschreibung* und/oder *Schutz des Smartphones* liegen.

6. Synthese der Arbeiten

Welche Forschungsfragen die Arbeiten bearbeitet haben und zu welchen Schlussfolgerungen sie gekommen sind, ist Teil dieses Kapitels. Zuerst werden die Arbeiten nochmals, basierend auf der Konzeptmatrix, eingeteilt. Dieser Schritt ist notwendig, da die Matrix sehr feingranular ist. Die erneute Einteilung basiert deshalb auf einer höheren Abstraktionsebene. So werden die übergeordneten Ziele bzw. adressierten Probleme identifiziert. Eine eineindeutige Zuordnung ist nicht immer möglich, da jede Arbeit auch (Teil-) Aspekte der übrigen Kategorien ansprechen kann. Die nachfolgenden Abschnitte sind nach dem Hauptziel der Forschungsarbeiten eingeteilt. Diese enthalten wiederum mehrere Unterpunkte. Im nachfolgenden Abschnitt geht es beispielsweise um Arbeiten, welche ihr Hauptaugenmerk auf dem Thema *Berechtigungen von Apps* haben. Innerhalb dieses Kontextes konnten weitere Gruppierungen vorgenommen werden. So wurde beispielsweise das *Verständnis der Nutzer* zum Thema Berechtigungen von Apps untersucht.

6.1. Berechtigungen von Apps

Für den Betrieb benötigen Apps verschiedene Berechtigungen, um auf gewisse Ressourcen zugreifen zu dürfen. Die Arbeiten in diesem Abschnitt befassen sich mit dem Themengebiet der Berechtigungen von Apps.

Verständnis der Nutzer: Wird dieses Berechtigungssystem von den Nutzern verstanden? In einigen Studien wird die Effektivität des Berechtigungsdialogs zur Installationszeit untersucht. Fast alle Android-Nutzer klicken den Dialog weg [17]. Viele Anwender

können nicht erklären, was sich hinter einer Berechtigung verbirgt ([17]–[19]) und welche Möglichkeiten eine Anwendung hierdurch hat.

Zu viele Berechtigungen: Furini et al. [20] nennen ein weiteres Indiz dafür, dass das Berechtigungssystem nicht verstanden wird. Die Apps der Probanden besitzen in vielen Fällen weit mehr Berechtigungen, als sie für die Ausführung ihrer Kernaufgabe benötigen. Ein Beispiel hierfür liefert Alani [21]. Fast 40 % seiner Befragten haben eine Taschenlampen-App installiert, welche insgesamt über 20 Berechtigungen einfordert. Hierunter auch solche, um den Inhalt des Speichers zu lesen. Derartige Berechtigungen werden laut Alani meist für Werbezwecke genutzt. Gratis-Apps haben im Vergleich zur Bezahlvariante meist mehr Berechtigungen, mit denen personenbezogene Daten ausgelesen werden können [22].

Hilfe bei Berechtigungen: Ein Teil der Arbeiten möchte dem Nutzer im Hinblick auf die von einer App geforderten Berechtigungen helfen. Diese Arbeiten betrachten fast alle Smartphones mit Android als Betriebssystem. Eine Vielzahl von ihnen (beispielsweise [23]–[25]) erfordert für die korrekte Funktionsweise Anpassungen am Betriebssystem selbst oder dem von Google betriebenen Play Store ([26], [27]). Die Entscheidung, welche Berechtigung entzogen respektive gewährt wird, kann vom Framework auf zwei unterschiedliche Arten getroffen werden. [23]–[25], [28] und [29] setzen zur Lösungsfindung auf **Crowdsourcing**, also die Intelligenz der Masse an Nutzern. Diese werden in verschiedene Privatsphäre-Profile eingeteilt. Somit sollen die Vorschläge für die Einstellungen von Nutzern mit ähnlichen Vorstellungen bezüglich der Privatsphäre kommen, um die Akzeptanz zu erhöhen. Agarwal und Hall haben in ihrer Arbeit untersucht, ob es einen Unterschied macht, ob ein initiales Set an Experten die Vorschläge macht oder diese gänzlich durch Crowdsourcing entstehen. Die Ergebnisse sind in beiden Fällen identisch. Ähnlich gute Vorschläge lassen sich auch **lokal** erzeugen, indem der Nutzer einige Fragen beantwortet. So werden sie bei [30] und [31] gefragt, für welche App-Kategorie (Art von App) sie welche Berechtigungen plausibel finden. [32] erstellen anhand von allgemeineren Fragen ein Privatsphäre-Profil. So wird den Nutzern aufgezeigt, welche Anwendungen ihren eigenen Vorstellungen widersprechen. Alle Apps unterbreiten dem Anwender daraufhin passende Vorschläge oder setzen diese teils automatisiert um.

6.2. Auswahl von Apps

Der Nutzer kann sein Smartphone um weitere Apps aus den App Stores für Android (Google Play Store) und iOS (Apple App Store) erweitern. Welche Ansätze es gibt, um dem Nutzer bei der Wahl einer App mit Fokus auf Sicherheit und Datenschutz zu helfen, ist Teil dieses Abschnittes. Die Notwendigkeit solcher Maßnahmen wird in Abschnitt 6.3 nochmals explizit aufgefasst. Neben der Begrenzung der Berechtigungen zur Laufzeit (vgl. vorherigen Abschnitt) setzen einige Arbeiten einen Schritt vorher an, nämlich bei der Auswahl einer App.

Berechtigungen und deren Auswirkungen visualisieren: Die Motivation dieser Arbeiten ist identisch: Die benötigten Berechtigungen werden entweder nicht verstanden oder ignoriert. Kelley et al. [27] rücken die benötigten Berechtigungen in den Fokus,

indem sie diese direkt auf der Detailseite einer App im Google Play Store anzeigen. Zudem wandeln sie die Liste an Berechtigungen, welche teils technisch klingen, in eine Checkliste um. So soll auch der Vergleich mehrerer Anwendungen vereinfacht werden. Harbach et al. [26] belassen die Anzeige dort wo sie ist (nach dem Klick auf Download), erweitern diese aber. Sie zeigen die Auswirkungen einer Berechtigung an. Beim Zugriff auf den Speicher wird beispielsweise ein Bild aus der Galerie des Nutzers gezeigt. Somit wird ihm verdeutlicht, welche Tragweite die einzelnen Berechtigungen haben. Benton et al. [33] versuchen Ähnliches. Auch sie passen den Dialog an, erweitern ihn aber mit mehr Text. Dies zeigt allerdings keinen statistisch signifikanten Effekt. Eine visuelle Darstellung wie von [34] ist effektiver.

Beschreibung anpassen: Wu et al. [35] passen die Beschreibung für vom Nutzer als kritisch empfundene Berechtigungen an seine Personalität an. Die so generierten Texte sind verständlicher und reicher an Informationen für den Nutzer. Zhang et al. [36] analysieren das Verhalten der App und geben diese (technischen) Informationen zusätzlich zur vom Entwickler bereitgestellten Beschreibung an den Nutzer weiter. Die Forscher geben an, dass durch ihren Text die tatsächliche Funktion der App beschrieben wird. Beide Arbeiten geben an, dass sie dabei helfen können, maliziöse Apps zu identifizieren.

Anleitung für die Wahl einer App: In [37] wird ein Flyer entworfen, welcher Hinweise darauf gibt, auf was man bei der Wahl im Google Play Store achten sollte. Er informiert die Leser weiterhin darüber, dass auch ihre Daten von Interesse sind und nicht nur diejenigen von bekannten Persönlichkeiten. Neben der Plausibilitätsprüfung der Berechtigungen sollte man zusätzlich auf weitere Apps des Entwicklers und das Datum der letzten Aktualisierung der App, zwecks Sicherheitsupdates, achten. Die Entscheidungen der Probanden werden dadurch nicht zwingend besser, aber fundierter.

Risiko einer App darstellen: Eine weitere Hilfestellung ist die Berechnung und Visualisierung des Risikos einer App. Es wird im Store bei der Auswahl von Apps angezeigt. Hiermit sollen ähnliche Anwendungen vergleichbar werden. Alle Darstellungsvarianten nutzen als Basis für die Berechnung die Anzahl kritischer Berechtigungen. [38] und [39] nutzen eine Skala zur Kommunikation des Risikos. [22] nutzt einen numerischen Wert. Zusätzlich wird bei einem Update im Google Play Store angezeigt, ob die App nun mehr oder weniger in die Privatsphäre eingreift. Ungeachtet der Art der Darstellung geben alle Autoren an, dass sich die Anwender für die bessere Anwendung entscheiden. Zumeist ist das diejenige mit weniger sensitiven/gefährlichen Berechtigungen bei gleicher oder ähnlicher Funktionalität.

6.3. Security Awareness der Nutzer

Die Erhebung der Daten erfolgt in fast allen Fällen subjektiv durch die Nutzung von Fragebögen oder der Durchführung von Interviews. Lediglich 2 Arbeiten ([40] und [41]) erfassen die benötigten Daten automatisiert und objektiv. Die Messung konzentriert sich in den meisten Fällen darauf, ob die Nutzer gängige Sicherheitsmechanismen kennen und auch anwenden. Einige in vielen Arbeiten abgefragte Punkte werden nachfolgend erläutert.

Zugriffsschutz: Ein simpler Schutz ist die Einrichtung einer Displaysperre, um den Zugriff auf das Smartphone und dessen Daten vor unbefugten Dritten zu schützen. Laut Breitinger et al. [42] nutzen 93 % der Generation Z/Y einen Lockscreen. Er merkt weiterhin an, dass solche Maßnahmen eher ergriffen werden als der Schutz der Privatsphäre. Gut 7 Jahre zuvor nutzten lediglich rund 60 % der Nutzer einen Lockscreen (siehe [43] und [44]).

Fehlendes Wissen über Schutzmechanismen: Ein oft genannter Punkt ist, dass den Nutzern das Wissen über die Existenz bestimmter Schutzmaßnahmen oder empfohlener Verhaltensweisen fehlt (beispielsweise [45]–[48]). Vecchiato und Martins [41] fordern von den Herstellern sicherere Standardeinstellungen. Denn laut ihnen haben selbst Personen, welche viele Einstellungen korrekt getätigt haben, weniger als die Hälfte aller Empfehlungen umgesetzt. Die Tendenz, dass IT-affine Personen die notwendigen Einstellungen eher kennen und anwenden, bestätigen auch Watson und Zheng [49]. Sie fordern, die Nutzer gezielt über solche Vorkehrungen zu informieren.

Benutzbarkeit muss bedacht werden: Dass eine Schutzmaßnahme nicht genutzt wird, kann verschiedene Gründe haben. Einige Befragte geben an, dass die Anpassungen im Alltag lästig sind [50]. Andere finden, dass diese eine schlechte Benutzbarkeit aufweisen [45], [51]. Sie geben an, dass sie solche Schutzmechanismen nutzen würden, wenn deren Benutzbarkeit steigt. In der Studie von Alsaleh et al. [46] halten einige Probanden die Nutzung solcher Mechanismen für Zeitverschwendung, da die Daten auf dem Gerät nicht relevant sind. Weiterhin wird angegeben, dass einem noch nie das Smartphone entwendet wurde und man aus diesem Grund beispielsweise auf einen Lockscreen verzichten kann.

Virenschutzprogramme: Ein im Zuge einiger Umfragen oft gefragter Punkt ist, ob die Nutzer auf ihrem Smartphone eine Antiviren-Software installiert haben. Dies ist bei wenigen Befragten der Fall. Die Forscher sind über dieses Ergebnis verwundert. Viele Teilnehmer geben an, dass sie wichtige oder gar sensible Daten auf dem Smartphone speichern, vgl. [42]. Weiterhin seien sie um die Sicherheit ihrer Daten besorgt, vgl. [51]. Die Nutzung von Antivirus-Apps ist dennoch eher gering (vgl. [42], [43], [52] und andere). Während Computer in der Regel mit einem Virenschutz ausgestattet sind, ist dies bei Smartphones eher die Ausnahme (beispielsweise [44]).

Sicherheitsempfinden von PC und Smartphone: Die Anwender schätzen Smartphones im Vergleich zum PC weniger sicher ein (beispielsweise [45]). Sie nennen auch mögliche Gründe hierfür. So seien PCs schon länger auf dem Markt und die Technik ausgereift und bekannt. Zudem seien vielen die Möglichkeiten von Smartphones auch mit Hinblick auf deren Rechenleistung nicht bekannt. Banking-Angelegenheiten werden laut Chin et al. bewusst am PC erledigt [53]. Sozialversicherungsnummer, Gesundheitsdaten und ähnlich sensible personenbezogene Daten werden ungern am Smartphone eingegeben, anders als am PC, so die Studie weiter. Gründe hierfür seien unter anderem die Angst vor einem Verlust des Gerätes.

Auch wenn die Nutzer mögliche Gefahren kennen, so handeln sie nicht immer entsprechend sicher. Einige Studierende aus der Studie von Jones et al. nutzen Online-Banking, haben aber nicht zwingend einen Lockscreen gesetzt. [54]

Auswahl von Apps: In Bezug auf die Security Awareness fragen viele Studien ihre Teilnehmer, nach welchen Kriterien sie sich für oder gegen eine App entscheiden. Für viele stehen die Punkte Sicherheit und Privatsphäre nicht im Mittelpunkt [44]. Vielmehr spielt die Empfehlung von Bekannten ([27]) oder die Bewertungen und Beliebtheit sowie Werbeanzeigen ([27], [53]) einer App eine primäre Rolle bei der Wahl. Viele nehmen fälschlicherweise an, dass es in den offiziellen App-Stores keine bösartigen Apps gibt (beispielsweise [18], [51]). Die Dialoge, welche die Berechtigungen einer App auflisten (seit Android Marshmallow gibt es diese Dialoge nicht mehr), werden von den meisten Nutzern ignoriert und helfen ihnen somit nicht bei der Entscheidung für oder gegen eine App [17], [18], [43]. Liccardi et al. [22] geben an, dass die reine Anzahl an Berechtigungen dazu führen kann, dass eine App nicht genutzt wird. Es sind allerdings nicht alle Berechtigungen eine Gefahr für die Sicherheit oder Privatsphäre. Selbst Apps mit legitimen Gründen für die Nutzung bestimmter Ressourcen werden nicht installiert.

Einflüsse und Zusammenhänge: Laut Bitton et al. [40] lassen demografische Eigenschaften keinen Rückschluss auf die Security Awareness zu. Sie haben aber einen Zusammenhang zwischen der Installation eines Virenscanners und der Security Awareness identifiziert. Sprache und Kultur haben keinen Einfluss [44]. Auch Bagga et al. [51] konnten keine signifikanten Unterschiede zwischen diversen Altersgruppen feststellen, ähnlich auch Jones et al. [54]. Wobei Jones et al. einräumen, dass sie nur eine Altersgruppe (17 bis 24 Jahre) untersucht haben. Diese Aussagen werden aber nicht von allen Autoren geteilt. So verhalten sich jüngere Menschen laut Alsaleh et al. [46] und Parker et al. [55] tendenziell sicherer. Weiterhin widersprechen Parker et al. der These, dass Sprache, Ethnie und Alter keinen Einfluss haben. Sie geben allerdings zu bedenken, dass deren Probanden meist junge Nutzer von Android-Smartphones sind und das demografische Profil somit nicht ausgewogen sei. Personen mit mittlerem Interesse an Cybersicherheit weisen sichereres Verhalten auf [42]. So erstellen diese beispielsweise häufiger Backups, was im Schadensfall hilfreich sein kann. Einige Studien (beispielsweise [46], [51], [55]) geben an, dass Männer risikoreicher handeln und Frauen nicht alle (tief in den Einstellungen versteckten) Sicherheitsmechanismen aktivieren. Die einzige Studie mit Kindern [56] zeigt ein umgekehrtes Bild. Hier sind die weiblichen Probanden sorgloser im Umgang mit ihrem Smartphone.

6.4. Aufklärung und Schulung der Nutzer

Arbeiten aus dieser Kategorie haben zum Ziel, den Nutzer aufzuklären.

Häufigkeit von Zugriffen: Schlegel et al. [34] visualisieren dem Nutzer am Beispiel seines Standortes, wie oft Anwendungen auf diese Ressource zugreifen. Trotz legitimer Gründe kann eine gehäufte Abfrage die Privatsphäre der Anwender verletzen. Durch die Visualisierung der Häufigkeit kann der Nutzer entscheiden, ob er diese Information weiterhin teilen möchte.

Auswirkungen von Berechtigungen: Meist bleibt es dem Anwender verborgen, welche Informationen übertragen werden. Eling et al. [19] zeigen mit ihrer App auf, welche Daten durch die gewährten Berechtigungen gelesen werden können. In ihrer Arbeit zeigt sich erneut, dass die Nutzer die Tragweite der Berechtigungen nicht kennen oder diese gar nicht erst gelesen haben. Zeigt die App an, auf welche Informationen sie zugreifen möchte, lehnen gut 60 % diese Anfrage ab. Dies kommt laut den Autoren einer Ablehnung der durch die Installation gewährten Berechtigung gleich. Dies bedeutet, dass der Nutzer nicht (mehr) mit seiner Entscheidung einverstanden ist oder schlicht nicht wusste, dass diese Informationen gelesen und versendet werden können. Furini et al. [20] möchten den Nutzer ebenfalls über Berechtigungen aufklären. Auch hier zeigte sich, dass die Vorstellungen der Nutzer von den Möglichkeiten der Berechtigungen abweichen. Zusätzlich analysieren sie die installierten Apps und kommen zu dem Schluss, dass viele Berechtigungen nicht zur primären Funktion der App benötigt werden. Sie fordern, den Nutzer weiter aufzuklären, damit er seine Privatsphäre schützen kann.

Dedizierte Lern-Apps: Bahrini et al. [57] klären den Nutzer spielerisch über Einstellungen zum Schutz seiner Privatsphäre auf. Die zumeist jungen (25 Jahre) Probanden haben durch die App weniger Neues gelernt, da Android in den letzten Jahren viel zur benutzbaren Sicherheit beigetragen hat, so die Autoren. Ähnliches machen Gerber et al. [58]. Sie wollen den Nutzer allerdings dazu animieren, stetig die App zum Lernen zu benutzen. Dies wollen sie mit einem Belohnungssystem erreichen.

Nudging: Almuhimedi et al. [59] weisen den Nutzer auf den in Android 4.3 bis 4.4.2 integrierten (experimentellen) Berechtigungsmanager hin. Durch Nudging bekommen sie regelmäßig Hinweise, wie oft eine App in der letzten Zeit welche Ressource genutzt hat. Der reine Hinweis auf den Berechtigungsmanager hat viele Nutzer dazu gebracht, diesen zu nutzen. Ähnlich verfahren Liu et al. [30]. Auch sie zeigen dem Nutzer regelmäßig an, welche App welche Zugriffe fordert und wollen ihn so stetig dazu animieren, die erteilten Berechtigungen zu prüfen.

Nutzer zu sichererem Verhalten animieren: Wie man die Nutzer dazu bewegen kann, gängige Sicherheitsmechanismen einzusetzen wurde auch untersucht. Van Bruggen et al. [60] wollen durch gezielte Nachrichten zur Nutzung einer Displaysperre animieren. Die verschiedenen Arten von Nachrichten (abschreckend, moralische oder eine Belohnung) haben keinen wirklichen Erfolg erzielt. Die Autoren geben an, dass der Aufwand nicht im Verhältnis zum Ergebnis steht. Albayram et al. [50] zeigen den Nutzern ohne Lockscreen ein Video, welches die möglichen Folgen eines ungeschützten Smartphones zeigen. Die Aufklärung über mögliche Risiken hat gut die Hälfte der Testgruppe zum Umdenken bewegt.

7. Diskussion der Arbeiten

Aus den Limitierungen und Forschungslücken aller Publikationen können Erkenntnisse für die zukünftige Ausrichtung und Fokussierung der Forschung gewonnen werden. Weiterhin werden jene Arbeiten oder Konzepte identifiziert, welche einen praktischen Einfluss hatten.

7.1. Limitierungen und Forschungslücken der Arbeiten

Ziel dieses Abschnittes ist es, die über alle Arbeiten hinweg aufgefallenen Limitierungen (ob explizit genannt oder durch die Analyse erkannt) und Forschungslücken aufzuzeigen. Wären einige dieser Einschränkungen nicht vorhanden, würde sich das mutmaßlich positiv auf die Forschung auf dem Themengebiet auswirken.

Verlass auf die korrekte Aussage von Befragten: Dass es einen Unterschied zwischen dem genannten Verhalten (subjektiv) der Probanden und deren tatsächlichem Handeln (objektiv) gibt, haben Bitton et al. [40] in ihrer Arbeit herausgefunden. Sie haben zuerst das geplante Handeln erfragt und anschließend in einem mehrwöchigen Experiment am Gerät des Nutzers das tatsächliche Handeln automatisiert erfasst. Sie kommen zu dem Schluss, dass die objektiven Werte präziser als die Selbsteinschätzung sind. Als Grund nennen sie, dass die Angaben auf Fragebögen entweder falsch sind oder die Frage falsch verstanden wurde. Dieser Punkt wird auch von [54] als mögliche Limitierung genannt.

Vernachlässigung von iOS: Von den analysierten Arbeiten hat nur [25] iOS-Nutzer betrachtet. Alle anderen Veröffentlichungen beschäftigen sich mit Android-Nutzern. Dies wird zumeist damit begründet, dass die Mehrheit Android nutze (beispielsweise [39], [58] und [23]). Bei allgemeineren Befragungen waren dennoch iOS-Nutzer vertreten. Die Vernachlässigung bezieht sich daher zumeist auf Arbeiten, welche eine praktische Implementierung vornehmen.

Kurze Studienzeit: Weiterhin fällt der Untersuchungszeitraum vieler Arbeiten kurz aus. Das haben beispielsweise [26] in ihrer Arbeit genannt und verweisen darauf, dass hiermit nur belegt werden kann, dass der Nutzer das vermittelte Wissen im Kurzzeitgedächtnis gespeichert hat. Ob man sein Verhalten langfristig ändern konnte, sei ungewiss. Es kann auch nicht ausgeschlossen werden, dass die Probanden sich durch die Forschungssituation anders verhalten als sonst (beispielsweise [38]). Die kurze Zeit zum Sammeln von Daten nennen auch [30] als mögliche Einschränkung.

Nicht repräsentative Wahl der Probanden: Viele der untersuchten Arbeiten, welche sich nicht explizit mit jungen Studierenden beschäftigen wollen, taten dies durch die Auswahl der Probanden dennoch (beispielsweise [58]). Einige dieser Arbeiten nennen die Probandenwahl ausdrücklich als limitierenden Faktor. Anders sieht es bei Veröffentlichung von [45] und anderen aus, welche gezielt junge Menschen im Fokus haben. Es sollte nicht nur im universitären Umfeld nach Teilnehmern gesucht werden, denn diese seien laut [39] zumeist höher gebildet. Aus diesem Grund sind die Aussagen meist nicht generalisierbar. Nur wenige Arbeiten treffen die Aussage, dass die Wahl der Probanden in etwa der Nutzerschicht von Android entspricht, etwa [53].

Beschränkung auf technisch versierte Nutzer: Ein ähnliches Problem ergibt sich dadurch, dass die Experimente nur mit einer eingeschränkten Gerätebasis stattfinden können. So müssen beispielsweise für einige Arbeiten die Geräte gerootet sein [25] oder eine spezielle Version des Betriebssystems genutzt werden [59]. Ein regulärer Nutzer wird somit ausgeschlossen, da dies in der Regel nur ein meist auf Sicherheit spezialisierter Nutzerkreis tut und diese meist mehr Wissen in Bezug auf Sicherheit vorweisen können [30].

Fehlende Alltagstauglichkeit: Viele Arbeiten machten sinnvolle Vorschläge, wie man das Sicherheitsbewusstsein der Nutzer verbessern oder ihnen helfen kann. Es sind einige Anwendungen, wie *Protect my privacy* [25], entstanden und erfolgreich mit echten Nutzern erprobt worden. Viele dieser Ideen haben sich aber mutmaßlich nicht durchgesetzt. Das mag am oben genannten Problem liegen, dass man für die Installation selbst Expertenwissen benötigt. Neben den Änderungen am Betriebssystem werden oft Anpassungen an Komponenten gefordert, auf die man keinen direkten Einfluss hat, etwa die Oberfläche des Google Play Store.

7.2. Einflussreiche Forschungsbeiträge

Die Vorschläge, Konzepte und Kritiken einiger Arbeiten haben seit ihrer Veröffentlichung einen praktischen Einfluss gehabt. Ob die Änderungen beispielsweise am Betriebssystem tatsächlich auf diese Arbeiten zurückzuführen sind, ist ungewiss. Dennoch dürfte es nicht abwegig sein, dass auch die Entwickler von Google (Android) und Apple (iOS) diese Publikationen gelesen haben und dadurch zu einigen Anpassungen motiviert wurden.

Verbessertes Berechtigungskonzept bei Android: Felt et al. [17] unterbreiten in ihrer Arbeit von 2012 den Vorschlag, dass Berechtigungsabfragen zur Laufzeit besser wahrgenommen werden könnten. Eling et al. [19] merken ebenfalls in ihrer Arbeit an, dass (feingranulare) Berechtigungsabfragen zur Laufzeit für den Nutzer nützlicher sind, als jene zum Installationszeitpunkt. Mit Android 6 (Marshmallow) hält diese Funktion erstmals offiziell Einzug in das Betriebssystem[2]. Somit muss der Nutzer nicht mehr, wie in früheren Versionen üblich, bereits mit der Installation allen Berechtigungen zustimmen. Weiterhin ist es seit dieser Android-Version möglich, erteilte Berechtigungen auch wieder zu entziehen. Dies haben beispielsweise die Arbeiten im Abschnitt 6.1 (Hilfe bei Berechtigungen) bereits ermöglicht.

Prüfung von Anwendungen durch App-Stores: In einigen Arbeiten gaben die Nutzer an, dass sie annehmen, dass es in den offiziellen App-Stores von Apple und Google keine schädlichen Apps gibt. Die Autoren dieser Arbeiten geben an, dass es einen solchen Mechanismus nicht gibt (beispielsweise [18], [51]). Google hat seitdem viel getan, um schädliche Anwendungen im Google Play Store zu identifizieren [61]–[63]. Diese Maßnahmen gehen alle in die richtige Richtung. Auch Apple gibt an, dass diverse Sicherheitsmechanismen dafür sorgen, dass Apps frei von bekannten Schadprogrammen sind [64]. Allerdings kann die Zuverlässigkeit dieser Mechanismen angezweifelt werden, da es immer wieder Beiträge über den Fund von Schadsoftware gibt. Laut einem Blogartikel von Avast aus dem Oktober 2020 wurden 21 Malware-Apps (Adware) im Google Play Store gefunden [65]. Die Sicherheitsforscher von Wandera haben im Oktober 2019 17 maliziöse Apps im App Store von Apple gefunden [66].

[2] https://www.android.com/intl/de_de/versions/marshmallow-6-0/ (besucht am 27.08.2020)

8. Fazit und Forschungsagenda

In diesem letzten Kapitel wird diese Arbeit nochmals reflektiert. Zuerst wird ein Fazit gezogen. Anschließend werden die Forschungsfragen beantwortet und basierend auf den vorherigen Kapiteln eine Empfehlung für die zukünftige Forschung gegeben.

8.1. Fazit

Diese Arbeit hat mithilfe der Durchführung eines SLR und der Analyse von 46 Arbeiten den Stand der Forschung auf dem Themengebiet der Security Awareness von Smartphone-Nutzern aufgezeigt. Es wurden die verschiedenen Forschungsschwerpunkte beleuchtet. Einige Schwächen zeigten sich bei der Wahl der Probanden, der Art der Datenerhebung und der Vernachlässigung von iOS-Nutzern. Meist lag der Fokus der Publikationen auf den Themengebieten Berechtigungen von Apps, Auswahl von Apps, Security Awareness der Nutzer und Aufklärung und Schulung der Anwender. Einen neuralgischen Punkt stellt das Berechtigungssystem von Android dar. Den Nutzern war es meist nicht bewusst, auf welche Daten eine Anwendung durch die erteilten Berechtigungen zugreifen darf. Wie im vorherigen Abschnitt erwähnt, wurden einige Vorschläge aus den Arbeiten in die Betriebssysteme integriert. Weiter ist aufgefallen, dass den Nutzern das Wissen für einen sicheren Umgang mit ihrem Smartphone fehlt oder sie ihre Daten nicht als schützenswert ansehen. In der Synthese hat sich ergeben, dass hier bessere und sicherere Standardwerte bei einigen Einstellungen hilfreich sein können. Hiervon profitieren alle Nutzer.

8.2. Beantwortung der Forschungsfragen

Auf die Fragen wurde in den vorangegangenen Kapiteln implizit eingegangen. Diese Erkenntnisse werden nun explizit genannt.

Welche Ansätze gibt es, um die Security Awareness zu erhöhen bzw. zu verbessern? Im Bereich Berechtigungen versuchen viele Arbeiten den Nutzer zu entlasten, indem sie Aufgaben automatisieren. Einen anderen Ansatz bilden Apps, welche den Nutzer spielerisch an das Thema Sicherheit heranführen. Meist genügt es auch, mögliche Risiken aufzuzeigen und Lösungsmöglichkeiten zu nennen, wie man diese effektiv verhindert.

Mit welchen Methodiken wurde die Security Awareness gemessen? Zumeist wurde auf Fragebögen oder ähnliche Methoden gesetzt. Hierbei müssen sich die Probanden selbst einschätzen. Einen allgemeinen Standard oder gar eine Skala gibt es mutmaßlich nicht. Das Messen der Security Awareness durch die Auswertung des tatsächlichen Verhaltens über einen längeren Zeitraum scheint die sinnvollste Variante zu sein (vgl. [40]).

Wie wird die Gültigkeit der Aussagen belegt? Welche typischen Limitierungen gibt es? Meist werden die getroffenen Aussagen, etwa zur Wirksamkeit einer App, mit einem Nutzertest belegt. Vielfach werden auch Fragebögen benutzt oder Interviews durchgeführt. Typische Limitierungen bilden meist die Anzahl und die Auswahl der Probanden oder die Korrektheit der getätigten Aussagen und Angaben. Weiterhin bilden viele Studien nur eine (kurze) Momentaufnahme ab. Länger angesetzte Studien bilden die Ausnahme.

Wie hat sich die Security Awareness über die Jahre verändert? Wie bereits bei der zweiten Forschungsfrage angesprochen gibt es keine einheitliche Skala. Somit ist die Beantwortung dieser Frage nicht trivial. An einigen Eckpunkten aus Abschnitt 6.3 lässt sich jedoch eine Tendenz ablesen. Ein in vielen Studien abgefragter Aspekt ist, ob eine Displaysperre gesetzt ist. Dieser ist über die Jahre von rund 60 % auf 93 % (Generation Z/Y) gestiegen. Hier ist ein positiver Trend zu erkennen. Bahrini et al. [57] ziehen in ihrer Arbeit das Fazit, dass die Betriebssysteme vermehrt den Fokus auf benutzbare Sicherheit legen. Damit begründen sie, dass ihre Probanden durch ihre App nicht viel Neues gelernt haben, da sie bereits über dieses Wissen verfügen.

8.3. Forschungsagenda

Aus den vorherigen Kapiteln lassen sich einige Empfehlungen für zukünftige Arbeiten auf diesem Gebiet ableiten. So sollten bei der Durchführung von Studien möglichst viele Daten automatisiert erhoben werden, um deren Aussagekraft zu erhöhen. Eine manuelle bzw. subjektive Erfassung durch die Nutzer kann zusätzlich erfolgen, um deren geplantes mit dem tatsächlichen Verhalten zu vergleichen. Weiterhin sollte die Auswahl von Probanden mehr an die tatsächlichen demografischen Rahmenbedingungen der Nutzer von Smartphones angepasst werden. Hierbei kann es zusätzlich von Interesse sein, verschiedene Nationalitäten zu betrachten, um etwaige regionale Unterschiede aufzudecken. Einige Arbeiten mussten aus der Betrachtung ausgeschlossen werden, da diese den Nutzer nicht involvierten. So haben einige Arbeiten plausible Konzepte und Ansätze geliefert, diese aber nicht durch die Nutzer verifiziert. Allgemein sollten bei Entwicklungen für eine bestimmte Zielgruppe immer Tests mit dieser durchgeführt werden, um zu überprüfen, ob diese auch alltagstauglich sind. Für die Feststellung der Security Awareness von Nutzern sollte ein allgemeingültiges Rahmenwerk erarbeitet werden. Dieses sollte so gestaltet werden, dass die Ergebnisse miteinander vergleichbar sind. Hierdurch kann die Veränderung der Security Awareness über die Jahre oder auch der Vergleich von verschiedenen Gruppen oder Nationalitäten einfacher erfolgen. Hierzu könnte das von Bitton et al. [40] vorgeschlagene automatisierte Framework zur Berechnung des Information Security Awareness Scores als Basis dienen.

Literaturhinweise

[1] „Newzoo's Global Mobile Market Report: Insights into the World's 3.2 Billion Smartphone Users, the Devices They Use & the Mobile Games They Play", *Newzoo*. https://newzoo.com/insights/articles/newzoos-global-mobile-market-report-insights-into-the-worlds-3-2-billion-smartphone-users-the-devices-they-use-the-mobile-games-they-play/ (zugegriffen Okt. 16, 2020).

[2] Bundesamt für Sicherheit in der Informationstechnik, „IT-Grundschutz-Kompendium". Zugegriffen: Okt. 16, 2020. [Online]. Verfügbar unter: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Kompendium/IT_Grundschutz_Kompendium_Edition2020.pdf?__blob=publicationFile&v=6.

[3] Bundesamt für Sicherheit in der Informationstechnik, „Die Lage der IT-Sicherheit in Deutschland 2014". Dez. 15, 2014, Zugegriffen: Okt. 21, 2020. [Online]. Verfügbar unter: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2014.pdf?__blob=publicationFile&v=2.

[4] M. Wilson und J. Hash, „Building an Information Technology Security Awareness and Training Program", National Institute of Standards and Technology, NIST Special Publication (SP) 800-50, Okt. 2003. doi: https://doi.org/10.6028/NIST.SP.800-50.

[5] M. Helisch, „Definition von Awareness, Notwendigkeit und Sicherheitskultur", in *Security Awareness: Neue Wege zur erfolgreichen Mitarbeiter-Sensibilisierung*, M. Helisch und D. Pokoyski, Hrsg. Wiesbaden: Vieweg+Teubner, 2009, S. 9–28.

[6] Fredrik Eriksson, „Proposal for a definition of smartphone", *International Telecommunication Union (ITU)*, Aug. 28, 2017. https://www.itu.int/en/ITU-D/Statistics/Documents/events/egh2017/EGH%202017%20background%20document%203%20-%20Definition%20of%20smartphone.pdf (zugegriffen Okt. 21, 2020).

[7] „Duden | Smartphone | Rechtschreibung, Bedeutung, Definition, Herkunft". https://www.duden.de/rechtschreibung/Smartphone (zugegriffen Okt. 21, 2020).

[8] P. D. I. Sjurts, „Definition: Smartphone", Feb. 19, 2018. https://wirtschaftslexikon.gabler.de/definition/smartphone-52675/version-275793 (zugegriffen Okt. 21, 2020).

[9] J. Brocke *u. a.*, „RECONSTRUCTING THE GIANT: ON THE IMPORTANCE OF RIGOUR IN DOCUMENTING THE LITERATURE SEARCH PROCESS", *ECIS 2009 Proceedings*, Jan. 2009, [Online]. Verfügbar unter: https://aisel.aisnet.org/ecis2009/161.

[10] M. Tahaei und K. Vaniea, „A Survey on Developer-Centred Security", in *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, Juni 2019, S. 129–138, doi: 10.1109/EuroSPW.2019.00021.

[11] H. M. Cooper, „Organizing knowledge syntheses: A taxonomy of literature reviews", *Knowledge in Society*, Bd. 1, Nr. 1, S. 104–126, März 1988, doi: 10.1007/BF03177550.

[12] J. Webster und R. T. Watson, „Analyzing the Past to Prepare for the Future: Writing a Literature Review", *MIS Quarterly*, Bd. 26, Nr. 2, S. xiii–xxiii, 2002.

[13] „Apple erfindet mit dem iPhone das Mobiltelefon neu", *Apple Newsroom*. https://www.apple.com/de/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/ (zugegriffen Nov. 02, 2020).

[14] Bundesamt für Sicherheit in der Informationstechnik, „Die Lage der IT-Sicherheit in Deutschland 2011“. Mai 31, 2011, Zugegriffen: Okt. 16, 2020. [Online]. Verfügbar unter: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2011.pdf?__blob=publicationFile&v=3.

[15] Bundesamt für Sicherheit in der Informationstechnik, „Wie sicher sind Smartphones?“, Feb. 04, 2011. https://www.bsi-fuer-buerger.de/BSIFB/DE/Service/Aktuell/Informationen/Artikel/aeltere_Artikel/Smartphones_24022011.html (zugegriffen Mai 18, 2020).

[16] „Smartphones: Information security risks, opportunities and recommendations for users“. https://www.enisa.europa.eu/publications/smartphones-information-security-risks-opportunities-and-recommendations-for-users (zugegriffen Okt. 15, 2020).

[17] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, und D. Wagner, „Android permissions: user attention, comprehension, and behavior“, in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, Washington, D.C., Juli 2012, S. 1–14, doi: 10.1145/2335356.2335360.

[18] P. G. Kelley, S. Consolvo, L. F. Cranor, J. Jung, N. Sadeh, und D. Wetherall, „A conundrum of permissions: installing applications on an android smartphone“, in *Proceedings of the 16th international conference on Financial Cryptography and Data Security*, Bonaire, März 2012, S. 68–79, doi: 10.1007/978-3-642-34638-5_6.

[19] N. Eling, S. Rasthofer, M. Kolhagen, E. Bodden, und P. Buxmann, „Investigating Users' Reaction to Fine-Grained Data Requests: A Market Experiment“, in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Jan. 2016, S. 3666–3675, doi: 10.1109/HICSS.2016.458.

[20] M. Furini, S. Mirri, M. Montangero, und C. Prandi, „Privacy perception and user behavior in the mobile ecosystem“, in *Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good*, Valencia, Spain, Sep. 2019, S. 177–182, doi: 10.1145/3342428.3342690.

[21] M. Alani, „Android Users Privacy Awareness Survey“, *International Journal of Interactive Mobile Technologies (iJIM)*, Bd. 11, S. 130, Apr. 2017, doi: 10.3991/ijim.v11i3.6605.

[22] I. Liccardi, J. Pato, D. J. Weitzner, H. Abelson, und D. De Roure, „No technical understanding required: helping users make informed choices about access to their personal data“, in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, London, United Kingdom, Dez. 2014, S. 140–150, doi: 10.4108/icst.mobiquitous.2014.258066.

[23] R. Liu, J. Cao, L. Yang, und K. Zhang, „PriWe: Recommendation for Privacy Settings of Mobile Apps Based on Crowdsourced Users' Expectations“, in *2015 IEEE International Conference on Mobile Services*, Juni 2015, S. 150–157, doi: 10.1109/MobServ.2015.30.

[24] B. Rashidi, C. Fung, A. Nguyen, T. Vu, und E. Bertino, „Android User Privacy Preserving Through Crowdsourcing“, *IEEE Transactions on Information Forensics and Security*, Bd. 13, Nr. 3, S. 773–787, März 2018, doi: 10.1109/TIFS.2017.2767019.

[25] Y. Agarwal und M. Hall, „ProtectMyPrivacy: detecting and mitigating privacy leaks on iOS devices using crowdsourcing“, in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, Taipei, Taiwan, Juni 2013, S. 97–110, doi: 10.1145/2462456.2464460.

[26] M. Harbach, M. Hettig, S. Weber, und M. Smith, „Using personal examples to improve risk communication for security & privacy decisions“, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, Ontario, Canada, Apr. 2014, S. 2647–2656, doi: 10.1145/2556288.2556978.

[27] P. G. Kelley, L. F. Cranor, und N. Sadeh, „Privacy as part of the app decision-making process“, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris, France, Apr. 2013, S. 3393–3402, doi: 10.1145/2470654.2466466.

[28] Q. Ismail, T. Ahmed, A. Kapadia, und M. K. Reiter, „Crowdsourced Exploration of Security Configurations“, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea, Apr. 2015, S. 467–476, doi: 10.1145/2702123.2702370.

[29] J. Lin, B. Liu, N. Sadeh, und J. I. Hong, „Modeling users' mobile app privacy preferences: restoring usability in a sea of permission settings“, in *Proceedings of the Tenth USENIX Conference on Usable Privacy and Security*, Menlo Park, CA, Juli 2014, S. 199–212, Zugegriffen: Juli 08, 2020. [Online].

[30] B. Liu *u. a.*, „Follow my recommendations: a personalized privacy assistant for mobile app permissions“, in *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security*, Denver, CO, USA, Juni 2016, S. 27–41, Zugegriffen: Juli 08, 2020. [Online].

[31] Y. Jing, G.-J. Ahn, Z. Zhao, und H. Hu, „Towards Automated Risk Assessment and Mitigation of Mobile Applications“, *IEEE Transactions on Dependable and Secure Computing*, Bd. 12, Nr. 5, S. 571–584, Sep. 2015, doi: 10.1109/TDSC.2014.2366457.

[32] C. B. Jackson und Y. Wang, „Addressing The Privacy Paradox through Personalized Privacy Notifications“, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Bd. 2, Nr. 2, S. 68:1-68:25, Juli 2018, doi: 10.1145/3214271.

[33] K. Benton, L. J. Camp, und V. Garg, „Studying the effectiveness of android application permissions requests“, in *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, März 2013, S. 291–296, doi: 10.1109/PerComW.2013.6529497.

[34] R. Schlegel, A. Kapadia, und A. J. Lee, „Eyeing your exposure: quantifying and controlling information sharing for improved privacy“, in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, Pittsburgh, Pennsylvania, Juli 2011, S. 1–14, doi: 10.1145/2078827.2078846.

[35] T. Wu *u. a.*, „Catering to Your Concerns: Automatic Generation of Personalised Security-Centric Descriptions for Android Apps“, *ACM Trans. Cyber-Phys. Syst.*, Bd. 3, Nr. 4, S. 36:1-36:21, Sep. 2019, doi: 10.1145/3317699.

[36] M. Zhang, Y. Duan, Q. Feng, und H. Yin, „Towards Automatic Generation of Security-Centric Descriptions for Android Apps“, in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, Denver, Colorado, USA, Okt. 2015, S. 518–529, doi: 10.1145/2810103.2813669.

[37] O. Kulyk, P. Gerber, K. Marky, C. Beckmann, und M. Volkamer, „Does This App Respect My Privacy? Design and Evaluation of Information Materials Supporting Privacy-Related Decisions of Smartphone Users“, gehalten auf der Workshop on Usable Security, San Diego, CA, 2019, doi: 10.14722/usec.2019.23029.

[38] C. S. Gates, J. Chen, N. Li, und R. W. Proctor, „Effective Risk Communication for Android Apps“, *IEEE Transactions on Dependable and Secure Computing*, Bd. 11, Nr. 3, S. 252–265, Mai 2014, doi: 10.1109/TDSC.2013.58.

[39] J. Kang, H. Kim, Y. G. Cheong, und J. H. Huh, „Visualizing Privacy Risks of Mobile Applications through a Privacy Meter“, in *Information Security Practice and Experience*, Cham, 2015, S. 548–558, doi: 10.1007/978-3-319-17533-1_37.

[40] R. Bitton, K. Boymgold, R. Puzis, und A. Shabtai, „Evaluating the Information Security Awareness of Smartphone Users“, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, Apr. 2020, S. 1–13, doi: 10.1145/3313831.3376385.

[41] D. Vecchiato und E. Martins, „Experience report: A field analysis of user-defined security configurations of Android devices“, in *2015 IEEE 26th International Symposium on Software Reliability Engineering (ISSRE)*, Nov. 2015, S. 314–323, doi: 10.1109/ISSRE.2015.7381824.

[42] F. Breitinger, R. Tully-Doyle, und C. Hassenfeldt, „A survey on smartphone user’s security choices, awareness and education“, *Comput. Secur.*, Bd. 88, 2020, doi: 10.1016/j.cose.2019.101647.

[43] A. Mylonas, A. Kastania, und D. Gritzalis, „Delegate the smartphone user? Security awareness in smartphone platforms“, *Computers & Security*, Bd. 34, S. 47–66, Mai 2013, doi: 10.1016/j.cose.2012.11.004.

[44] J. Ophoff und M. Robinson, „Exploring end-user smartphone security awareness within a South African context“, in *2014 Information Security for South Africa*, Aug. 2014, S. 1–7, doi: 10.1109/ISSA.2014.6950500.

[45] V. Gkioulos, G. Wangen, S. K. Katsikas, G. Kavallieratos, und P. Kotzanikolaou, „Security Awareness of the Digital Natives“, *Information*, Bd. 8, Nr. 2, Art. Nr. 2, Juni 2017, doi: 10.3390/info8020042.

[46] M. Alsaleh, N. Alomar, und A. Alarifi, „Smartphone users: Understanding how security mechanisms are perceived and new persuasive methods“, *PLOS ONE*, Bd. 12, Nr. 3, S. e0173284, März 2017, doi: 10.1371/journal.pone.0173284.

[47] R. C. Jisha, R. Krishnan, und V. Vikraman, „Mobile Applications Recommendation Based on User Ratings and Permissions“, in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Sep. 2018, S. 1000–1005, doi: 10.1109/ICACCI.2018.8554691.

[48] I. Androulidakis und G. Kandus, „Mobile Phone Security Awareness and Practices of Students in Budapest“, in *Proceedings of the 6th International Conference on Digital Telecommunications*, 2011, S. 17--22, Zugegriffen: Juli 08, 2020. [Online].

[49] B. Watson und J. Zheng, „On the User Awareness of Mobile Security Recommendations“, in *Proceedings of the SouthEast Conference*, Kennesaw, GA, USA, Apr. 2017, S. 120–127, doi: 10.1145/3077286.3077563.

[50] Y. Albayram, M. M. H. Khan, T. Jensen, und N. Nguyen, „“...better to use a lock screen than to worry about saving a few seconds of time”: Effect of Fear Appeal in the Context of Smartphone Locking Behavior“, 2017, S. 49–63, Zugegriffen: Juli 16, 2020. [Online]. Verfügbar unter: https://www.usenix.org/conference/soups2017/technical-sessions/presentation/albayram.

[51] T. Bagga, J. Sodhi, B. Shukla, und M. A. Qazi, „SMARTPHONE SECURITY BEHAVIOUR OF THE INDIAN SMARTPHONE USER“, *MAN IN INDIA*, S. 13, 2017.

[52] I. Androulidakis und G. Kandus, „Differences in users' state of awareness and practices regarding mobile phones security among EU countries", in *Proceedings of the 5th WSEAS international conference on Communications and information technology*, Juli 2011, S. 294–300.

[53] E. Chin, A. P. Felt, V. Sekar, und D. Wagner, „Measuring user confidence in smartphone security and privacy", in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, Washington, D.C., Juli 2012, S. 1–16, doi: 10.1145/2335356.2335358.

[54] B. H. Jones, A. G. Chin, und P. Aiken, „Risky business: Students and smartphones", *TECHTRENDS TECH TRENDS*, Bd. 58, Nr. 6, S. 73–83, Nov. 2014, doi: 10.1007/s11528-014-0806-x.

[55] F. Parker, J. Ophoff, J.-P. V. Belle, und R. Karia, „Security awareness and adoption of security controls by smartphone users", in *2015 Second International Conference on Information Security and Cyber Forensics, InfoSec 2015, Cape Town, South Africa, November 15-17, 2015*, 2015, S. 99–104, doi: 10.1109/InfoSec.2015.7435513.

[56] N. Etaher und G. R. S. Weir, „Understanding children's mobile device usage", in *2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF)*, Juni 2016, S. 1–7, doi: 10.1109/ICCCF.2016.7740437.

[57] M. Bahrini, G. Volkmar, J. Schmutte, N. Wenig, K. Sohr, und R. Malaka, „Make my Phone Secure! Using Gamification for Mobile Security Settings", in *Proceedings of Mensch und Computer 2019*, Hamburg, Germany, Sep. 2019, S. 299–308, doi: 10.1145/3340764.3340775.

[58] N. Gerber *u. a.*, „FoxIT: enhancing mobile users' privacy behavior by increasing knowledge and awareness", in *Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust*, Orlando, Florida, USA, Dez. 2018, S. 53–63, doi: 10.1145/3167996.3167999.

[59] H. Almuhimedi *u. a.*, „Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging", in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea, Apr. 2015, S. 787–796, doi: 10.1145/2702123.2702210.

[60] D. Van Bruggen, S. Liu, M. Kajzer, A. Striegel, C. R. Crowell, und J. D'Arcy, „Modifying smartphone user locking behavior", in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, Newcastle, United Kingdom, Juli 2013, S. 1–14, doi: 10.1145/2501604.2501614.

[61] „Google ‚Bouncer' Now Scanning Android Market for Malware". https://uk.pcmag.com/mobile-apps/66697/google-bouncer-now-scanning-android-market-for-malware (zugegriffen Nov. 05, 2020).

[62] „Google Play Protect für den Schutz vor schädlichen Apps - Google Play-Hilfe". https://support.google.com/googleplay/answer/2812853?hl=de (zugegriffen Nov. 05, 2020).

[63] „Google Play Protect", *Android*. https://www.android.com/intl/de_de/play-protect/ (zugegriffen Nov. 05, 2020).

[64] „Sicherheit bei Apps – Übersicht", *Apple Support*. https://support.apple.com/de-de/guide/security/sec35dd877d0/web (zugegriffen Nov. 30, 2020).

[65] „New Malware Apps on Google Play | Avast". https://blog.avast.com/new-malware-apps-on-google-play-avast (zugegriffen Nov. 28, 2020).

[66] „Trojan malware infecting 17 apps on the App Store“, *Wandera*, Okt. 24, 2019. https://www.wandera.com/ios-trojan-malware/ (zugegriffen Nov. 30, 2020).

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Zukünftige harmonisierte Zertifizierung in Europa am Beispiel 5G, CC und IACS

Dr. Helge Kreutzmann [1]

Kurzfassung:
Der Beitrag schildert, wie der neue Rechtsrahmen "Cybersecurity Act" zur Harmonisierung der IT-Sicherheitszertifizierung in Europa aufgebaut ist, wie damit Schemata entstehen und wie dabei die Beteiligung der Experten erfolgt. Dieser Prozess wird an drei Beispielen, die in unterschiedlichen Stadien sind (5G/NESAS, Common Criteria und Industrielle Automatisierungstechnik) exemplarisch vorgestellt, wobei auf Besonderheiten der drei Schemata eingegangen wird. Der Beitrag endet mit einem Überblick über weitere Kandidaten und einem Ausblick in Form von offenen Fragen.

Stichworte: 5G, Common Criteria, CSA, EUCC, Europa, IACS, Zertifizierung

1. Motivation für die Cybersicherheitszertifizierung

Vertrauen in die Sicherheitsleistung von IT-Produkten, -Diensten und -Prozessen zu schaffen, ist für Anbieter und Kunden essenziell. In der einfachsten Variante versucht der Hersteller durch Eigenerklärungen dieses Vertrauen zu erreichen, was zwar einen geringen Aufwand für alle Beteiligten bedeutet, aber vom Kunden genau jenes Vertrauen bereits verlangt, was eigentlich erworben werden soll. Im anderen Extremfall prüft jeder Kunde jedes Produkt intensiv selbst (ggf. mit Audits beim Hersteller), was zwar zu sehr hohem Vertrauen führt, aber auch sehr hohen Aufwand bedeutet und in der praktischen Durchführung unrealistisch ist.

Daher hat sich die Zertifizierung von Produkten und Prozessen durch neutrale Drittparteien („Zertifizierungsstellen“) nach transparenten IT-Sicherheitskriterien am Markt bewährt. Durch die Überwachung der Zertifizierungsstellen (z.B. im Rahmen einer Akkreditierung) wird deren Unabhängigkeit von den Herstellern und ihre fachliche Qualifikation gewährleistet, sodass sie als stellvertretende Stelle die Produkte, Dienste und Prozesse prüfen (einmal, statt durch jeden) und das Einhalten der Kriterien durch ein Zertifikat bestätigen können, dem dann alle Kunden vertrauen können.

Damit das Zertifizierungsschema aber von allen Parteien akzeptiert wird, muss es dem Anbieter einen möglichst großen Markt bieten (d.h. eine Zertifizierung muss in einem möglichst großen Gebiet anerkannt werden) und zum Weiteren auch eine flexible Prüftiefe/ein flexibles Vertrauensniveau ermöglichen, denn ein Produkt für „Otto Normalverbraucher“ braucht im Allgemeinen nicht so hohe Prüftiefen wie ein Produkt, das zum Beispiel kritische Infrastrukturen absichern muss.

Um möglichst große Märkte im Bereich der IT-Sicherheitszertifizierung zu adressieren, wurde bisher für jeden Einzelfall auf bilaterale oder multilaterale Abkommen gesetzt.

[1] Bundesamt für Sicherheit in der Informationstechnik

So schließen beispielsweise die europäischen oder internationalen Akkreditierungsorganisationen (European Accreditation (EA) bzw. International Accreditation Foundation (IAF)) für ihre Mitglieder Verträge zur gegenseitigen Anerkennung, die für „populäre“ Zertifizierungsnormen wie beispielsweise die ISO/IEC 27001 [2] gelten. Andererseits gibt es für andere Normen wie die ISO/IEC 15408 („Common Criteria”) [3] sowohl europäische (SOG-IS) als auch internationale Abkommen (CCRA) zur gegenseitigen Anerkennung zwischen staatlichen Stellen. Wieder andere Gremien (z.B. die Global System for Mobile Communications (GSMA) für den Mobilfunkstandard 5G) entwickeln dazu eigene Strukturen/Stellen.

Problematisch hierbei ist, dass jedes Abkommen seine eigenen Regeln entwirft, die Vertrauenswürdigkeitsstufen (so vorhanden) individuell definiert und die Anerkennung nur für die dem jeweiligen Vertrag beigetretenen Länder anwendbar ist. Um dieses zu vereinheitlichen, hat die Europäische Union 2019 den „Cybersecurity Act“ [1] als Vorgabe für europäische Cybersicherheits-Zertifizierungsschemata entwickelt[2].

2. Der rechtliche Rahmen „Cybersecurity Act“

Im CSA wurde darauf verzichtet, konkrete Cybersicherheitszertifizierungsschemata zu beschreiben, stattdessen wurden Anforderungen an und die Entwicklung von europäischen Schemata geregelt.

Kern eines Schemas ist eine idealerweise internationale oder europäische Cybersicherheitsnorm, deren Einhaltung zertifiziert wird. Die Zertifizierung kann dabei durch akkreditierte Stellen für einen der drei Vertrauenswürdigkeitsstufen „niedrig“ (basic), „mittel“ (substantial) oder „hoch“ (high) erfolgen, wobei in letzterem Falle im Grundsatz staatliche, ansonsten privatwirtschaftliche Zertifizierungsstellen zuständig sind. Der CSA sieht auch die Möglichkeit vor, dass Hersteller die Bewertung bei der Vertrauenswürdigkeitsstufe niedrig selbst durchführen. Ob dies möglich ist und welche Vertrauenswürdigkeitsstufen unterstützt werden, muss im jeweiligen Schema definiert werden.

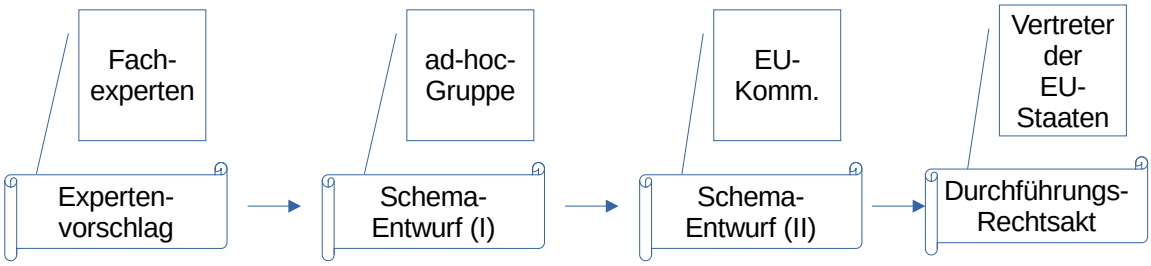
Für welche Normen, Bereiche oder Produktgruppen ein europäisches Schema entwickelt wird, legt der CSA wie erwähnt nicht fest, sondern definiert ein Verfahren, wie diese potenziellen Schemata ausgesucht und entwickelt werden. Standardmäßig wird dazu ein sogenanntes „Union Rolling Work Programme” (URWP) von der EU-Kommission aufgelegt, in dem mögliche Zertifizierungsschemata mit ihrer jeweiligen Priorität aufgeführt werden. Dieses Programm wird mit Hilfe von Industrievertretern in der „Stakeholder Cyber Certification Group” (SCCG) gepflegt. Darüber hinaus sind die Mitgliedstaaten in der „European Cyber Certification Group” (ECCG) vertreten und beraten die EU-Kommission in Fragen rund um den CSA.

Die EU-Kommission beauftragt dann die ENISA, eine „ad hoc“-Gruppe zur Erstellung eines Schemas zu gründen. Hierzu beruft die ausgewählte ENISA Experten nach einem fairen Verteilschlüssel aus den betroffenen Kreisen ein, die dann den Schemaentwurf

[2] Der CSA enthält zwei Teile. Neben dem hier beschriebenen Anteil begründet er auch ein neues und permanentes Mandat für die ENISA.

erstellen. Der Schema-Entwurf kann dann, ggf. nach weiterer Kommentierung, durch durchführende EU-Rechtsakte in Kraft gesetzt werden.

Das so initiierte Schema entfaltet anschließend europaweit Wirkung, d.h. nationale Zertifizierungs-systeme für diesen Bereich müssen zurückgezogen werden und die im Schema erteilten Zertifikate gelten unmittelbar in der ganzen EU. Eine detailliertere Übersicht über den CSA bietet auch [8].



Da also der CSA nur einige grundlegende Anforderungen sowie den Prozess der Erstellung der Schemata beschreibt, soll dies an drei Beispielen, die sich in unterschiedlichem Status befinden, beleuchtet werden.

3. Fallstudien

3.1. Fallstudie 1: 5G-Zertifizierung / NESAS

Das erste Beispiel ist die 5G-Zertifizierung basierend auf den GSMA-Vorgaben „Network Equipment Security Assurance Scheme (NESAS)". Die GSMA bietet derzeit bewusst keine Zertifizierung an, hat ihr eigenes Schema aber ansonsten an Zertifizierungsschema angelehnt. Die beiden Systeme (der Einfachheit halber GSMA-NESAS und NESAS-CSA genannt) werden im Folgenden vorgestellt und verglichen.

Die NESAS-Schema arbeitet in zwei Stufen. Zuerst wird der Entwicklungsprozess des Herstellers auditiert. Dabei wird geprüft, ob der Hersteller geeignete Entwicklungs- und Produktpflegeprozesse eingeführt hat und diese auch betreibt. Ist dies der Fall, so definiert der Auditor eine Liste von Nachweisen, die für die nachfolgenden Produktzertifizierungsprozesse benötigt werden. Diese Nachweise belegen dann, dass der Hersteller sich an die auditierten Prozesse gehalten hat. Da im Audit der Prozess und nicht der Standort oder beteiligte Personen betrachtet werden, kann bei global agierenden Herstellern das Audit auch für Standorte „am anderen Ende der Welt" (im Vergleich zum Ort der Auditierung) gelten, solange die Prozesse dort auch etabliert sind.

In der zweiten Stufe erfolgt die eigentliche Produktprüfung. Basis sind hierbei die SCAS-Dokumente der 3GPP, die je nach Produkt und Konfiguration zur Anwendung kommen (es ist hierbei die Funktionalität bzw. der Einsatzzweck entscheidend).

Stellt ein Hersteller nun einen Antrag, so muss er sich entweder einem Audit (Stufe 1) unterwerfen oder er muss, falls für diese Produktkategorie bei ihm bereits ein Audit nicht älter als 2 Jahre vorliegt, die definierten Nachweise für das konkrete Produkt beifügen. Damit kann nachvollzogen werden, dass das Produkt wie angegeben entwickelt wurde.

Das Schema enthält also als Besonderheit auch eine ausdrückliche Audit-Komponente, die unter Umständen sogar von einer anderen Zertifizierungsstelle (bzw. deren Auditoren) beigesteuert werden kann.

In Zusammenarbeit zwischen der GSMA und dem deutschen Bundesamt für Sicherheit in der Informationstechnik wurde basierend auf diesem Konzept ein Entwurf für NESAS-CSA erstellt, wobei das Ziel war, NESAS selbst so wenig wie möglich anzupassen.

Im Entwurf ist vorgesehen, dass diese beiden Komponenten (Audit und Produktprüfung) bei einer Zertifizierungsstelle zusammengeführt werden. Diese zertifiziert zum einen (optional) die Auditoren, die die Audits durchführen, und Produkte, wobei als Voraussetzung die Auditnachweise bzw. die Herstellernachweise für das Produkt die Anforderungen erfüllen müssen. Die eigentliche Produktprüfung erfolgt in akkreditierten und von der Zertifizierungsstelle lizenzierten Prüfstellen.

Eine weitere Besonderheit ist, dass der 5G-Schemaentwurf so ausgestaltet ist, dass der Prozess die simultane Erstellung von Zertifikaten sowohl im EU-Schema als auch die Verwendung der Ergebnisse im GSMA-Schema ermöglicht, wenn die beteiligten Stellen und Personen in beiden Schemata akkreditiert bzw. zertifiziert sind. Somit wird der Aufwand für Hersteller minimiert, die in beiden Welten ihr Ergebnis nutzen wollen. Auch ist angedacht, das NESAS-GSMA-Schema so weiterzuentwickeln, dass weitere lokale Zertifizierungsvarianten ermöglichen werden – damit wird dann auch die Frage der gegenseitigen Anerkennung relevant.

NESAS-CSA kennt nur die Zertifizierung gemäß Vertrauenswürdigkeitsstufe "basic", wobei angedacht ist, bei zukünftigen Überarbeitungen des Schemas eine Entwicklung Richtung der Vertrauenswürdigkeitsstufe „substantial“ vorzunehmen.

Eine Besonderheit stellt auch die Aufrechterhaltung der Zertifizierung bei Sicherheitskorrekturen („minor updates“) dar. Hier soll das Zertifikat gültig bleiben, ohne dass eine Reevaluierung erfolgen muss, da ja bereits der Prozess entsprechend auditiert ist. Bei größeren Änderungen (z.B. neue Funktionalitäten) ist weiterhin eine Rezertifizierung notwendig.

Die Entwicklung von NESAS-CSA hat auch bereits positive Auswirkungen auf das NESAS-GSMA. Der CSA verlangt von Schemata gewisse Grundanforderungen (siehe Artikel 51 des CSA [1]), die so nicht vollständig in NESAS-GSMA umgesetzt sind und im CSA-Schema-Entwurf ergänzt wurden. Beispielsweise hat die Forderung nach „Sicherheit durch Voreinstellungen“ („Secure by default“) bereits zu Überlegungen geführt, dies auch in NESAS-GSMA zu integrieren.

Neben der Frage, wann (formal auch noch ob) die ad-hoc-Gruppe für 5G eingesetzt wird, ist in diesem Schema sicherlich der Geltungsbereich noch abschließend zu diskutieren. Reicht „basic“ für Funkkomponenten aus? Hierfür kann das etablierte NESAS eingesetzt werden. Es gibt aber auch Forderungen, andere Produktbereiche (z.B. die SIM-Karte (UICC)) mit aufzunehmen oder z.B. die Prozesse bei den Mobilfunkbetreibern, was mit NESAS so nicht abbildbar ist. Auch stellt sich die Frage, ob ein Schema,

das sich ausschließlich auf die Vertrauenswürdigkeitsstufe „niedrig“ fokussiert, hinreichend zukunftssicher ist.

3.2. Fallstudie 2: Transposition von SOG-IS (Common Criteria)

Das zweite Beispiel ist das Schema für die „Common Criteria” (ISO/IEC 15408) [3]. Nach dieser Norm wird bereits weltweit durch staatliche Stellen gemäß des europäischen SOG-IS bzw. internationalen CCRA-Abkommens umfangreich zertifiziert, zudem gibt es nationale Varianten in weiteren Märkten wie Russland oder China.

Die CC bauen auf einer Normenserie auf, die seit mehr als 20 Jahren fortentwickelt wird (und selbst wieder auf langjährigen Normen beruht) und im Laufe des Jahres 2021 in einer neuen Version erscheinen wird. Aufgrund dieser Reife und Marktbedeutung hatte die EU-Kommission schon vor der Erstellung des URWP die ENISA beauftragt, eine ad-hoc-Gruppe für die Schema-Entwicklung einzuberufen, die ihr Ergebnis vorgestellt hat und das bereits mehrere anschließende Kommentierungsrunden durchlaufen hat.

Die Evaluierung und Zertifizierung von Produkten gemäß den CC basiert im Kern auf zwei Dimensionen. Zum einen wird die Sicherheitsleistung des Produktes in einer semiformalen Sprache beschrieben („Funktionale Sicherheitsanforderungen“ / „Security Functional Requirements“, SFR), die eine gute Modellierung der Sicherheitsleistung erlauben. Zum anderen wird die Bedrohungslage in einem sogenannten „Angriffspotenzial“ modelliert, bei der Faktoren wie Expertise, Werkzeug, Angriffsdauer eingehen, um die benötigte Vertrauenswürdigkeit in der Evaluierung (systematisch) zu ermitteln. Dazu gibt es 7 Standardvertrauenswürdigkeitsstufen („EAL“), die aber zudem individuell angepasst (augmentiert) werden können. So hat sich beispielsweise für Produkte mit hohem Schutzbedarf die Stufe EAL 4+ etabliert. Bei höheren Stufen erfolgt dann zudem eine (semi)formale Beschreibung der Sicherheitsleistung bis hin zu beweisbaren Modellen, was sehr aufwendig und nur bei sehr hohem Schutzbedarf angemessen ist.

Ein weiteres wichtiges Konzept ist die Entwicklung von „Vorlagen“ für ganze Produktklassen. Diese Vorlagen heißen Schutzprofile (PP) und beschreiben abstrakt, welche Sicherheitsleistung Produkte erfüllen müssen, beispielsweise zur Absicherung der Kommunikation. Dabei werden die erlaubten Parameter im PP definiert und dann in den konkreten Vorgaben für das jeweils zu zertifizierende Produkt in den Sicherheitsvorgaben (ST) konkretisiert, d.h. aus den Optionen die jeweils zutreffenden ausgewählt. Dies ermöglicht es Anwendern (z.B. Industriegremien oder staatlichen Stellen bei Beschaffungen für den hoheitlichen Bereich), ihre Sicherheitsleistung allgemein zu formulieren und dann bei der Zertifizierung die Erfüllung des PPs zu verlangen, ohne die Sicherheitsleistungen jedes einzelnen Produktes individuell vergleichen zu müssen. Die Hersteller wissen zudem genau, welche Sicherheitsleistungen für ihr Produkt in einem bestimmten Markt gefragt ist.

Der Schema-Entwurf [6] hat zwar das SOG-IS-Schema als Grundlage genommen, es aber in einigen Punkten fortentwickelt. Während die Struktur der CC eine Selbstevaluierung ausschließt, ist die zuständige ad-hoc-Gruppe auch zu dem Schluss gekommen, dass die sieben EAL-Stufen nur auf die zwei Vertrauenswürdigkeitsstufen „mittel“ und

„hoch“ abgebildet werden sollen, d.h. eine Zertifizierung nach „niedrig“ nicht möglich sein soll.

Zum Zeitpunkt der Beitragserstellung (Ende 2020) waren aber noch nicht alle Fragen abschließend geklärt, die wichtigsten werden nachfolgend angerissen.

Die Grenze zwischen „mittel“ und „hoch“ ist noch umstritten, d.h. ab welcher EAL-Stufe die CSA-Vertrauenswürdigkeitsstufe „hoch“ beginnt bzw. grundsätzlich staatliche Zertifizierungsstellen zuständig sind. Damit in Zusammenhang steht auch die Frage, wie mit sehr hohen EAL-Stufen (z.B. EAL 6) umgegangen werden soll, die derzeit grundsätzlich nicht der gegenseitigen Anerkennung des EUCC unterliegen. Ist hier eine nationale Anwendung auch jenseits etablierter technischer Domänen (das Konzept wurde aus SOG-IS übernommen) möglich?

Eine wesentliche Weiterentwicklung, allerdings nur „versuchsweise“, ist die Aufnahme von Patch-Management-Konzepten, um effizient die Zertifizierung aufrechtzuerhalten oder schnell wieder zu erlangen, wenn nach der Veröffentlichung aufgrund von Sicherheitsproblemen neue Produktversionen notwendig werden. Hier wird die Praxis zeigen müssen, ob und wenn ja welches der zwei möglichen Verfahren sich in der Praxis bewährt und vom Markt angenommen wird.

Eine interessante offene Frage ist auch die Migration der technischen SOG-IS-Strukturen, wie beispielsweise der JIL-Arbeitsgruppen. Diese könnten beispielsweise an die ECCG andocken, wären dann aber recht starr in ihrer Teilnehmerschaft. Oder sie könnten Information Sharing and Analysis Center sein, die eigenständig agieren können, um z.B. effizient mit Industrie- und Prüfstellenvertretern Angriffs- und Evaluierungsmethodiken weiterzuentwickeln. Allerdings stellt sich dann die Frage der Verbindlichkeit der Ergebnisse.

Jenseits des Schemas stellt sich auch die Frage, wie andere Schemata sich davon abgrenzen bzw. das EUCC-Schema nutzen, da die CC produktagnostisch („horizontal“) aufgebaut sind. So könnten beispielsweise SIM-Karten für 5G (siehe Abschnitt 3.1) genauso wie Kernkomponenten für industrielle Steuersysteme (siehe Abschnitt 3.3) auch nach den CC evaluiert und zertifiziert werden. Die Frage ist, wann (und nach welchen Regeln) vertikale Domänen wie 5G oder IoT eigene Evaluierungsmethodiken entwickeln können oder sollten.

Diese Fragen werden in den nächsten Monaten intensiv diskutiert und schließlich durch die Kommission in enger Abstimmung mit der ECCG geklärt, damit dann voraussichtlich Mitte 2021 die Durchführungsrechtsakte verabschiedet werden können. Spätestens dann wird auch klar, wie lange der Übergang von den nationalen bzw. dem europäischen SOG-IS-Schema zum EUCC-Schema erfolgt und ob dieser abrupt (Stichtag) oder graduell (z.B. nationale Reevaluierungen und Reassessments auch später noch möglich) eingeführt wird.

3.3. Fallstudie 3: Industrielle Automatisierung / Kritis

Das dritte Beispiel ist ein Schema-Entwurf für industrielle Steuerungssysteme. (IACS). „Industrielle Steuersysteme“ ist ein großer Bereich, der von Sensoren und Aktoren über

Maschinensteuerungen bis hin zu Leitsystemen Produkte in einer Vielzahl von Branchen, einschließlich kritischer Infrastrukturen, abdeckt. Auch der Energieerzeugungs- und -verteilsektor bis hin zu Smart Metern gehört dazu. Teilweise wird der Bereich auch mit „Industrial Internet of Things“ (IIoT) bezeichnet.

Aufgrund der großen Breite gibt es historisch gewachsen eine Reihe von Normen und Standards, mit der die Sicherheitsleistung beschrieben werden kann. Hinzu kommt, dass solche Systeme oft über längere Zeit historisch gewachsen sind, sodass die Sicherheitsleistung oft über das System erbracht werden muss, da einzelne Komponenten diese nicht erbringen können (bzw. auch über zu wenig Rechenleistung oder Energie dafür verfügen).

Speziell die Normenreihe IEC 62443 [4] wurde für solche Systeme entwickelt. Teil 1 betrachtet primär Grundlagen und Begriffe, Teil 2 fokussiert primär auf den Anlagenbetreiber, Teil 3 ist primär für den Integrator relevant und Teil 4 für die Produktsicherheit. IEC 62443-4-1 ist dabei für die Anforderungen an den Entwicklungsprozess, -4-2 für die Produktsicherheitsanforderungen zuständig. Was der Normenreihe derzeit fehlt ist eine Evaluationsmethodologie, sodass unklar ist, wie die bereits existierenden Zertifizierungen hier erfolgen.

In einigen Teilen des Bereichs werden auch die Common Criteria [3] eingesetzt, z.B. für Smart Meter (Gateways), wo es nationale und europäische Schutzprofile gibt. Schließlich wurden in einer europäischen Untersuchung auch penetrationsgestützte Evaluierungen mit festem Zeitbudget (wie die französische Zertifizierung CSPN) eingesetzt.

Aufgrund der großen gesellschaftlichen Relevanz hatte die EU-Kommission bereits 2014 (d.h. weit vor der Entwicklung des CSA) das „Joint Research Centre“ (JRC) beauftragt, ein europäisches Zertifizierungsverfahren für IACS zu entwerfen. Eine europäische Arbeitsgruppe mit Vertretern aus Industrie, Forschung und Regierungen entwickelte im Laufe der Jahre, u. a. über eine CSPN-basierte Machbarkeitsstudie, schließlich 2020 der Entwurf eines Schemas für IACS [5] gemäß den Anforderungen des CSA. Aufgrund der großen Produkt- aber auch Kritikalitätsbandbreite deckt er, anders als die beiden vorhergehenden Schemata, alle CSA-Vertrauenswürdigkeitsstufen inkl. der Selbstbewertung ab.

Ein weiterer Entwicklungsschwerpunkt war die notwendige Normenoffenheit (Agnostizität), da es wie beschrieben die Norm für IACS nicht gibt. Diese Entscheidung bedingt, dass der Schema-Entwurf abstrakt aber doch hinreichend genau beschreiben muss, welche Evaluierungstätigkeiten pro Vertrauenswürdigkeitsstufe des CSA durchgeführt werden müssen. Dazu gehört auch, dass wünschenswerte Konzepte aus den im Feld bereits verwandten Normen (z.B. das Konzept der Sicherheitsvorgaben und Schutzprofile, inkl. deren Evaluierung, aus den CC sowie die Beurteilung des Entwicklungsprozesses aus der IEC 62443-4-2 [4]) in abstrakter und sprachlich neutraler Form im Schemaentwurf wiedergegeben werden mussten, statt nur auf die Normen zu verweisen, wie dies bei anderen Schemata der Fall ist. Schließlich mussten Konzepte vorbereitet werden, um sicherzustellen, dass die Aussagen der Zertifizierung unabhängig

von der zugrundeliegenden Norm zu vergleichbaren Aussagen kommen. Diese Fragestellung beinhaltet auch die Kompetenzfeststellung der involvierten Personen (insb. Evaluatoren) und muss wesentlich auch noch bei der weiteren Schemabefassung in der ad-hoc-Arbeitsgruppe behandelt werden. Auch ist noch nicht klar, ob dieses Konzept in dem finalen CSA-Schema realisiert werden kann oder ob eine Festlegung auf ein (oder zwei) zugrundeliegende Normen nicht zwingend erfolgen muss.

Ein weiterer nicht abschließend behandelter Punkt ist die Frage der Aufrechterhaltung von Zertifizierungen. Der Entwurf sieht vor, dass Herstellerselbsterklärungen (einmalig) leicht auf neue Versionen ausgedehnt werden können, aber erkennt auch die Notwendigkeit an, für insbesondere Sicherheitsaktualisierungen eine effiziente Möglichkeit zu schaffen, das Zertifikat zu erhalten oder zeitnahe für die Version mit Patch wiederzuerlangen. Um dies zu ermöglichen, wird wahrscheinlich eine Kombination der Informationen aus dem Entwicklungsprozess gemäß IEC 62443-4-1 mit den Erfahrungen aus dem Patch-Management der CC (siehe Abschnitt 3.2) bei der Arbeit in der ad-hoc-Gruppe die Basis für die weitere Behandlung darstellen.

Es ist davon auszugehen, dass IACS zudem eines der nächsten Schemata wird, die in einer ad-hoc-Gruppe entwickelt werden wird, da im Entwurf des URWP IACS neben Produkten des „Internet of Things“ (IoT) als eine der zwei Top-Themen für nächste Schemata eingetragen ist. Es wird dabei auch eine spannende Frage, wie diese beiden Schemata (IoT und IACS/IIoT) genau voneinander abgetrennt werden.

3.4. Übersichtsvergleich

Zur besseren Vergleichbarkeit der Schemata sind einige Kernparameter in der nachfolgenden Tabelle aufgeführt:

	NESAS-CSA	EUCC	IACS
Formaler Stand	Im URWP-Entwurf angedeutet	Vorbereitung der delegierten Rechtsakte	Im URWP-Entwurf mit hoher Priorität geplant
Dokumentenstand	Vorschlag GSMA und BSI	Version 1.1 aus ad hoc-Gruppe und Kommentierung	Vorschlag aus europäischer JRC-Gruppe liegt vor
Horizontal/Vertikal	Vertikal	Horizontal	Vertikal mit Bezug zu horizontalen Schemata
Basis-Norm	SCAS der 3GPP	EN ISO/IEC 15408 [CC]	agnostisch, beispielhaft CC, IEC 62443, CSPN
CSA-Stufen	Basic	Substantial, high	alle
Selbsterklärung	Nein	Nein	Ja
Prüfstellen akkreditiert	ja	ja	nein
Steuerungsgremium?	ja (im Schema)	über ECCG	ja (im Schema)

4. Weitere Schemata und Ausblick

Derzeit wird neben den finalen Arbeiten am EUCC-Schema auch an dem Schema für Cloud-Zertifizierungen gearbeitet. Wesentliche Eingaben für dieses Schema sind das französische SECNUM Cloud-Schema [10] (für die Vertrauenswürdigkeitsstufe hoch) und die deutsche C5-Testierung [9] (mit Fokus auf der Vertrauenswürdigkeitsstufe „mittel“). Dabei entsteht das Problem, wie das „Upgrade“ eines Cloud-Anbieters effizient realisiert werden kann, da sich die beiden Systeme doch wesentlich unterscheiden. Wenn sich ein Cloud-Anbieter, der die Anforderungen der CSA-Vertrauenswürdigkeitsstufe „mittel“ erfüllt, entscheidet, die Stufe „hoch“ zu erreichen, sollten sich die Arbeiten auf die Bereiche konzentrieren, bei denen ein Mehr an Sicherheit notwendig ist, es sollte vermieden werden, dass er sämtliche technischen und prozeduralen Verfahren erneut überarbeiten muss, lediglich weil sich die Herangehensweisen der zugrundeliegenden Verfahren unterscheiden. Im aktuellen Entwurf [12] wurden die Anforderungen beider Konzepte herangezogen und die Evaluation basierend auf der DIN EN ISO/IEC 17021-1 [11] und die ISAE3402 [13] gemischt. (Bei Konzentration auf eine Norm wie beispielsweise die IEC 62443 oder die CC tritt dieses Problem nicht auf, da das Konzept in diese Normen konsistent adressiert ist).

Wie in Abschnitt 3.3 erwähnt, ist IoT neben IACS eines der zwei im URWP vorgeschlagenen nächsten Schemata. Hierzu gibt es bereits Vorarbeiten bei ETSI für Consumer IoT und die Vertrauenswürdigkeitsstufe niedrig. Es wird interessant zu sehen sein, ob hier ähnlich zum aktuellen Ansatz bei 5G ein stark fokussiertes Schema („besser eine Sache gut und gründlich statt alles irgendwie zu machen“) erstellt wird, oder ob, ggf. unter Zuhilfenahme von horizontalen Normen/Verfahren wie den CC oder der CSPN (bzw. einer europäisierten Variante davon) auch die anderen Vertrauenswürdigkeitsstufen des CSA adressiert werden.

Neben der Schema-Erstellung sind auch der Betrieb und die Fortentwicklung der Schemata eine wichtige Fragestellung, weil diese durch den CSA nicht vollständig festgelegt werden. Hierzu haben die einzelnen Schema-Entwürfe verschiedene Vorschläge erstellt, auch die ECCG hat sich hierzu bereits Gedanken gemacht. Das 5G- und IACS-Schema sehen eine operative Steuerungsgruppe auf Ebene der Mitgliedsstaaten vor, um beispielsweise neue Normenversionen zu akzeptieren oder technische Fragestellungen im Evaluations- und Zertifizierungsprozess zu klären. Das am weitesten fortgeschrittene EUCC-Schema wird sich nach aktueller Planung einer Mischung von weiterentwickelten Gremien aus der SOG-IS-Welt sowie von ECCG-Untergruppen bedienen. Wie genau diese weiterentwickelten Gremien organisiert werden, ist derzeit noch nicht abschließend geklärt.

Aus Sicht des Autors gibt es aber vor allem noch fundamentale Fragestellungen, die über das einzelne Schema hinausgehen, und noch einer Willensbildung und ggf. der Rechtsetzung bedürfen. Zwei davon sind in den nächsten Absätzen skizziert. Eine ergänzende Diskussion dazu findet sich auch in [7].

Die erste Frage ist die der Anzahl und dem Zuschnitt der CSA-Schemata. Technisch (Kapazität bei ENISA, Verfügbarkeit von Experten) können wahrscheinlich mittelfristig

1 bis 2 Schemata pro Jahr erstellt werden. Die Frage ist jedoch, wie viele werden benötigt? Horizontale Zertifizierungsnormen wie die CC, die Familie der zeitbegrenzten, penetrationsbasierten wie der CSPN oder ggf. auch perspektivisch die IEC 62443 können eine große Bandbreite an Produkten abdecken, analog ist dies auch für eine Prozess- und Dienstleistungszertifizierung denkbar. Gleichzeitig gibt es viele Wünsche für vertikale (produktbereichsbezogene) Zertifizierungsschemata. Wie wird entschieden, wann welches gilt bzw. ein solches entwickelt wird? Dürfen die (vertikalen) Schemata sich überlappen? Sollte es sich, wie im IACS-Beispiel auf ein oder mehrere horizontale Schemata abstützen? Und wenn ein Produkt in mehreren Bereichen (bspw. ein IoT-Produkt für IACS aber auch Medizintechnik) eingesetzt wird? Ist dann sichergestellt, dass ein Zertifikat der Vertrauenswürdigkeitsstufe „mittel“ in allen relevanten Schemata die gleiche Aussage hat? Vielleicht liefert ja die im IACS-Entwurf angelegte Diskussion zwischen den grundlegenden Normen (insbesondere der IEC 62443 und der ISO/IEC 15408) zu dieser Diskussion wichtige Impulse.

Die für den Autor zweite wichtige Frage ist die der Aufrechterhaltung der Zertifizierung. In Zeiten von agilen Entwicklungsmethoden, „continuous delivery“ und weltweit vernetzten Geräten und Angreifern ist ein sicheres Produkt (für ein sicheres System) nur möglich, wenn erkannte Schwachstellen schnell und umfassend behoben werden. Hierbei darf der Endanwender nicht vor der Frage stehen, ob er eine zertifizierte oder eine sichere (aktuellere) Version einsetzen will. Dieses Problem beantworten die Schemata unterschiedlich. 5G und IACS erlauben für die Vertrauenswürdigkeitsstufe „basic“, unter gewissen Randbedingungen die Ausdehnung der Zertifikatsaussage auf neue Versionen, indem der Entwicklungsprozess (einschließlich des Patch-Management-Prozesses) stärker in den Fokus genommen wird. Das EUCC-Schema experimentiert mit verschiedenen „neuartigen“ Patch-Management-Ansätzen, um eine schnelle Bereitstellung zertifizierter Versionen zu ermöglichen. Hier stellt sich für den Autor die Frage, ob es nicht Zeit wäre, das Thema „dauerhafte Vertrauenswürdigkeit“ (continuous assurance) grundsätzlich zu behandeln, wozu auch klare rechtliche Fragestellungen gehören. Wie kann aus der Evaluierung der Version X auf eine oder mehrere zukünftige Versionen X+1, X+2 usw. geschlossen werden, und wer haftet, wenn dieser Schluss nicht trägt? Diese Fragestellung stand vor rund 10 Jahren im Umfeld der CC schon einmal auf der Agenda, wurde aber leider damals nicht weiterverfolgt.

5. Abkürzungsliste

Abkürzung	Langform	Abkürzung	Langform
3GPP	3rd Generation Partnership Project	ANSSI	Agence nationale de la sécurité des systèmes d'information
BSI	Bundesamt für Sicherheit in der Informationstechnik	CC	Common Criteria
CSA	Cybersecurity Act	CSPN	Certification de Sécurité de Premier Niveau
EA	European Accreditation	EAL	Evaluation Assurance Level
ECCG	European Cybercertification Coordination Group	ENISA	Agentur der Europäischen Union für Cybersicherheit
EUCC	Europäisches CC(-Schema)	ETSI	Europäische Institut für Telekommunikationsnormen
GSMA	GSM Association	IACS	Industrial Automation and Control Systems
IAF	International Accreditation Forum	IEC	International Electrotechnical Commission
IIoT	Industrial Internet of Things	IoT	Internet of Things
ISO	International Organization for Standardization	JRC	Joint Research Centre
NESAS	Network Equipment Security Assurance Scheme	SCAS	Security Assurance Specification
SCCG	Stakeholder Cybersecurity Certification Group	SFR	Security Functional Requirements
URWP	Union Rolling Work Programme		

Literaturhinweise

[1] Verordnung (EU) 2019/881 des Europäischen Parlaments und des Rates vom 17. April 2019 über die ENISA (Agentur der Europäischen Union für Cybersicherheit) und über die Zertifizierung der Cybersicherheit von Informations- und Kommunikationstechnik und zur Aufhebung der Verordnung (EU) Nr. 526/2013 (Rechtsakt zur Cybersicherheit) (Text von Bedeutung für den EWR)

[2] DIN EN ISO/IEC 27001, „Informationstechnik - Sicherheitsverfahren - Informationssicherheitsmanagementsysteme - Anforderungen (ISO/IEC 27001:2013 einschließlich Cor 1:2014 und Cor 2:2015)

[3] Common Criteria, DIN EN ISO/IEC 15408, „Informationstechnik - IT-Sicherheitsverfahren - Evaluationskriterien für IT-Sicherheit"

[4] DIN EN/IEC 62443, „IT-Sicherheit für industrielle Automatisierungssysteme"

[5] JRC/ERNCIP, „Recommendations for the Implementation of the Industrial Automation & Control Systems Components Cybersecurity Certification Scheme (ICCS)", https://erncip-project.jrc.ec.europa.eu/documents/recommendations-implementation-industrial-automation-control-systems-components

[6] Entwurf für ein CSA-Schema für die CC, die Fassung der ad-hoc-Gruppe kann unter https://www.enisa.europa.eu/publications/cybersecurity-certification-eucc-candidate-scheme gefunden werden.

[7] Sebastian Fritsch/Dietmar Bremser: EU Cybersecurity-Act, <kes> 2020#2

[8] Dietmar Bremser/Sebastian Fritsch: Europäische Cybersicherheitszertifizierung – der große Sprung nach vorn?, <kes> 2020#3

[9] Cloud-Testierungskatalog C5 des BSI, https://www.enisa.europa.eu/publications/eucs-cloud-service-scheme/

[10] SecNumCloud-Schema der ANSSI: https://www.ssi.gouv.fr/administration/qualifications/prestataires-de-services-de-confiance-qualifies/prestataires-de-service-dinformatique-en-nuage-secnumcloud/

[11] DIN EN ISO/IEC 17021-1,

[12] Entwurf für ein Cloud-Schema gem. CSA, https://www.enisa.europa.eu/publications/eucs-cloud-service-scheme/

[13] International Standard on Assurance Engagements (ISAE) 3402 Assurance reports on controls at a service organization

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Entwicklung der Informationssicherheit im Bereich der automobilen Lieferkette auf dem Weg zum TISAX-Standard in der Version 5.0

Dipl.-Ing. Thomas Salvador[1], Martina Petersen[2]

Kurzfassung:
Als Leitindustrie steht die Automobilindustrie seit Jahrzehnten unter besonderem Wettbewerbs- sowie Innovations-, Kosten- und Effizienzdruck. Aufgrund dieser Rahmenbedingungen wirkte die internationale Automobilindustrie nachweislich seit über 34 Jahren als Initiator für die ersten Zertifizierungen im Bereich Managementsysteme hin. Auch auf dem Gebiet der Informationssicherheit war die Automobilindustrie entsprechend eine der führenden zivilen Industrien, die die Harmonisierung, Einführung und Durchsetzung von Standards für Managementsysteme der Informationssicherheit (kurz ISMS) zum Ziel hatte. Über die ersten Anfänge im Jahr 2005 als Anforderungskatalog der Arbeitsgruppe für Informationssicherheit des Verbandes der Automobilindustrie (kurz VDA-ISA), entstand über verschiedene Zwischenstufen bis zum Jahr 2017 dann der erste international zertifizierbare Automobilstandard für Informationssicherheit, der TISAX-Standard. TISAX steht für Trusted Information Security Assessment eXchange und ist eine eingetragene Marke der ENX Association. Der letzte Stand des TISAX-Anforderungskatalogs in der Version 5.0 steht im Fokus dieses Fachbeitrages.

Stichworte: Automobilindustrie, ENX Association, Grundschutz-Tools, GSTools, Informationssicherheit, Informationssicherheit auf Reisen, ISMS-Tools, Lieferkette, Management von Informationssicherheit, OEMTISAX-Standard, TISAX, Verschlüsselung, Zertifizierung, Zulieferer.

1. Einführung und historische Entstehung der Informationssicherheit in der Automobilindustrie

Als Leitindustrie steht die Automobilindustrie seit Jahrzehnten unter besonderem Wettbewerbs- sowie zugleich Innovations-, Kosten- und Effizienzdruck. Aufgrund dieser Rahmenbedingungen wirkte die internationale Automobilindustrie nachweislich seit über 34 Jahren als Initiator für die ersten Zertifizierungen im Bereich Managementsysteme (Erstveröffentlichung der ISO 9001[3] im Jahr 1987) hin. Auch auf dem Gebiet der Informationssicherheit war die Automobilindustrie entsprechend eine der führenden zivilen Industrien, die die Harmonisierung, Einführung und Durchsetzung von Standards für Managementsysteme der Informationssicherheit (kurz ISMS) zum Ziel hatte. Aufgrund der besonderen Anforderungen einer automobilen Lieferkette mit einer üblichen Fertigungstiefe[4] auf Seiten der Zulieferer von 75%, Tendenz steigend, spielen Themen wie Geheimhaltung, Informationssicherheit, IT-Sicherheit samt technischer organisatorischer Maßnahmen (TOMs) in der Zusammenarbeit mit der Lieferkette eine eminent wichtige Rolle. Gleichzeitig bedürfen sie jedoch auch weiterreichender Lösungen und Maßnahmen, wobei diese im Bereich ziviler Infrastrukturen bereits vor dem Jahr 2000

[1] Geschäftsführer, QM Experts GmbH, München

[2] Geschäftsführende Gesellschafterin, CertVision GmbH, Nürnberg

[3] https://www.beuth.de/de/norm/din-iso-9001/3053678, aufgerufen am 19.01.2021

[4] Anteil der Wertschöpfung am Endprodukt https://www.vda.de/de/themen/automobilindustrie-und-maerkte/mittelstand/zulieferindustrie-und-mittelstand.html, aufgerufen am 19.01.2021

jenseits international gültiger Standards lagen. Standards, die zu diesem Zeitpunkt die Spezifika der Automobilindustrie nicht abdeckten.

Die Schlüsse aus den Erfahrungen mit internationalen Normen und Standards ab 1987 und davor, sowie der steigende Wettbewerbsdruck mit zunehmend bedrohtem technischem Informationsaustausch führte deshalb im Jahr 2000 zur Gründung der ENX Association, einem Zusammenschluss europäischer und nordamerikanischer Automobilhersteller, -zulieferer und -verbände. Primärziel war es zunächst, ein sicheres Industrienetzwerk zu schaffen. Dies führte im weiteren Verlauf allerdings auch zu einem weiterentwickelten Zertifizierungsmodell von Informationssicherheitsmanagementsystemen in Form eines vertrauenswürdigen Austauschs von Informationssicherheitsbewertungen, sogenannten TISAX-Assessments. Diese wiederum sollten als Eintrittskarte in eine besonders abgesicherte Infrastruktur mit reglementiertem Zugang sowohl für ausgewählte Zertifizierungsgesellschaften als auch für IT-Service Provider dienen.

Über die ersten Anfänge im Jahr 2005 als Anforderungskatalog der Arbeitsgruppe für Informationssicherheit des Verbandes der Automobilindustrie (kurz VDA-ISA), entstand über verschiedene Zwischenstufen bis zum Jahr 2017 dann der erste international zertifizierbare Automobilstandard für Informationssicherheit, der TISAX-Standard. TISAX® steht für Trusted Information Security Assessment eXchange und ist eine eingetragene Marke der ENX Association.

Aufgrund der schnellen Entwicklungen auf dem Gebiet Informationssicherheit und nicht zuletzt auch durch die globale Covid-19-Pandemie verstärkt, wurde seitens des VDA bereits 15 Monate nach Veröffentlichung der Vorgängerversion nun die Version 5.0 veröffentlicht. Somit kamen die neuesten Erkenntnisse aus der Zertifizierungs- und weltweiten Implementierungspraxis der Praxis außerordentlich schnell zugute. Dieser letzte Stand des TISAX-Anforderungskatalogs in der Version 5.0 soll im Fokus dieses Fachbeitrages stehen.

2. Hintergründe zum Aufbau der automobilen Lieferkette und der Fertigungstiefe von automotiven OEMs

Ein Originalausrüstungshersteller oder Erstausrüster (englisch Original Equipment Manufacturer, OEM) ist ein Hersteller von Komponenten oder Produkten, der diese nicht selbst in den Einzelhandel bringt.[5] Der Begriff ist nicht der Automobilindustrie zu eigen und kommt auch in vielen anderen Branchen zum Einsatz. Nachfolgende Abbildung zeigt den Aufbau der Automobilen Lieferkette am Beispiel einer Instrumententafel über einen direkten Lieferanten („1st Tier“, aus dem Englischen für „Rang“, bzw. „Lieferant ersten Ranges“) und entsprechender Unterlieferanten über die weitverzweigte Lieferkette. So setzt sich die Instrumententafel wie sie in ein Gesamtfahrzeug verbaut wird, wiederum aus verschiedenen Einzelkomponenten zusammen, wie Kabel oder Displays. Diese wiederum aus Einzelelementen wie beispielsweise im Fall des Kabels aus dem leitfähigen Element Kupfer, was üblicherweise zur Bildung eines Kabels mit einem Isolator aus Polyethylen/Polypropylen versehen wird und an Anfang und Ende einen

[5] BGH, Urteil vom 6. Juli 2000, I ZR 244/97

Stecker zur Herstellung einer sicheren Verbindung aufweist. Das Kupfer wiederum entspringt einer Metallerzmine usw. Hinter jedem Rohstoff, Element, Bauteil oder Komponente steht ein Lieferant, mit dem während des gesamten Produktentstehungsprozesses Informationen und Daten ausgetauscht werden. Es entsteht ein weit verzweigtes Netz an Lieferanten. Im Fall von Volkswagen sprechen wir über eine Baumstruktur mit mehr als 40.000 Lieferanten weltweit.

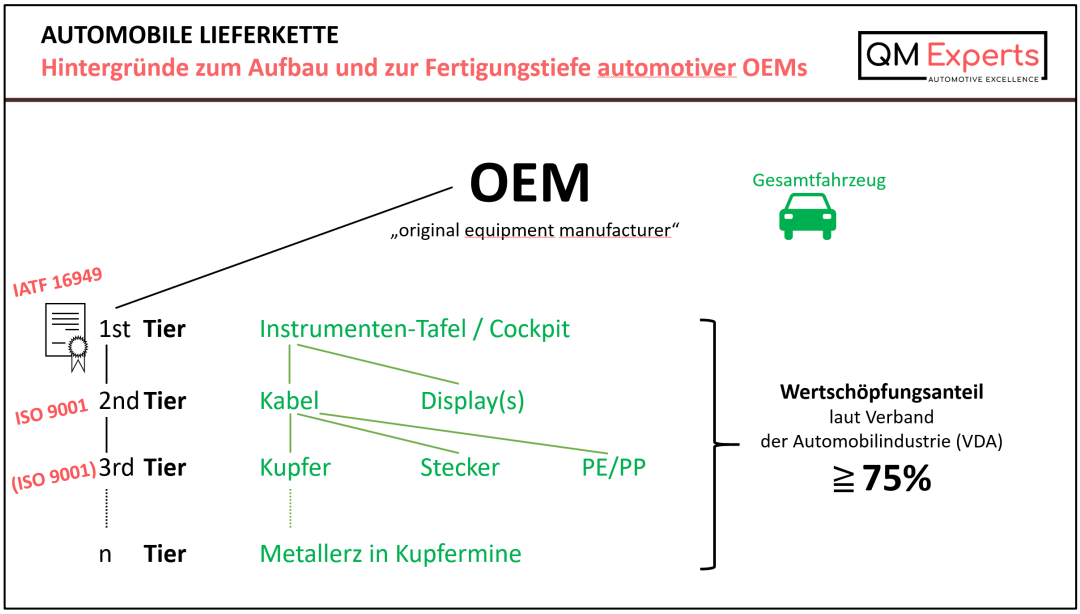


Abbildung 1: Automobile Lieferkette am Beispiek einer Instrumententafel

2.1. Besondere Anforderungen an Informationssicherheit in der Automobilindustrie

Im Rahmen von Entwicklungsprojekten werden viele vertrauliche und geheime Informationen (Entwicklungsdaten, Produktionsprozessdaten, Lieferanteninformationen, etc.) zwischen den OEMs und der Lieferkette ausgetauscht. Diese Daten benötigen zu ihrem Schutz ein entsprechendes Informationssicherheitsmanagementsystem (ISMS), welches Themen wie Datenschutz, Datensicherheit und IT-Sicherheit gleichermaßen berücksichtigt. Während OEMs im Automobilbereich bereits heute sehr viel in die Sicherheit ihrer IT-Systeme, Services, Infrastruktur aber auch in das Bewusstsein und die Kompetenz von Mitarbeitern investieren, resultiert auf Ebene des komplexen und weitverzweigten Lieferantennetzwerks jedoch eine große Vulnerabilität, da diese anteilig den größten Teil der Wertschöpfung ausmachen und heute häufig noch nicht über so umfangreiche Schutzmaßnahmen zur Abwehr von Bedrohungen hinsichtlich der Informationssicherheit verfügen. Deshalb ist der Verband der deutschen Automobilindustrie (VDA) bereits im Jahr 2005 dazu übergegangen, Anforderungen an die Informationssicherheit für Lieferanten herauszugeben, die dann im Jahr 2017 im zertifizierbaren TISAX-Standard mündeten. Für viele OEMs ist der Standard seitdem verbindlich für die Aufnahme einer Lieferantenbeziehung.

Diese Hintergründe sollen einen ersten Einblick geben, warum die Automobilindustrie mit der Arbeitsgruppe Informationssicherheit hier aufbauend auf der ISO/IEC 27001

einen auf die Automobilindustrie adaptierten und erweiterten Standard ins Leben gerufen hat, über dessen Struktur nachfolgend ein kurzer Überblick gegeben werden soll.

3. Besonderheiten des TISAX-Standards (in der Version 5.0) Stand August 2020

Seit Anfang August 2020 steht der neue VDA-ISA Katalog in der Version 5.0.1 für TISAX Assessments zur Verfügung. Für bereits laufende Assessments kann noch bis zum 31. März 2021 der alte Katalog verwendet werden. Bestehende Labels, welche ab Abschluss des letzten Assessments eine dreijährige Gültigkeit haben, müssen ab sofort im Rahmen eines Re-Assessments auch nach Version 5.0 überprüft werden.

Der Verband der Automobilindustrie spricht von einer „grundlegenden Überarbeitung" mit „strukturellen als auch inhaltlichen Optimierungen". Es zeigt sich, dass sich viele Detailanpassungen hinter den in der offiziellen Versionshistorie aufgeführten Änderungen verbergen. Was insbesondere bei einem Re-Assessment auch ein Nachziehen von Prozessen und Dokumenten notwendig machen wird.

Der Wegfall des bisher eigenständigen Moduls „Anbindung Dritter" und dessen Überführung in die Themengebiete des Moduls „Informationssicherheit" ist im Zusammenhang weitestgehend selbsterklärend. Ähnliches gilt für die drei neuen Controls „mobiles Arbeiten", „Umgang mit Identifikationsmitteln" und „Eignung von Mitarbeitern".

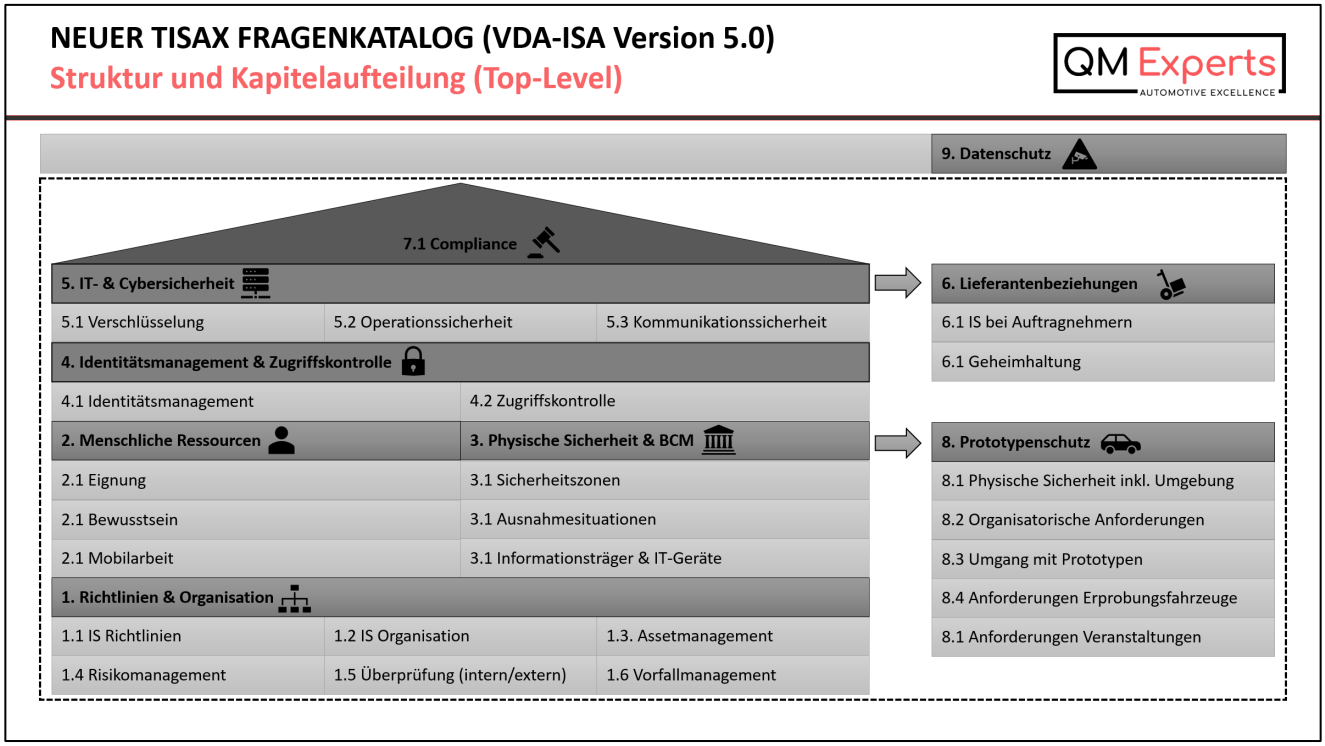


Abbildung 2: Übersicht des TISAX-Standards in der Version 5.0 (Veröffentlichungsdatum August 2020).

Die großen Änderungen verbergen sich im Hauptmodul „Informationssicherheit". Um diese greifen zu können, reicht der Blick in die Änderungshistorie jedoch leider nicht aus. Nicht nur Control Nummerierungen haben sich an vielen Stellen verschoben – was

sich gerade bei der bisherigen Kennzahlenmetrik bemerkbar macht - auch die Kontrollfragen und Anforderungen selbst wurden umformuliert und neu gruppiert. Neue Anforderungen sind dazugekommen und existierende Anforderungen können ihre Einstufung zwischen „Soll“ oder „Muss“-Anforderungen gewechselt haben. Die bisherige Einstufung „Kann“ wurde gestrichen und existiert nicht mehr. Das neue und eigenständige Control „mobiles Arbeiten“ geht nun stärker auf die aktuellen Anforderungen im Home-Office und Maßnahmen beim Reisen in sicherheitskritische Länder ein.

3.1. Exkurs: Umsetzungshinweise zum Informationsschutz auf Reisen

Aus Sicht eines gelebten Informationssicherheitsmanagementsystems gibt es insbesondere beim Reisen und mobilen Arbeiten eine Vielzahl an Dingen zu beachten. Die nachfolgende Aufzählung beleuchtet dazu exemplarisch einige wenige, dennoch wichtige Aspekte:

Insbesondere gibt es Länder, die eine Entschlüsselung bei der Einreisekontrolle verlangen können (z.B. USA) sowie Länder in denen die Einfuhr von Verschlüsselungssoftware – auch auf Datenträgern – generell verboten ist (z.B. China, Russland). Bitte erkundigen Sie sich daher vor der Reise diesbezüglich über die individuellen Einreisebestimmungen des jeweiligen Landes, um Gesetzesübertretungen zu vermeiden.

Wenn Sie Ihren Koffer packen, achten sie darauf, dass vertrauliche oder sensible Informationen nicht ins Reisegepäck gehören. Es ist zu empfehlen, so wenig technische Geräte wie möglich mitzunehmen, um die Angriffsfläche grundsätzlich zu reduzieren. Nutzen Sie einen Reise-Laptop und speichern Sie darauf ausschließlich die Daten, die Sie für die Reise benötigen.

Beachten Sie: Ein Hotelzimmer ist kein privater Raum; Lassen Sie Unterlagen und elektronische Geräte nicht unbeaufsichtigt im Hotelzimmer. Zimmer- und Hotelsafes an der Rezeption bieten keinen echten Schutz, da Angestellte mit einem Generalschlüssel oder Code Zugang zu Wertsachen erhalten können. In manchen Ländern werden versteckte Aufzeichnungen von allen Aktivtäten in Hotelzimmern angefertigt. Daher sollten Sie vertrauliche und sensible Gespräche nicht im Hotelzimmer führen.

Ausstattung, Verschlüsselung und Datenübertragung: Monitore sollten stets mit einer Sichtschutzfolie ausgestattet werden. Alle Programm-Updates für Geräte sollten Sie vor Reiseantritt aufspielen. Bluetooth und WLAN bieten Einfallswege für Angreifer, daher gilt bei Nichtgebrauch diese abzuschalten. Öffentliche WLAN-Netze sollten Sie nach Möglichkeit nicht nutzen. Wann immer möglich, Internet und E-Mail über eine eigene VPN-Verbindung nutzen. USB-Sticks oder andere externe Speichermedien, die Ihnen Dritte angeboten haben, sollten Sie weder auf Reisen noch nach der Rückkehr nutzen. Achten Sie darauf, dass elektronische Endgeräte im Ausland keine kritischen (anzüglichen, religiösen, politischen oder gar pornografischen) Inhalte in Form von Texten oder Bildern enthalten. Dies kann zur Konfiszierung führen und Ihnen erhebliche Probleme bereiten.

4. Das TISAX Label als modernes Zertifikat

Streng genommen dürfte man das TISAX-Verfahren gar nicht Zertifizierung nennen. Eigentlich handelt es sich bei TISAX um ein Assessment, also eine Art Gutachten. Während „Zertifizierung“ den Prozess zur Erlangung eines Managementsystemzertifikats beschreibt, bezeichnet „Assessment“ hier das Ergebnis einer Informationssicherheitsüberprüfung eines Unternehmens. Diese mündet in einen „Assessment Report“ und schließt mit einem sogenannten Label ab. Wenn man so möchte, ist dieses Label das Äquivalent zu einem Zertifikat.

Die weltweit bekannteste und erste Managementsystem-Norm ist die aus dem Jahr 1987 stammende ISO 9001. Sie bescheinigt Unternehmen mit einem Zertifikat, dass diese die internationalen Anforderungen an Qualitätsmanagementsysteme erfüllen. Unternehmen werben gerne damit. Sie transportieren so nach außen, dass Sie ein gut geführtes Unternehmen mit entsprechenden Nachweisen sind.

International ist diese Praxis auf Ebene der ISO 9001 jedoch leider nicht sehr transparent. Es gibt bis heute keine zentrale und überwachte Datenbank, in der sich ein Kunde über die Rechtmäßigkeit eines ISO-9001-Zertifikats erkundigen kann. Aus diesem Grund sind einige Zertifizierungsgesellschaften wie TÜV, DEKRA, etc. mit Kundenportalen dazu übergegangen, zentrale Abfragedatenbanken für die Öffentlichkeit anzubieten.

Diesen grundsätzlichen Nachteil hat die „ENX Association“ als Governance und Träger-Organisation, die hinter dem TISAX-Standard steht, von Anfang an vermieden. So wurde mit den ENX YELLOW PAGES eine Datenbank und ein zentrales sowie öffentliches Register geschaffen, in dem alle gültigen nach TISAX bewerteten Unternehmen mit Standort und Registrierungsnummer abgefragt werden können. Der Vorteil einer derartigen zentralen Lösung ist ein tagesaktueller Überblick über die Gültigkeit des Informations-Sicherheitsstatus eines Lieferanten oder eines Unternehmens.

Um den modernen Ansatz zu unterstreichen, hat die „ENX Association“ bewusst auf ein neues Wording gesetzt. Was in der ISO-Welt das Zertifikat ist, wird in der TISAX-Welt als Label bezeichnet. Im Rahmen der Begutachtung eines Managementsystems für Informationssicherheit sind für die Bereiche Informationssicherheit, Prototypenschutz und Datenschutz in Summe die Begutachtung von bis zu 8 Anforderungsniveaus in Form von zugeordneten Labels möglich.

5. Von der Anforderungsseite zur Umsetzung in der Praxis am Beispiel CertVision GmbH

5.1. Marktüberblick ISMS-Tools vom zivilen bis militärischen Einsatz

Historisch betrachtet war 1998 das GSTool, welches vom Bundesamt für Sicherheit in der Informationstechnik entwickelt und herausgegeben wurde, die erste toolgesteuerte Möglichkeit, Sicherheitskonzepte nach der Vorgehensweise des IT-Grundschutzes um-

zusetzen. Ende 2014 wurde wegen mangelnder Wirtschaftlichkeit der Vertrieb und damit einhergehend Ende 2016 auch der Support eingestellt[6].

Daraufhin stellte das BSI Alternativen zum GSTool vor, die 2015 von der CSC in einer Studie untersucht wurden[7]. Aus dem damals noch überschaubaren Kreis an ISMS-Tools sind mittlerweile über 34 gelistete Alternativen zum GSTool geworden[8] und zudem noch weit mehr ISMS-Tools in der DACH-Region verfügbar.

Wenn man sich dieses rasche Wachstum anschaut, wäre vermutlich eine aktuelle Betrachtung der verfügbaren Tools und gewachsenen Bedürfnisse sinnvoll. Das stetig wachsende Software-Angebot zeigt aber auch, welche Bedeutung der Einsatz von ISMS-Tools eingenommen hat und dass die Nachfrage nach solchen enorm gestiegen ist.

Längst liegt der Schwerpunkt der Tools und des Anwendungsspektrums nicht mehr auf dem Gefährdungsmanagement und dem damit verbundenen Maßnahmentracking. Vielmehr wird heutzutage ein gesamtheitlicher Ansatz in Betrachtung des kontinuierlichen Verbesserungsprozesses (KVP) und einem gelebten Informationssicherheitskonzept verfolgt.

Um zu veranschaulichen, welche Teilbereiche ein ISMS-Tool heutzutage erfüllen sollte, schauen wir uns die Inhalte des NormTrackers an, eines in der Microsoft Cloud auf Azure gehosteten Dienstes der CertVision GmbH.

Inhalte / Funktionen:

- Strukturiertes Abarbeiten eines Anforderungskataloges mithilfe von Checklisten
- Aufgabentracking mit Rollenkonzept
- Gefährdungsanalyse und -behandlung mit Maßnahmenverfolgung
- Wirksamkeitsprüfungen und protokollierte End- und Freigabekontrollen
- Testprotokollierung und Änderungshistorie
- Dokumentationsupload und Verlinkungsmöglichkeiten von externen Speicherorten
- Aufwärtskompatibilität und flexible Skalierbarkeit

Anhand dieser Inhalte kann ein ganzheitliches ISMS aufgebaut und optimal betrieben werden. Andere Software-Hersteller bieten ähnlich umfangreiche Tools und ISMS-Konzepte an.

5.2. Exemplarisches Praxisbeispiel an einem cloudbasierten ISMS-Tool

Compliance-Vorgaben dokumentieren und deren Einhaltung prüfen, definierte Sicherheitslevel mittels getroffener technischer und organisatorischer Maßnahmen erreichen,

[6] Wikipedia „GSTOOL“ siehe unter https://de.wikipedia.org/wiki/GSTOOL

[7] CSC-Studie: GSTOOL Quo Vadis? „Alternativen zum GSTool“ siehe unter https://www.kronsoft.de/download/free/csc-studie.pdf, aufgerufen am 08.01.2021

[8] Alternative IT-Grundschutz-Tools; siehe unter https://www.bsi.bund.de/DE/Themen/ITGrundschutz/GSTOOL/Andere-Tools/anderetools_node.html, aufgerufen am 08.01.2021

dabei unterstützen ISMS-Tools. Wir möchten Ihnen anhand unseres NormTrackers den Prozess der Implementierung bis hin zur Umsetzung etwas veranschaulichen.

Kunde	Software-Hersteller (Bsp. CertVision)	Cloud-Provider (Bsp. Mircosoft Azure)
1. Beauftragung ISMS-Tool / Cloud-Dienste	Bereitstellung Kundeninstanz (Datenbankskalierung, Mandantentrennung, Storageverschlüsselung)	Wartung, Verfügbarkeit, Betrieb der PaaS Infrastruktur im europäischen Rechenzentrum
2. Login / Authentifizierung über Identitätsdienst	Implementierung der Azure AD PaaS Lösung als zentraler Identitätsmanagementdienst in die Anwendung	Entwicklung, Bereitstellung und Wartung des globalen Identitätsdienstes Azure Active Directory (Anmeldung mit 2-Faktor-Authentifizierung, Single Sign On, Conditional Access)
3. Umsetzung ISMS	Wissen, Vorlagen, Assistenten im Tool, Integration in andere Clouddienste	Backup, Datensicherheit, Bereitstellung Container und fertiger App-Services als Webserver
4. Validierung / Zertifizierung	Regelmäßige Updates / Erweiterungen	Sicherstellung der Governance und Sicherheit der bereitgestellten IaaS und PaaS Dienste sowie Weiterentwicklung

Abbildung 3: Exemplarisches Praxisbeispiel cloudbasiertes ISMS Tool CertVision

5.3. Implementierungsvorteile von cloudbasierten ISMS-Tools

Die Nachfrage nach cloudbasierten Lösungen nimmt stetig zu. Anbieter werben mit mehr Flexibilität, unkomplizierter Einführung und reduziertem Wartungs- und Administrationsaufwand gegenüber On-Premise Lösungen. Die Angst vor unkontrolliertem Datenabfluss sinkt, nicht zuletzt durch die hohen Anforderungen, denen Cloud-Provider unterliegen. Doch cloudbasierte ISMS-Tools bieten noch viel mehr Vorteile, die wir Ihnen in der Darstellung auf der folgenden Seite erläutern möchten.

Zusammenfassend darf behauptet werden, dass cloudbasierte ISMS-Lösungen durchaus Potenzial für einen geregelten und sicheren Security- und Geschäftsprozess bieten und dabei optimale Einsparungsmöglichkeiten auf technischer und organisatorischer Ebene zulassen.

Automatische Updates, Upgrades und Weiterentwicklungen	Hersteller sorgt für regelmäßige Erweiterungen und Updates der Software sowie für die Behebung von Bugs. Dies geschieht zumeist unauffällig außerhalb der Geschäftszeiten, sodass der Anwender durch die Einspielungen keinen Ausfall registriert. Somit haben Anwender immer die aktuelle Softwarevision im Einsatz.
Wartungen	Für die Wartung der gehosteten Lösung sorgt der Provider, auch diese Aufgabe entfällt beim Anwender.
Datenschutz, Datensicherheit	Die Verantwortung über die Einhaltung von Datenschutzanforderungen wird vom Hersteller / Provider erfüllt.
Orts- und zeitunabhängiges Arbeiten	Mehrere Standorte können problemlos und zeitgleich in einer Anwendung arbeiten, das verbessert die Zusammenarbeit.
Skalierbarkeit	Umfang und User sind innerhalb kurzer Zeiträume dynamisch nach oben und unten skalierbar.
Rechenleistung, Kapazität	Eigene Rechner werden entlastet. Die Cloud ist ohne große Mehrkosten beliebig skalierbar, speziell für Testing und BigData-Szenarien. Dies erhöht die Produktivität und senkt die Time to Market.
Bindung	Umstieg auf andere Modelle und Tools sind schneller möglich.
Backup	Der benötigte Speicherplatz ist ausgelagert, es wird keine eigene Hardware benötigt. Hierdurch entfällt die Betreuung durch Administratoren. Backupmöglichkeiten gibt es als SaaS-Angebote.
Implementierungs- und Administrierungsaufwand	Bei der Einführung der cloudbasierten Anwendung sind deutlich reduzierte Implementierungsaufwände erforderlich, da die gesamte Basis vom Cloud-Provider standardisiert zur Verfügung gestellt wird. Dadurch gelingt eine schnelle und unkomplizierte Implementierung.
Kostenminimierung	Kosten nach Nutzen. Finanzierung ist besonders für kleine Organisationen attraktiv. Kosten für eine eigene Infrastruktur werden reduziert. Zudem lassen sich Capex- in Opex-Kosten verwandeln.
Organisatorische Flexibilität	Unternehmen können Sicherheits-, Auslastungs-, Know-how und Technologierisiken an den Dienstleister auslagern, nach Bedarf in verschiedene Technologien und Skalierungsstufen investieren.
Fokus	Erhöhte Konzentration auf das Kerngeschäft.

Abbildung 4: Vorteile von cloudbasierten ISMS-Tools am Beispiel CertVision

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Self-Sovereign Identity – Vertrauensbasis für selbstbestimmte Identitätsnetzwerke

Paul Bastian[1], Micha Kraus[1], Jörg Fischer[1], Christoph Bösch[1]

Kurzfassung:
Der digitale Nachweis einer Identität ist seit den Anfängen des Internets eine Herausforderung. Bislang konnte sich kein weltweiter Standard durchsetzen. Fehlende digitale Nachweise bleiben somit eines der größten Digitalisierungshemnisse[2]. Derweil werden Rufe nach dezentralen und nutzerzentrierten Lösungen immer lauter. Dies zeigten jüngst die kontroversen Diskussionen über den Datenschutz bei der Umsetzung der Corona-Warn-App. Bereits 2016 formulierte Christopher Allen die Kernthesen der Self-Sovereign Identity (SSI) und fordert eine verteilte, selbstbestimmte, die Privatsphäre schützende und sichere Identität in der Hand des Nutzers. Die Entwicklung dieser Technologie schreitet seitdem rasant voran, begleitet von offenen Standards und vielversprechenden Open-Source-Projekten. Getrieben wird sie von einer enthusiastischen, breiten Community aus Industrie, Finanzwelt und öffentlicher Verwaltung. In diesem Beitrag betrachten wir die verschiedenen Facetten dieser jungen Technologie und die Vertrauensbeziehungen innerhalb des SSI-Ökosystems.

Stichworte: Aries, Blockchain, Datenschutz, Dezentrale Identitäten, Digitale Identität, Governance, Hyperledger Indy, IDunion, Lissi (Let's initiate SSI), OPTIMOS, Self-Sovereign Identity, SSI (Self-Sovereign Identity), Ursa, Verifiable Credentials, Vertrauen

1. Einführung

In einer zunehmend vernetzten Welt kommen der sicheren digitalen Identität und digitalen Nachweisen eine tragende Bedeutung zu. Der Grundstein ist mit der eID-Funktion des Personalausweises gelegt. Notwendig ist ein ganzheitliches Ökosystem digitaler Identitäten, das einen nahtlosen, sicheren und datensparsamen Austausch von unterschiedlichen digitalen Nachweisen ermöglicht.

Die Verwendung digitaler Identitäten hat einerseits das Potenzial Geschäftsprozesse, Arbeitsabläufe und die technischen Systeme effektiver zu gestalten. Andererseits stellen die persönlichen Daten ein sensibles, schützenswertes Gut dar. Um eine hohe Nutzerakzeptanz zu erreichen erfordern sie insbesondere die Beachtung der Privatsphäre. Während sich einige Protokolle wie TCP/IP oder HTTP langfristig etabliert haben, hat sich seit den Anfängen des Internets kein weltweiter Standard für die Identifizierung und Authentifizierung durchgesetzt.

Mit OpenID Connect wurde ein erfolgreicher Standard geschaffen, der einen interoperablen Austausch überprüfbarer Identitätsdaten ermöglicht. Dabei ist aber die zentrale Rolle des Identity-Providers zu beachten, der bei jedem Authentifizierungsvorgang involviert ist und somit allwissend über die Aktivitäten der Nutzer Bescheid weiß. Die Diskussionen über den Datenschutz rund um die Einführung der Corona-Warn-App zeigen, dass dezentrale und nutzerzentrierte Lösungen von der Gesellschaft gefordert werden.

1 Bundesdruckerei GmbH, Kommandantenstraße 18, 10969 Berlin
2 https://www.bundesregierung.de/breg-de/aktuelles/digitale-identitaet-1824658

10 Principles of Solf-Sovreign Identity

1. **Existence**. Users must have an independent existence.
2. **Control**. Users must control their identities.
3. **Access**. Users must have access to their own data.
4. **Transparency**. Systems and algorithms must be transparent.
5. **Persistence**. Identities must be long-lived.
6. **Portability**. Information and services about identity must be transportable.
7. **Interoperability**. Identities should be as widely usable as possible.
8. **Consent**. Users must agree to the use of their identity.
9. **Minimalization**. Disclosure of claims must be minimized.
10. **Protection**. The rights of users must be protected.

Schon am Anfang der Jahrtausendwende wurden mit idemix[1] und U-Prove[2] Techniken vorgestellt, die dem Nutzer mehr Kontrolle und Anonymität in digitalen Identifizierungsvorgängen geben. Diese sogenannten *attribute-based credentials* wurden in Forschungsprojekten wie ABC4Trust[3] untersucht und prototypisch umgesetzt. Eine weitläufige Umsetzung dieser Systeme scheiterte bisher auch mangels eines öffentlich zugänglichen und dezentralen Speichers für Prüfdaten, der gleichzeitig die Privatsphäre der Nutzer gewährleistet. Genau hier kann die aufstrebende Distributed Ledger Technologie (DLT) ihre Stärken ausspielen und das fehlende Puzzleteil für ein umsetzbares Gesamtsystem darstellen. Im Jahr 2016 schrieb Christopher Allen seine Prinzipien für selbstbestimmte Identitäten (Self-Sovereign Identity - SSI) und legte damit das Fundament für ein dezentrales, selbstbestimmtes Identitätsnetzwerk (siehe Kasten).

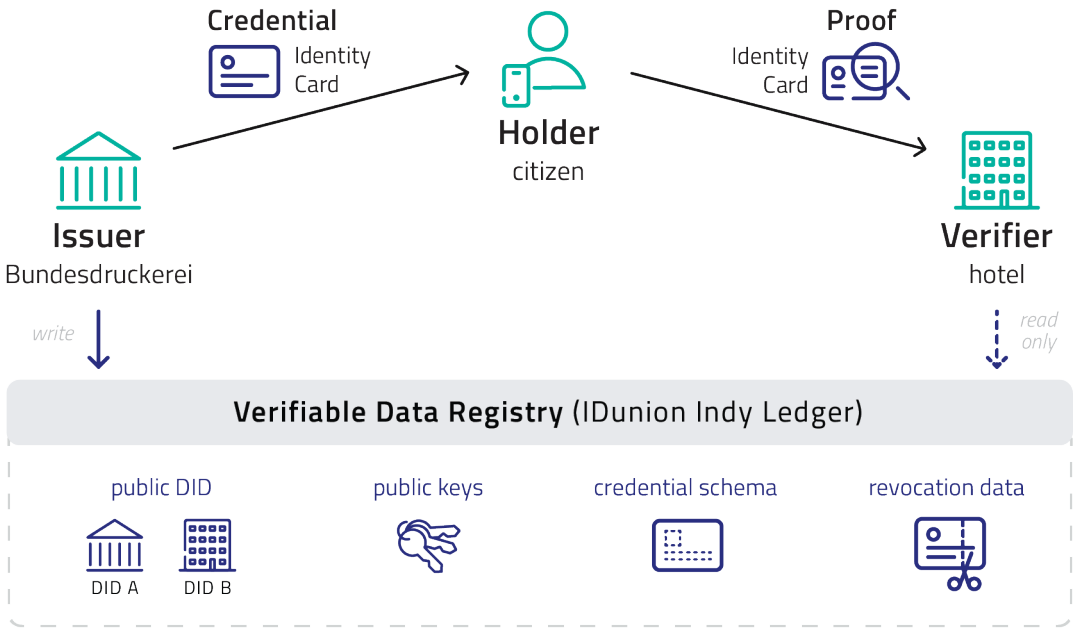
1.1. Das magische Dreieck

Wir stellen die Grundmechanismen von SSI anhand eines Beispiels vor, bei dem ein Dienstreisender in ein Hotel einchecken möchte. Nach dem Bundesmeldegesetz ist das Hotel verpflichtet einen Meldeschein mit den Identitätsdaten jedes Übernachtungsgastes zu erstellen. Für die Rechnung soll außerdem die Adresse des Arbeitgebers des Reisenden angegeben werden. Der Reisende benötigt also einen authentischen Nachweis seiner Identität und die Daten des Arbeitgebers, die er dem Hotel zeigen kann. Dieser Vorgang kann analog durchgeführt werden, indem der Reisende seinen Personalausweis und seine Visitenkarte aus seinem Portemonnaie entnimmt und an der Rezeption des Hotels vorlegt. Dadurch lassen sich zumindest die Echtheit des Personalausweises prüfen und die Daten anschließend in den Meldeschein übertragen.

Nun betrachten wir wie ein digitaler kontaktloser Check-In mit SSI aussieht. Dabei finden wir neben vielen Parallelen zu der analogen Ausprägung zusätzliche Vorteile. Im magischen SSI-Dreieck sind drei Rollen beschrieben: *issuer* (Aussteller), *holder* (Inhaber) und *verifier* (Prüfer). In unserem Beispiel nimmt der Reisende die Rolle des *holder* ein. Wie in seinem richtigen Portemonnaie, verwaltet er bei SSI verschiedene *credentials* (Nachweise) in seiner digitalen *wallet* (virtuelles Portemonnaie). Diese *credentials*

[3] https://www.abc4trust.eu/

wurden von verschiedenen *issuers* ausgestellt und beinhalten in der Regel *claims* (Aussagen, z.B. Name und Alter) über den *holder*. In unserem Beispiel gibt es zwei *issuer*. Zum einen die Bundesdruckerei, die ein *credential* mit persönlichen Meldedaten ausstellt und zum anderen der Arbeitgeber, dessen *credential* Daten der Betriebszugehörigkeit und der Rechnungsadresse beinhaltet.



Das magische Dreieck – Schematische Darstellung des Self-Sovereign Identity Konzepts

Der *holder* kann nun selbstbestimmt entscheiden welche *claims* er einem *verifier* zur Prüfung übermittelt. Der *verifier* kann auf Grundlage öffentlicher Prüfdaten die Authentizität und Unverfälschbarkeit der übermittelten *claims* sicherstellen. Die Prüfdaten werden dabei von einer dezentralen Infrastruktur (hier IDunion *ledger*) des SSI-Ökosystems bereitgestellt. Die Prüfdaten umfassen neben Informationen über die Identität des *issuer* und über das *credential schema* auch Angaben zu revozierten *credentials*. Personenbezogene Daten müssen und werden keinesfalls öffentlich zur Verfügung gestellt. Der Ausstellungsvorgang ist somit von dem Authentifizierungsvorgang entkoppelt, da für die Prüfung keine Kontaktaufnahme zum *issuer* erforderlich ist. Wie in der analogen Welt üblich, erfährt der Aussteller der Nachweise also nichts von deren Verwendung, was ein wichtiger Beitrag zur Selbstbestimmung der Nutzer darstellt.

Im Gegensatz zum analogen Prozess können bei SSI zusätzliche Privatsphäre fördernde Mechanismen zum Einsatz kommen. Die selektive Offenlegung von *claims* eines *credentials* oder logische Aussagen über *claims* (z.B. Altersverifikation "über 18") fördern die Datenminimierung. So würde in unserem Beispiel der Reisende nur die für den Check-In erforderlichen Daten aus den beiden *credentials* präsentieren und dabei nicht benötigte Daten verbergen, wie z.B. den Geburtsort oder die Dienstrufnummer. Die Offenlegung der *claims* wird bei SSI in der Regel mit einem interaktiven kryptografischen Beweis umgesetzt, sodass der *holder* häufig auch als prover bezeichnet wird. Bei jeder Verbindung zu einer Entität tritt der *holder* unter einem neuen Pseudonym auf, so dass er nur so viel wie nötig und so wenig wie möglich über sich bekannt geben muss und

somit eine unerwünschte Nachverfolgbarkeit erschwert wird. Höchstmögliche Kontrolle und Transparenz über die Daten schaffen Vertrauen beim Nutzer. Durch offene Standards gesicherte Portabilität und Interoperabilität ermöglichen den langfristigen Erhalt eines SSI-Ökosystems.

Self-Sovereign Identity bietet aber auch für Unternehmen und die öffentliche Verwaltung einen großen Mehrwert – es schafft einen einheitlichen Standard zum Austausch von Identitätsinformationen, bietet eine gemeinsame, verteilte und damit ausfallsichere Infrastruktur und gewährleistet die einfache und sichere Überprüfung von Daten ohne dabei selbst immer den Aussteller kontaktieren zu müssen. Damit ist es bestens für eine Beschleunigung der Digitalisierung geeignet, bei der über eine Schnittstelle mehrere Identitätsinformationen verschiedener Aussteller gleichzeitig überprüft werden können und sich somit für Unternehmen kostensparende und schlanke Prozesse ermöglichen lassen. Zudem können viele Entwicklungskomponenten repliziert werden, wenn weitere Arbeitsabläufe ebenfalls auf SSI migriert werden.

2. SSI-Ökosystem und Kernkomponenten

2.1. Standardisierung

Derzeit gibt es in mehreren Standardisierungsorganisationen Aktivitäten zu den unterschiedlichen Bausteinen eines dezentralen Identitätssystems. In der W3C werden die *decentralized identifiers*[4] (*DIDs*) und v*erifiable credentials*[5] (*VC*) spezifiziert. Diese Standards ermöglichen die Identifizierung des *DID subject*, einer Entität (Nutzer, Organisation, ...) ohne von zentralen Diensten abhängig zu sein. Mit v*erifiable credentials* können außerdem überprüfbare Identitätsattribute zwischen einem Aussteller und Besitzer bzw. Besitzer und einem Überprüfenden direkt ausgetauscht werden. In der Decentralized Identity Foundation[6] (DIF) findet u.a. die Standardisierung zu *DIDComm*[7] statt, eine sichere und auf *DIDs* aufbauende dezentrale Kommunikationsschicht. Aus dem Bereich der IETF werden viele kryptografische Verfahren, Kodierungen und andere Protokollbausteine wiederverwendet oder als Drafts neu verabschiedet. In den Standardisierungsgremien ISO/CEN wurden ebenfalls Arbeitsgruppen gegründet, die eine Harmonisierung der o.g. Protokolle herstellen sollen, damit SSI ein weltweiter/europäischer Standard auch für das eGovernment werden kann.

Abseits der technischen Standardisierung legt die Trust over IP Foundation[8] den Fokus auf den Governance Layer. Ziel ist es einen Standard zu schaffen, auf Basis dessen Vertrauensbeziehungen in einem dezentralen Identitätsökosystem etabliert werden können.

2.2. Community und Projekte

Aus diesen standardisierten Grundpfeilern heraus hat sich eine vielfältige Community gebildet, die verschiedenste Ausprägungen von SSI umgesetzt hat. Diese werden als *DID methods* in der *DID method registry* der W3C aufgelistet. Die dort enthaltenen SSI-

4 https://www.w3.org/TR/did-core/
5 https://www.w3.org/TR/vc-data-model/
6 https://identity.foundation/
7 https://github.com/decentralized-identity/didcomm-messaging
8 https://trustoverip.org/

Derivate unterscheiden sich mitunter deutlich, zum Beispiel in der verwendeten Organisationsstruktur, der technischen Infrastruktur und Distributed Ledger Technology, den kryptografischen Methoden und Datenschutzkonzepten. Neben der eigentlichen Infrastruktur bildet sich auch ein ganzes Ökosystem um die *wallets* und Softwareagenten, welche die angestrebte Interoperabilität der Netzwerke ermöglichen sollen.

Die Bundesdruckerei beschäftigt sich bereits seit 2018 mit SSI und konnte in einer Kooperation mit anderen Unternehmen aus der Industrie im Projekt Lissi[9] ("Let's initiate SSI") wertvolle Erfahrungen sammeln. Die frühen Ergebnisse führten zu einer Fokussierung auf den SSI-Stack der Projekte Hyperledger Indy und Aries, auf dem auch in diesem Beitrag der Schwerpunkt liegt. Darüber hinaus werden die Aktivitäten derzeit im Rahmen des Förderprojekts "Schaufenster sichere digitale Identitäten" des BMWi weitergeführt. Die Bundesdruckerei ist Teil des Konsortiums IDunion[10], welches das Ziel hat mittels SSI ein dezentrales, vertrauenswürdiges und europäisches Identitätsökosystem aufzubauen und zu etablieren. Der Aufgabenschwerpunkt ist die sicherheitsrelevante Analyse und Weiterentwicklung des technischen Gesamtsystems und die Schaffung eines Frameworks für die Nutzung im behördlichen Umfeld. Die Projektpartner planen außerdem die Umsetzung von Anwendungsbeispielen aus Industrie, Finanzwelt und Verwaltung, wie zum Beispiel den automatisierten Austausch von Stammdaten, Bankdaten, digitale Zeugnissen, digitale Personalausweise oder Zulassungsbescheinigungen.

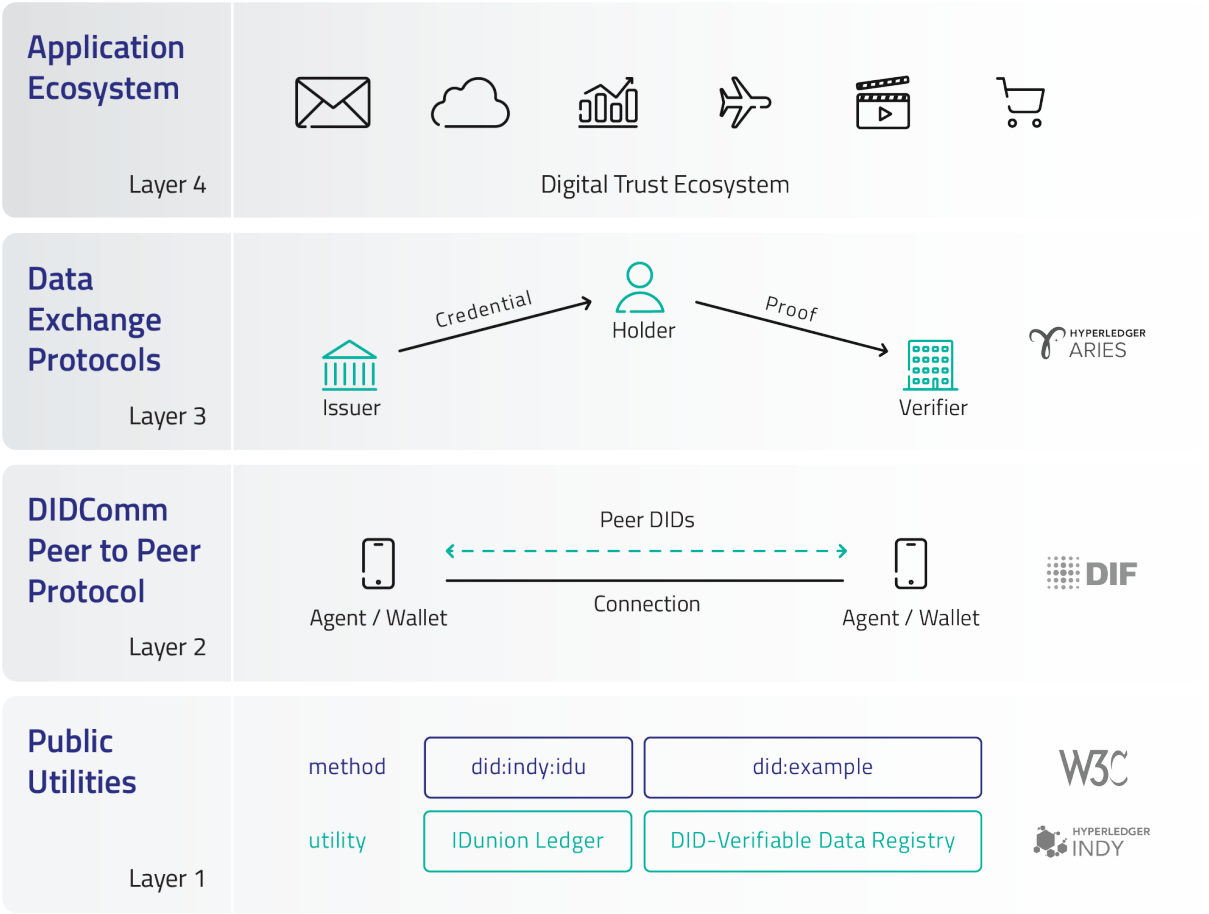
2.3. Hyperledger Indy und Aries

Unter dem Dach der Linux Foundation ist in den letzten Jahren ein sehr vielversprechender Open-Source Stack für dezentrale Identitäten entstanden. Seit 2017 wird das Projekt Hyperledger Indy[11] entwickelt, welches mit einem dedizierten public *permissioned ledger* Netzwerk ein besonders Privatsphäre förderndes Identitätsnetzwerk darstellt. Da auf dem *ledger* lediglich die öffentlichen Informationen der institutionellen *issuer* und *verifier* gespeichert sind, jedoch keine nutzerbezogenen Daten, werden so auch Probleme wie Skalierung, Effizienz und Energieverbrauch gelöst, wie sie bei früheren Blockchainprojekten gegenwärtig waren.

Mit den Projekten Hyperledger Ursa[12] und Aries[13] sollen gemeinsame, implementationsunabhängige Komponenten und technische Spezifikationen aus der Open-Source Community geschaffen werden, die eine Interoperabilität von verschiedenen Identitätssystemen vereinfachen sollen. In Ursa werden die kryptografischen Primitiven gebündelt. In Aries wird die Kommunikation zwischen zwei Agenten betrachtet, dazu gehört z.B. der Verbindungsaufbau und der Austausch von verifiable credentials. Um Eigenschaften der selektiven Offenlegung von Attributen oder der multi-show unlinkability

9 https://lissi.id/
10 https://idunion.org/
11 https://www.hyperledger.org/use/hyperledger-indy
12 https://www.hyperledger.org/use/ursa
13 https://www.hyperledger.org/use/aries

zu erreichen, werden besondere Signaturverfahren mit kenntnisfreien Beweisen (Zero-Knowledge Proof - ZKP) Beweisen verwendet.



Das Schichtenmodell von SSI

2.4. Technologie-Stack

Der typische SSI-Technologie-Stack teilt sich in vier Schichten auf. Das Fundament und die Infrastruktur werden durch einen *distributed ledger* (verteilter Datenspeicher) gebildet. Der *ledger* wird von ausgewählten Knoten betrieben und enthält die öffentlichen Daten der teilnehmenden Entitäten. Die Identitäten der Teilnehmer werden über Decentralized Identifier (*DIDs*) realisiert. Eine *DID* ist auflösbar (vergleichbar einer URL) zu einem *DID document*, welches auf dem *ledger* gespeichert ist. Im *DID document* sind Daten wie Service-Schnittstellen oder öffentliche Schlüssel hinterlegt. Die *issuer* legen zudem Metainformationen zu Credentials auf dem *ledger* ab, zum Beispiel Schemata oder Revozierungsdaten.

Die über *DIDs* repräsentierten Identitäten können auf dem darüber liegenden Layer eine kryptografisch sichere Verbindung aufbauen. Das als *DIDComm* standardisierte Protokoll umfasst alle Aktivitäten vom Verbindungsaufbau bis zum Austausch von Nachrichten. Die *DIDs* und das zugehörige Schlüsselmaterial werden in *wallets* gespeichert und von Softwareagenten verwendet.

In Layer drei sind der eigentliche Austausch und die Verbindung der drei Parteien des SSI-Ökosystems dargestellt. *issuer* beziehungsweise *verifier* stehen in direkter Verbindung zum *holder*, der *issuer* stellt dem *holder* Verifiable Credentials aus, diese können vom *verifier* angefragt und vom *holder* präsentiert werden. Das Vertrauen zwischen *verifier* und *issuer* wird indirekt über das Netzwerk abgebildet.

Schlussendlich stellt der vierte Layer die Anwendung der SSI-Technologie in Geschäftsprozessen, Behördengängen und anderen Nachweisen von Identitäten und Beglaubigungen dar. Der technologische Stack wird begleitet von einem organisatorisch-rechtlichen Rahmen, der die Regeln des Netzwerks festlegt, sowie die rechtliche und regulatorische Verbindung zur realen Welt klärt. Im Beitrag werden die verschiedenen Vertrauensbeziehungen innerhalb des SSI-Stacks näher erläutert und erklärt, welche Implikationen sich für Sicherheit und Privatsphäre daraus ableiten lassen.

3. Vertrauensbeziehungen im SSI-Ökosystem

Vertrauen ist ein subjektives und psychologisch gesellschaftsgeprägtes Empfinden. Das Entstehen von Vertrauen im institutionellen Rahmen hat ökonomische, politologische und soziologische Gründe. Insbesondere im digitalen Kontext unterscheidet sich der Vertrauensaufbau im Vergleich zu erlernten Vorgängen aus einer analogen Erziehungswelt. Dabei ist Vertrauen die Basis jedes Identitätsmanagementsystems, jedoch wurde bei der Entwicklung der Netzwerkprotokolle in den 1980er Jahren die Sicherheit und Identifizierung der Kommunikationspartner vernachlässigt, weshalb dies heute von darüber liegenden Anwendungsschichten umgesetzt werden muss. Self-Sovereign Identity kann diese Lücke als "spanning layer"[3] füllen und eine interoperable und wiederverwendbare Ebene für Identifizierungs- und Authentifizierungsprozesse abbilden.

Im Folgenden betrachten wir das Vertrauensmodell von SSI und die darin zugrunde liegenden Vertrauensbeziehungen. Eine Vertrauensbeziehung besteht dabei aus einem Vertrauensgeber und einem Vertrauensnehmer, zum Beispiel gibt der *verifier* dem *issuer* das Vertrauen eine reale Organisation zu verkörpern und nur legitime Credentials auszugeben, beziehungsweise rechtzeitig zu revozieren. Die Vertrauensbasis kann dabei sowohl organisatorischer als auch technischer Natur sein, so zum Beispiel über bereits existierende, kryptografisch gesicherte Kommunikationskanäle, oder über durch eine Governance geregelte Verträge. Wir klassifizieren im Folgenden die Vertrauensbeziehungen nach den jeweiligen Rezipienten (Vertrauensnehmer), neben den drei klassischen Akteuren *issuer*, *holder* und *verifier* kommt hier zusätzliche die Infrastruktur hinzu, also der *ledger* und die zugehörige Governance-Organisation.

3.1. Vertrauen in die Infrastruktur

Eine dezentrale Infrastruktur ist die gemeinsame Basis, auf der Identitätstransaktionen eines SSI-Ökosystems ausgeführt werden. In dem hier betrachteten Ökosystem wird der Betrieb der Infrastruktur von einem festen Kreis an Netzwerkbetreibern und durch organisatorische und technische Maßnahmen sichergestellt. Um als solide Basis für die beteiligten Akteure (*issuer*, *holder*, *verifier*) zu fungieren, muss die Infrastruktur insbesondere folgende Eigenschaften erfüllen [4]:

- Unverfälschbarkeit und eindeutige Zuordnung der Transaktion zum Autor (Authentizität der Daten)
- Überprüfbarkeit des aktuellen Zustands des *ledger* (Integrität der Daten)
- Ausfallsicherheit des Netzwerks auch bei mehreren fehlerhaften Knotenbetreibern (Verfügbarkeit, Redundanz & Robustheit)

Da es sich bei den Daten, die im öffentlich Netzwerk gespeichert werden, um nicht personenbezogene Prüfdaten handelt, die gerade öffentlich zugänglich und zu einer öffentlichen Entität zuordenbar sein sollen, spielen die beiden Schutzziele Vertraulichkeit und Anonymität keine Rolle.

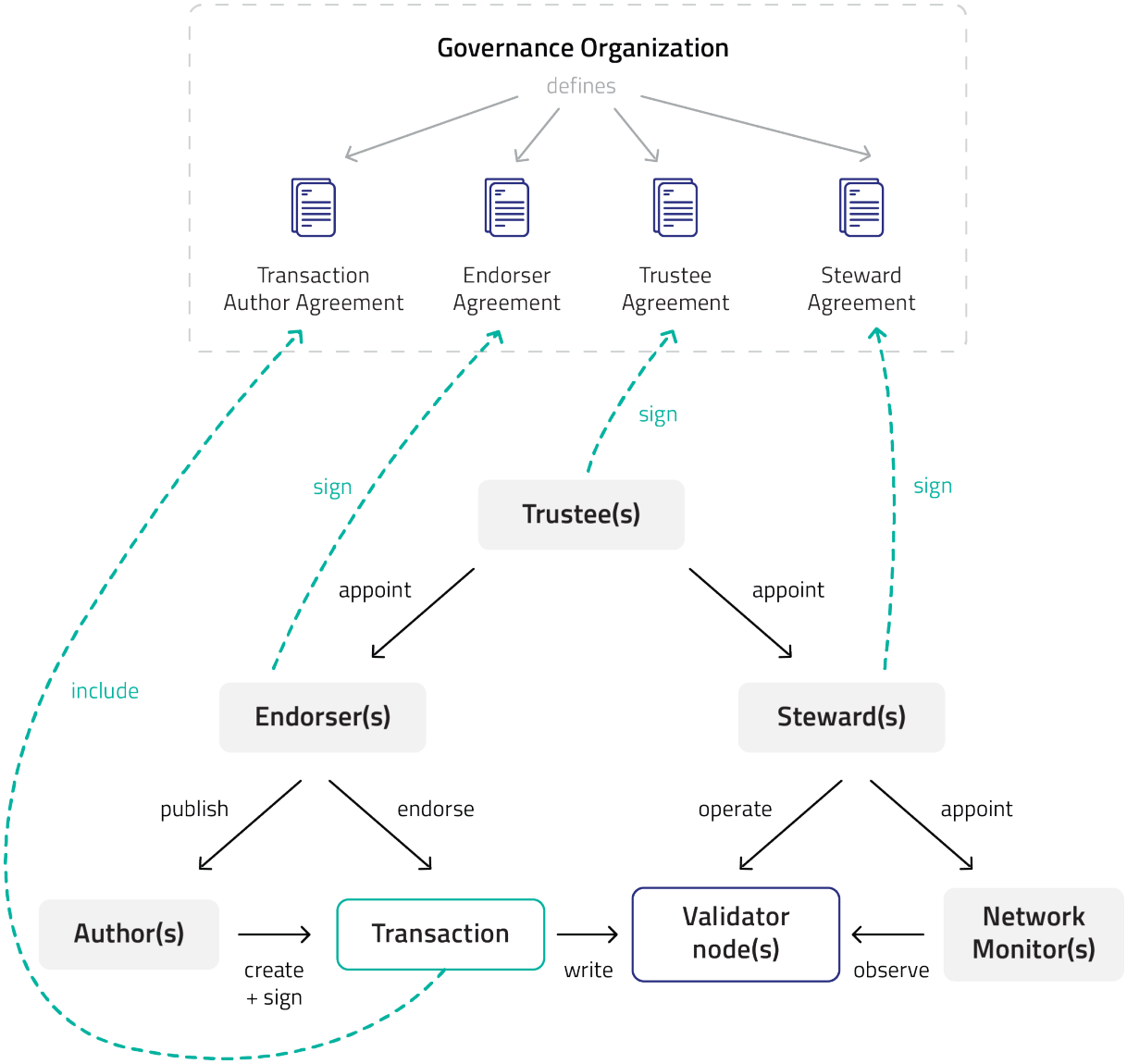
Bei der verwendeten Infrastruktur handelt es sich um Hyperledger Indy – eine öffentliche, genehmigungspflichtige (*public permissioned*) Blockchain. Diese Ledger-Technologie verbindet bestimmte Eigenschaften der öffentlichen, genehmigungslosen (*public permissionless*) Varianten wie Bitcoin und Ethereum mit jenen der privaten, genehmigungspflichtigen (*private permissioned*) Varianten wie R3 Corda und Hyperledger Fabric. Das Netzwerk besteht aus mehreren Netzwerkknoten (*validator nodes*), die einem genehmigungspflichtigen Netzwerk entsprechend von speziell befähigten beziehungsweise ernannten Akteuren (*stewards*) betrieben werden. Die validator nodes bilden untereinander Vertrauen durch den Konsensalgorithmus. Bei Hyperledger Indy wird Konsens mit dem Plenum-Protokoll gebildet, ein Redundant Byzantine Fault Tolerance Algorithmus (RBFT), welcher eine redundante Version des als sicher evaluierten PBFT[5] darstellt. Dieser bedingt für eine maximale Anzahl F von bösartig oder fehlerhaft agierenden Knoten eine Anzahl $N = 3F + 1$ von Gesamtknoten im Netzwerk. Eine sinnvolle und angestrebte Größe des Netzwerks beinhaltet circa 25 validator nodes, sodass eine gute Ausfallsicherheit gewährleistet ist und der Konsensalgorithmus möglichst effizient arbeitet [6]. Die Effizienz von RBFT und der hohe Durchsatz bei geringem Ressourcenaufwand entspricht somit den Vorteilen eines genehmigungspflichtigen (*permissioned*) *ledger*.

Die Lesezugriffe auf das Netzwerk sind frei zugänglich und erfordern keine Authentisierung, analog zu den typischen Eigenschaften einer öffentlichen (*public*) Blockchain. Die Schreibzugriffe sind jedoch genehmigungspflichtig und bedürfen neben der Signatur des Autors auch noch der Zustimmung eines *endorser*. Als initiale Transaktion erstellt ein Autor seine *DID* und hinterlegt seinen abgeleiteten öffentlichen Schlüssel. Alle weiteren Transaktionen muss er nun selbst mit seinem privaten Schlüssel digital signieren, sodass die Authentizität der Daten gewährleistet und einfach überprüfbar ist. Die so entstehenden Transaktionen schickt der Autor an mehrere Knoten und sobald er mehr als $F + 1$ erfolgreiche Antworten erhält, ist die Finalität bzw. Konsistenz gegeben. Im Netzwerk werden dabei der Konsens über die Daten mittels einer Boneh–Lynn–Shacham (BLS)-Multisignatur [7] abgespeichert, welche die zweifelsfreie Konsistenz der validator nodes beweist, sodass bei einer Leseanfrage nur ein einziger Knoten angefragt werden muss. Die gespeicherten Transaktionen werden als geordnete Liste mit einem Merkle-Tree gespeichert. Der korrekte Zustand des *ledger* kann anhand des aktuellen state proof und des Auditierpfads des Merkle-Tree überprüft werden. Die Authentizität der *stewards* kann wiederum anhand ihrer öffentlichen Schlüssel aus dem *genesis*

file sichergestellt werden. Demgegenüber wird bei Bitcoin die Konsistenz der Hashwerte der Verknüpfungen zwischen den Blöcken ausgehend von einem *genesis block* geprüft.

Um die Regeln des genehmigungspflichtigen Netzwerks durchzusetzen, hat der *ledger* seine eigene Konfiguration und ein integriertes Rechte- und Regelsystem gespeichert. Die Rollen umfassen:

- *trustee* - verwaltet die Rechte anderer Mitglieder und kann das Regelwerk festlegen
- *steward* - betreibt einen Netzwerkknoten
- *endorser* - bestätigt den Schreibzugriff regulärer Datentransaktionen
- *network monitor* - überwacht und analysiert die Metadaten für ordnungsgemäßen Betrieb des Netzwerks



Governance und Rollensystem in Hyperledger Indy

Das Vertrauen in das Netzwerk basiert also neben den *stewards* auch auf den einzelnen *trustees*, eine Art Ältesten- oder Aufsichtsrat. Das Regelwerk ist jedoch so gestaltet, dass für grundlegende Änderungen auch eine Mehrheit der *trustees* notwendig wird. Um die vertrauengebenden Rollen der technischen Basis zu komplementieren, gibt es in der Regel eine Governance-Organisation. Diese repräsentiert das Netzwerk und bildet den rechtlichen Rahmen mit der realen Geschäftswelt ab. Sie schließt Verträge mit Personen und Organisationen und bindet sie an ihre technischen Rollen und die damit gegebenen Rechte und Verpflichtungen. Diese Pflichten sichern in erster Linie das Vertrauen, dass die Akteure nach bestem Wissen und Gewissen handeln und alle notwendigen Sicherheitsvorkehrungen treffen, um das Netzwerk zu schützen. Gegenüber genehmigungsfreien Netzwerken kann die Governance-Organisation gegebenenfalls zur Haftung gezogen werden. Durch den daraus resultierenden Anreiz haben die Netzwerkbetreiber ein berechtigtes Interesse, Expertise über die Ledgersoftware zu halten und Schwachstellen möglichst schnell zu beseitigen. Ebenso können in einem genehmigungspflichtigen Netzwerk Sicherheitslücken schneller beseitigt werden, da die geringe Anzahl der Knoten über die Governance-Organisation effizient gesteuert werden kann. Allerdings muss genauso wie bei zentralen Lösungen Vertrauen in die (Open-Source) Software erbracht werden.

3.2. Vertrauen in den Issuer

Das Vertrauen in den *issuer* und dessen ausgestellte Identitätsdaten ist die Grundlage für die Identifizierung des Nutzers durch den *verifier*.

3.2.1. Identität des Issuers

Die *DID* eines *issuer* wird zusammen mit einem kryptografischen öffentlichen Schlüssel im Netzwerk bekannt gemacht. Dessen Identität kann also innerhalb des Netzwerks über diese Schlüssel verifiziert werden. Für das Vertrauen in diesen Akteur ist aber im Wesentlichen die Bindung dieser Schlüssel zu einer konkreten Organisation von Interesse (also die Antwort auf die Frage: Ist die Bank wirklich die Bank als die sie sich ausgibt?). Je nach Anwendungsfall kann die Authentizität der beteiligten Akteure natürlich über einen bilateralen Weg sichergestellt werden. Um in einem wachsenden Ökosystem skalierbare Lösungen bereitzustellen, fokussieren wir uns hier aber auf Verfahren, die eine automatische Verifikation ermöglichen, ohne vorherigen Kontakt zu erfordern.

Web Reputation

Diese Herausforderung wird derzeit im Internet anhand von Organisationsvalidierungen und PKIs gelöst. Es gibt einige Vorschläge wie diese etablierten Mechanismen bzw. die vorhandene Reputation aus dem Web in ein dezentrales Identitätsnetzwerk übernommen werden können. Der Kern der Vorschläge ist dabei, eine Verknüpfung zwischen einer Domain und einer *DID* herzustellen, um zu zeigen, dass der bereits überprüfte Domaininhaber mit dem *DID controller* übereinstimmt. Folgende Lösungsansätze wurden bis jetzt präsentiert:

DNS Record: Mayrhofer und Sabadello beschreiben in einem IETF Draft ein Verfahren bei dem *DIDs* in DNS-Records referenziert werden. Dies ermöglicht eine Zuordnung

ausgehend von einer Domain oder E-Mail-Adresse zu einer *DID*. Da bei diesem Verfahren lediglich auf DNS mit den bekannten Schwächen[14] von DNSSEC aufgebaut wird, kann dieses Verfahren eher als einfacher Discovery Mechanismus gesehen werden, ohne große Garantien bezüglich Authentizität zu gewährleisten.

did:web: Mit der did:web Methode wird eine eigene *DID method* spezifiziert (Editor's Draft), bei der eine Web-Domain Teil der *DID* ist. Das *DID* document wird dabei über diese Domain aufgerufen. Diese *DID method* kommt also ohne eine geteilte Ablage, wie z.B. einen *ledger*, für die *DID* documents aus und benutzt mit TLS einen weit verbreiteten Mechanismus für die Überprüfung der Authentizität des *DID* documents.

well-known configuration: Die *Well Known DID Configuration* ist eine in der DIF spezifizierte Methode zur Herstellung einer kryptografischen und bidirektionalen Vertrauensbeziehung zwischen einem *DID controller* und dem Inhaber einer Webressource. Namensgebend ist dabei die Verwendung der Well-Known URIs, wie sie in IETF8615 definiert ist und bereits für ACME-Challenges, CalDAV-Verzeichnisse sowie OpenID-Connect Konfigurationen verwendet wird. *Die Well Known DID Configuration* definiert einen neuen Typ, der unter dem fest definierten ".well-known" Verzeichnis Informationen zur Verifizierung eines *DID controllers* bereitstellt. Im *DID* document ist dann ebenfalls ein Link zur verknüpften Domain signiert hinterlegt.

Die Vorteile der auf Web-Reputation basierenden Verfahren liegen im schnellen initialen Aufbau für Vertrauensbeziehungen im neuen SSI-Ökosystem. Die Verknüpfung zwischen Domaininhaber und *DID controller* ist je nach Verfahren unterschiedlich stark (kryptografisch abgesichert, uni-/bidirektional). Durch den Einsatz der etablierten Mechanismen geht man aber auch Kompromisse ein, wie einer zentralistischen PKI oder einem angreifbaren DNS, die teilweise konträr zu den Leitgedanken von SSI sind.

Inhärente Organisationsvalidierung

Eine Vertrauensstruktur lässt sich jedoch auch mit den Werkzeugen des SSI-Ökosystems etablieren. Dabei können v*erifiable credential*s und *DID* Documents als Nachweis einer Organisationsüberprüfung dienen:

Credential: Analog zu den X.509-Zertifikaten in klassischen PKIs können Verifiable Credentials (VC) an Organisationen von einem autoritativen *issuer* (vgl. CA) ausgestellt werden. Dieses VC kann der Inhaber nun einer dritten Partei präsentieren um die Authentizität seiner Identität zu untermauern. Zusätzlich könnten in einer öffentlichen Credential Registry alle überprüften Organisationen eines Netzwerks präsentiert werden (siehe OrgBook BC[15]).

DID Document: Ein *DID document* kann neben einem eindeutigen Identifier und kryptografischen Schlüsseln auch Metadaten wie Unternehmensname, Anschrift und URLs beinhalten. So kann eine autoritative Entität den Inhalt, vor allem aber die Schlüssel, eines *DID* Documents signieren. Diese Signatur kann nun vom *DID controller* als weitere Metadaten zum *DID* Document hinzugefügt werden. So können Dritte ohne direk-

14 https://tools.ietf.org/html/rfc7929#section-7
15 https://www.orgbook.gov.bc.ca/

ten Kontakt mit dem *DID* Controller dessen Authentizität prüfen. Das ist vor allem hilfreich bei der Verifikation eines Proofs, da hier nur eine Verbindung zwischen *holder* und *verifier*, aber nicht zum *issuer* aufgebaut wird.

Ledger-integrierte Lösung: Eine weitere Möglichkeit wäre die Einführung einer eigenen Netzwerk-Rolle für die Validierung von Organisationen. Analog zu den *stewards* und *endorsers* könnten diese Organisationvalidatoren von den *trustees* bestimmt werden. Die Organisationvalidatoren würden nun auf transaktionaler Ebene, anstatt wie vorher auf Ebene des *DID* documents, ihre Bestätigung tätigen. Somit wäre direkt beim Schreiben der Transaktionen eine Verifikation möglich, da die Signatur nicht manuell nachgereicht werden muss (vgl. vorheriger Abschnitt). Mit Hilfe eines eIDAS-konformen Siegels kann es zukünftig ermöglicht werden die Organisationsvalidierung in einen europaweit rechtlich sicheren Rahmen einzubetten.

3.2.2. Ausgestellte Credentials des Issuers

Authentizität und Integrität der Daten

Durch technische Maßnahmen wird die Unverfälschbarkeit der ausgestellten Daten gewährleistet (Integrität, Authentizität). Der *verifier* überprüft die Unverfälschbarkeit, indem er die Signatur des *issuer* validiert. Hierbei unterscheiden sich aber die für *verifiable credentials* genutzten Signaturschemata stark, was unmittelbare Auswirkung auf die Privatsphäre der Nutzer hat. Besonders hervorzuheben sind Signaturschemata, die Zero Knowledge Proofs (ZKPs) unterstützen (zum Beispiel Camenisch-Lysyanskaya [8] und BBS+ [9][16] Signaturverfahren). Im Rahmen dieser (meist interaktiven) Signaturprüfung, erfährt der *verifier* lediglich, dass der *holder* die Kenntnis einer gültigen Signatur der offengelegten Attribute beweist. Demgegenüber stehen klassische Signaturverfahren, bei denen der *verifier* für die Überprüfung die eigentliche Signatur benötigt. Dies ermöglicht wiederum, eine Korrelation zwischen verschiedenen Präsentationen herzustellen, da sich die Signatur in der Regel nicht ändert bzw. ansonsten nur einmal benutzt werden könnte. Außerdem erschweren die klassischen Signaturverfahren Mechanismen wie die selektive Offenlegung oder logische Aussagen für beispielsweise Altersbeweise.

Ein weiterer wichtiger Bestandteil ist die Möglichkeit, ausgestellte *credentials* zurückzurufen. Sei es, weil der *issuer* fehlerhafte Daten ausgestellt hat oder durch eine missbräuchliche Nutzung seitens des *holder* oder durch Verlust der privaten Schlüssel. Zwei Punkte zeichnen sich dabei als Privatsphäre fördernde Mechanismen aus. Zum einen sollte für die Überprüfung der Gültigkeit keine direkte Verbindung zum *issuer* aufgebaut werden müssen. Stattdessen sollten öffentlich zugänglich und hochverfügbare Prüfdaten aus z.B. dem *ledger* verwendet werden. Zum anderen sollten auch hier ZKPs benutzt werden, so dass auch bei dieser Prüfung eine Korrelation zwischen verschiedenen Vorgängen derselben Person verhindert werden. Bei Hyperledger Indy wird dies durch einen Prüfalgorithmus basierend auf einem kryptografischen Akkumulator [10] umgesetzt.

16 https://mattrglobal.github.io/bbs-signatures-spec/

Semantische Bedeutung der Daten

Ein gemeinsames Verständnis der Bedeutung der Daten ist entscheidend für ein offenes und interoperables Identitätsökosystem, bei dem mehrere *issuer* und *verifier* agieren und der Nutzer nicht für jede dieser Kombinationen ein eigenes Credential benötigen soll. Dieses Verständnis kann durch die Einigung auf gemeinsame Schemata, wie z.B. auf schema.org erreicht werden. So können diese für die Strukturierung bzw. Annotation der Daten genutzt werden. Im Rahmen der Verifiable Credentials bietet sich vor allem die Strukturierung mittels JSON-LD an, wie sie z.B. auch in *rich schemas*[17] verwendet wird.

Qualität der Daten

Neben der semantischen Bedeutung ist die Qualität und das daraus resultierende Vertrauensniveau der präsentierten Daten für den *verifier* entscheidend. Zum Beispiel könnte die Anschrift, die der *holder* einem Online-Shop (*verifier*) übermittelt, aus einem hoheitlichen Melderegister stammen oder aber ursprünglich vom *holder* selbst angegeben sein. Das in eIDAS spezifizierte Level of Assurance (LoA) ist beispielhaft ein Maß für die Qualität der Daten von notifizierten Identitätsschemata. Genauso liegt auch bei SSI die Verantwortung für Prüfung und Bekanntgabe der Qualität der Daten beim *issuer* bzw. bei Notifizierungsstellen. Die *eIDAS bridge*[18], die im Rahmen des eSSIF Projekts[19] entsteht, soll es ermöglichen das LoA substantial für Verifiable Credentials zu erreichen. Dabei benutzt der *issuer* für die Signatur der *credentials* ein nach eIDAS reguliertes Siegel. Da diese Siegel derzeit nicht kompatibel zu ZKPs sind, können die oben genannten Privatsphäre fördernden Eigenschaften bisher mit der *eIDAS bridge* nicht erreicht werden.

Diese Ansätze können letztendlich kombiniert werden, um in einem t*rust framework* branchenübergreifende Standards für vertrauenswürdige Organisationen und *credentials* zu definieren. Individuen oder Dachorganisationen können ihre jeweiligen qualitativen Anforderungen in sogenannten *trust schemes* transparent zusammenstellen und vertrauenswürdig auf dem *ledger* hinterlegen. Diese Strukturen existieren oft bereits in der realen Welt und können einfach in die digitale SSI-Welt überführt werden. Somit wird beispielsweise ermöglicht, dass ein Handelsunternehmen nicht eine Vielzahl von bilateralen Vertrauensbeziehungen zu verschiedenen Banken unterhalten muss, sondern diese über das *trust scheme* des Verbands europäischer Banken bezieht. Durch eine Harmonisierung der unterschiedlichen existierenden *trust schemes*[20] wurde die Grundlage geschaffen, um diese im Rahmen von eSSIF zu integrieren[21].

3.3. Vertrauen in den Verifier

Die Vertrauenswürdigkeit der Identität des *verifiers* ist vor allem für den *holder* von Interesse. Dem *verifier* übermittelt der *holder* letztendlich meist personenbezogene Daten, und gleichzeitig ist es eine schwierige Aufgabe, den *holder* in die Lage zu versetzen

17 https://github.com/hyperledger/aries-rfcs/tree/master/concepts/0250-rich-schemas
18 https://joinup.ec.europa.eu/collection/ssi-eidas-bridge
19 https://essif-lab.eu/
20 https://www.lightest.eu/
21 https://essif-lab.eu/essif-train-by-fraunhofer-gesellschaft/

selbst die Überprüfung durchzuführen. Im deutschen eID-System werden Diensteanbieter erst nach einer zentralen Prüfung dazu berechtigt Ausweisdaten auszulesen. Ein solches Berechtigungssystem ist flächendeckend für ein SSI-Ökosystem sicherlich schwierig umzusetzen und würde wahrscheinlich maßgeblich die Verbreitung einschränken. Nichtsdestotrotz geben die eingesetzten Mechanismen von SSI die Möglichkeit, dass sich der *verifier* zuerst gegenüber dem *holder* "ausweist", indem er selbst ein Credential über seine Institution präsentiert. Die Frage stellt sich nun, wer den Vertrauensanker für den *holder* darstellt und wie er in seiner *wallet* über die Authentizität der Identität des *verifier* informiert wird. Die *wallet* könnte diese Information dem *holder* in einfacher Weise vermitteln, wie es zum Beispiel mit HTTPS im Browser oder verifizierten Accounts auf Social Media Plattformen (z.B. blauer Haken bei Twitter) zeigen. Hier sind weitere Konzepte und Entwicklungen zu erarbeiten.

Datenminimierung

Datenminimierung ist einer der Grundsätze der Datenschutz Grundverordnung (DSGVO). Die von SSI bereitgestellten Mechanismen unterstützen deren Umsetzung und erschweren so den Missbrauch durch einen *verifier*, der mehr Daten anfordert als für seinen Zweck benötigt werden. Zum einen ermöglichen die eingesetzten Zero-Knowledge Proofs eine datensparsame selektive Offenlegung und Altersbeweise. Zum anderen werden beim Nutzer automatisch die Transaktionsdaten bei einer Offenlegung von persönlichen Daten gespeichert. Dies ermöglicht einen zweifelsfreien Nachweis, falls ein *verifier* gegen die DSGVO verstoßen sollte. Ergänzend empfehlen wir, dass der Zweck der Datenanfrage vom *verifier* angegeben werden muss und auch in den eingesetzten Protokollen hinterlegt wird.

3.4. Vertrauen in den Holder

Die offenen Protokolle der Self-Sovereign Identity-Technologie ermöglichen ein hohes Maß an Interoperabilität und bieten dem Nutzer die freie Wahl einer *wallet*. Dieses offene Ökosystem ermöglicht jedoch potenziell auch die Entwicklung missbräuchlicher *wallets*, die die Extraktion und Weitergabe von Credentials und derer Schlüsselmaterial möglich machen. Somit haben der *issuer* und der *verifier* ein berechtigtes Interesse, dass die ausgestellten Credentials sicher verwahrt und vor Missbrauch geschützt werden, denn der *issuer* riskiert sonst seine Glaubwürdigkeit und der *verifier* fundiert seine Geschäftsprozesse auf der Annahme richtiger Daten. Um eine Vertrauensbeziehung zwischen dem *issuer* und dem *holder* zu erreichen, muss eine Bindung zwischen der Identität des Nutzers und der ihm ausgestellten Credentials gewährleistet sein. Der technische Identitätsnachweis mittels kryptografischer Schlüssel und Signaturen ermöglicht eine Bindung an eine bestimmte *wallet*, nicht jedoch an den Besitzer der *wallet*. Dafür muss der Nutzer neben der *wallet* einen zusätzlichen Authentisierungsfaktor durch Wissen oder Biometrie erbringen. Die gängigen SSI-*wallets* sind heute bereits mindestens mit einer PIN oder Biometrie abgesichert, um den Zugang zur *wallet* freizuschalten; diese werden beim Einrichten der *wallet* selbst erstellt. Da das Ergebnis dieser Authentisierung weder dem *issuer* noch dem *verifier* mitgeteilt werden kann, schützt der Mechanismus nur den *holder* vor dem Missbrauch nach dem Verlust der *wallet*. An einer

weiterreichenden Lösung für das ganze Ökosystem wird derzeit gearbeitet und die verschiedenen Ansätze und Konzepte mit dem Fokus auf Biometrie dargestellt. Für einen vertrauenswürdigen *holder* müssen dabei folgende Eigenschaften erfüllt werden:

- Nutzerbindung der *credentials* (Authentizität der Nutzeridentität)
- Gerätebindung der *credentials* (Integrität der *wallet*)
- Nachweisbarkeit gegenüber *issuer* und *verifier* (Authentizität der *wallet*)

3.4.1. Authentizität der Nutzeridentität

Um eine Nutzerauthentisierung beispielsweise mittels Biometrie zu realisieren, muss die Architektur zwei wesentliche Designentscheidungen treffen: Zum einen muss geklärt werden, wo die schützenswerten biometrischen Referenzdaten gespeichert sind, und zum anderen muss festgelegt werden, welche Komponente des Ökosystems den Abgleich der Referenzdaten mit den Authentisierungsdaten vornimmt. Daraus ergeben sich dann direkte Implikationen für die Sicherheit der biometrischen Daten, die Anforderungen an die *wallet* und die Portabilität der *wallet*. Nach derzeitigem Forschungsstand ergeben sich daraus drei wesentliche Szenarien, die im Folgenden dargestellt werden.

Varianten	**Biometrische Daten**			**Wallet**	
	Speicherort	**Vergleichsort**	**Sicherheit**	**Anforderungen**	**Portabilität**
transparent wallet	im Credential	extern	niedrig	niedrig	hoch
trusted wallet	im Credential	*wallet* intern	mittel	mittel	hoch
secure wallet	in der *wallet*	*wallet* intern	hoch	hoch	mittel

In der ersten Variante ("transparent wallet") kommt der *wallet* des *holders* mit Hinblick auf die Nutzerauthentisierung keine sicherheitsrelevante Bedeutung zu. Die biometrischen Daten sind in den Credentials selbst verankert und der Abgleich findet durch eine externe Komponente namens "biometric service provider" statt[22]. Dadurch sind die sicherheitsspezifischen Anforderungen an die *wallet* gering, sie stellt hauptsächlich die Nutzerschnittstelle für den *holder* dar und ist somit vollständig portabel. Neben dem weiterhin erforderlichen Vertrauen in die *wallet*, müssen in diesem Szenario alle Entitäten des Ökosystems zusätzlich dem externen b*iometric service provider* vertrauen. Die biometrischen Daten werden in dieser Variante an *issuer* und *verifier* weit über das Ökosystem gestreut, was aus Sicht des Datenschutzes bedenklich ist.

In der zweiten Variante ("trusted wallet") sind die biometrischen Referenzdaten abermals in den *Credentials* verankert, aber der Vergleich der biometrischen Daten findet dieses Mal lokal in der *wallet* statt. Dieser Ansatz verbessert die Privatsphäre, da bei einem Beweis eines *credential* gegenüber dem *verifier* keine biometrischen Daten übertragen werden. Der *verifier* braucht allerdings eine neue Vertrauensbasis zum *holder*, da

22 https://github.com/hyperledger/aries-rfcs/tree/master/concepts/0231-biometric-service-provider

er den biometrischen Abgleich nicht selbst vornimmt, sondern der Software des Wallet-Herstellers vertrauen muss. Wir schlagen dafür die Spezifizierung einer *trusted wallet* vor. Nach diesem Verfahren können die Hersteller einen Auditierungsprozess durchlaufen und sich so ihre SSI-Wallet zertifizieren lassen. Im Laufe des c*redential* Nachweises kann der *verifier* die Zertifizierung der *wallet* überprüfen und somit der biometrischen Authentisierung vertrauen.

In einer dritten Variante ("secure wallet") findet sowohl die Speicherung der biometrischen Referenzdaten als auch der Vergleich innerhalb einer zertifizierten *wallet* statt. Hiermit findet eine indirekte Bindung der *credentials* an den Nutzer statt, da die *credentials* an die *wallet* gebunden werden und sich der *holder* gegenüber der *wallet* authentisiert. Die Anforderung der biometrischen Überprüfung wird hierbei in den Metadaten der *credentials* vermerkt und automatisch bei der Nutzung von der *wallet* veranlasst. Das hat den Vorteil, dass die biometrischen Daten bestmöglich geschützt sind. Allerdings wird dabei die freie Portabilität der *wallet* eingeschränkt, da alle biometrisch gebundenen *credentials* nun auch an die *wallet* gebunden sind.

Die Konzeption und Spezifizierung dieser Möglichkeiten wird zukünftig in der DIF vorangetrieben.

3.4.2. Integrität und Authentizität der Wallet

Die spezifizierten Schnittstellen einer "Secure Wallet" ermöglichen dem *issuer* eine Abschätzung über das Sicherheitsniveau der *wallet*. Diese ermöglicht neben der Speicherung der biometrischen Daten auch die Absicherung der kryptografischen Schlüssel des *credentials* in einer besonders geschützten Zone. Darauf basierend kann der *issuer* entscheiden ob und mit welchem "Level of Assurance" er sicherheitsrelevante *credentials* ausgibt. Dies ist besonders in regulierten Anwendungsumfeldern gefordert, z.B. GWG, eIDAS.

Die OPTIMOS-Technologie kann als Sicherheitsanker für die *wallet* des Nutzers verwendet werden. So können mit der Verwendung des Secure Elements die Daten/Schlüssel der *wallet* geschützt, eine Gerätebindung erreicht und eine missbräuchliche Nutzung durch Kopieren der sensiblen Daten verhindert werden. SSI bietet weitere vielversprechende Möglichkeiten für zukünftige dezentrale Schlüsselmanagementkonzepte[23]. Beispielsweise ermöglichen *hot wallets* die Verwendung von Cloud-basierten Hardware-Sicherheitsmodulen.

4. Zusammenfassung

Die Prinzipien von SSI bieten Erfolg versprechende Lösungen für die aktuellen Herausforderungen der digitalen Souveränität in einer zunehmend von globalen Technologieunternehmen bestimmten Welt. Das SSI-Ökosystem bietet eine universale Vertrauensschicht für dezentrales und digitales Identitätsmanagement. In unserem Beitrag zeigen wir, dass die vertrauensbildenden Komponenten stark miteinander verwoben sind und einen unterschiedlichen Reifegrad haben. Der Paradigmenwechsel zu einem dezentralen und nutzerorientierten System bringt neue Herausforderungen mit sich. Insbesondere

23 https://github.com/hyperledger/aries-rfcs/blob/master/concepts/0051-dkms/dkms-v4.md

die größere Verantwortung des Nutzers bringt neue Konzepte z.B. des Schlüsselmanagements und der Wiederherstellungsmechanismen hervor. Die derzeitigen Aktivitäten in der Standardisierung und in der Open-Source Entwicklung geben Anlass zu der Erwartung einer vielversprechenden Zukunft.

Literaturhinweise

[1] Camenisch, Jan, and Els Van Herreweghen. "Design and implementation of the idemix anonymous credential system." Proceedings of the 9th ACM conference on Computer and communications security. 2002.

[2] Paquin, Christian, and Greg Zaverucha. "U-prove cryptographic specification v1. Technical Report, Microsoft Corporation (2011).

[3] Clark, David D. "Interoperation, open interfaces and protocol architecture." The Unpredictable Certainty: White Paper (1995): 33-144.

[4] Berghoff, Christian, et al. "Blockchain sicher gestalten." Konzepte, Anforderungen, Bewertungen. Bundesamt für Sicherheit in der Informationstechnik (Hrsg) Bundesamt für Sicherheit in der Informationstechnik, Bonn (2019).

[5] Castro, Miguel, and Barbara Liskov. "Practical Byzantine fault tolerance and proactive recovery." ACM Transactions on Computer Systems (TOCS) 20.4 (2002): 398-461.

[6] Sedlmeir, Johannes, et al. "The DLPS: A New Framework for Benchmarking Blockchains." Proceedings of the 54th Hawaii International Conference on System Sciences.

[7] Boneh, Dan, Ben Lynn, and Hovav Shacham. "Short signatures from the Weil pairing." International conference on the theory and application of cryptology and information security. Springer, Berlin, Heidelberg, 2001.

[8] Camenisch, Jan, and Anna Lysyanskaya. "A signature scheme with efficient protocols." International Conference on Security in Communication Networks. Springer, Berlin, Heidelberg, 2002.

[9] Au, Man Ho, Willy Susilo, and Yi Mu. "Constant-size dynamic k-TAA." International conference on security and cryptography for networks. Springer, Berlin, Heidelberg, 2006.

[10] Camenisch, Jan, Markulf Kohlweiss, and Claudio Soriente. "An accumulator based on bilinear maps and efficient revocation for anonymous credentials." International workshop on public key cryptography. Springer, Berlin, Heidelberg, 2009.

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Quantencomputerresistente Kryptografie: Aktuelle Aktivitäten und Fragestellungen

Dr. Tobias Hemmert[1], Dr. Manfred Lochter[1], Prof. Dr. Daniel Loebenberger[2], Prof. Dr. Marian Margraf[2], Stephanie Reinhardt[1], Prof. Dr. Georg Sigl[2]

Kurzfassung:
Quantentechnologie stellt eine vielversprechende Innovation der nächsten Jahre dar. Auf die IT-Sicherheit hat diese Technologie zweierlei Auswirkungen: So stellen einerseits kryptanalytisch relevante Quantencomputer eine Disruption der gegenwärtig genutzten Kryptografie dar und machen es schon heute notwendig, entsprechend proaktiv tätig zu werden und alternative, quantenresistente Verfahren zu erproben. Die Erprobung dieser sogenannten *Post-Quanten-Kryptografie* geschieht im Rahmen zahlreicher Aktivitäten der angewandten Forschung national wie international, um die neuartigen Verfahren in die Praxis zu überführen. Andererseits kann die Quantentechnologie genutzt werden, um auf Basis quantenmechanischer Mechanismen kryptografische Systeme zu konstruieren. Diese sogenannte *Quantenkryptografie* wird beispielsweise im Kontext des Quantenschlüsselaustauschs (engl. *Quantum Key Distribution*) in vielen Projekten zur Anwendung gebracht, auch wenn sie Beschränkungen mit sich bringt und bezüglich der Sicherheit noch einige Fragen zu beantworten sind. In dieser Arbeit gehen wir auf die unterschiedlichen Projekte ein und erläutern, wie sich diese zueinander einordnen lassen.

Stichworte: Migrationsempfehlungen, Post-Quanten-Kryptografie, QKD, Quantenkryptografie, Quantum Key Distribution

1. Einleitung

Die fortschreitende Entwicklung von Quantencomputern stellt eine Bedrohung für die heute verbreitete Public-Key-Kryptografie dar [24], [37]. Dazu zählen insbesondere Schlüsseleinigungsverfahren und digitale Signaturen. Es ist eine etablierte Arbeitshypothese im Kontext hochsicherer Anwendungen [28], dass Anfang der 2030er Jahre ein entsprechend leistungsfähiger Quantencomputer existiert, der gängige kryptografische Verfahren, wie beispielsweise RSA-2048, brechen kann [23]. Deshalb ist es wichtig, quantencomputerresistente Verfahren zu entwickeln und rechtzeitig in die Praxis zu überführen.

Die Dringlichkeit dieser Herausforderung liegt darin begründet, dass die Umstellung auf quantencomputerresistente Verfahren letztlich eine Migration weiter Teile der Infrastruktur notwendig macht. Die hierbei anzupassenden Komponenten reichen von den angesprochenen kryptografischen Primitiven über die genutzten Kommunikationsprotokolle bis hin zu kryptografischer Hardware. Ferner haben sensible Daten oft einen sehr langen Schutzbedarf, zum Beispiel bis zu 30 Jahre für Dokumente mit einem Geheimhaltungsgrad von mindestens VS-NfD [4].

[1] Bundesamt für Sicherheit in der Informationstechnik, Bonn
[2] Fraunhofer AISEC, Garching bei München

Im Großen und Ganzen lassen sich zwei Ansätze für quantencomputerresistente Verfahren unterscheiden.

In der Post-Quanten-Kryptografie werden klassische kryptografische Algorithmen entwickelt, deren Sicherheit auf mathematischen Problemen beruht, für deren Lösung bisher weder ein effizienter Quantenalgorithmus noch ein effizienter klassischer Algorithmus bekannt ist. Das BSI erarbeitete im Frühjahr 2020 Handlungsempfehlungen einer "Migration zu Post-Quanten-Kryptografie" [11], in welchen konkrete Schritte genannt werden, die bei einer entsprechenden Anpassung der Infrastruktur heute schon denkbar sind (siehe auch den Beitrag "Standardisierung von Post-Quanten-Kryptografie und Empfehlungen des BSI" zum diesjährigen BSI-Kongress). Parallel dazu werden seit Herbst 2019 verschiedene Anwendungsprojekte des Bundesministeriums für Bildung und Forschung durchgeführt, die unterschiedliche Anwendungen im Kontext der Post-Quanten-Sicherheit praktisch erforschen [5].

In Abgrenzung hierzu werden in der Quantenkryptografie, insbesondere beim quantenbasierten Schlüsselaustausch (Quantum Key Distribution, kurz QKD), Verfahren entwickelt, deren Sicherheit auf quantenphysikalische Prinzipien zurückgeführt werden soll. Zurzeit wird QKD in zahlreichen Projekten im Rahmen von Teststrecken sowohl auf nationaler als auch internationaler Ebene erprobt. Darüber hinaus unternimmt das BSI in Zusammenarbeit mit dem Europäischen Institut für Telekommunikationsnormen (ETSI) erste Schritte zur Sicherheitszertifizierung von QKD-Produkten.

Die vorliegende Arbeit soll einen Überblick über aktuelle Projekte in den beiden Bereichen Post-Quanten-Kryptografie und Quantenkryptografie geben und wichtige Fragestellungen aufzeigen, die noch zu bewältigen sind.

2. Post-Quanten-Kryptografie

Wir gehen nun auf die aktuellen Aktivitäten im Bereich der Anwendbarkeit von Post-Quanten-Kryptografie in der Praxis ein. In Kapitel 3 behandeln wir entsprechende laufende Aktivitäten der Quantenkryptografie.

2.1. Laufende und abgeschlossene Förderprojekte

Sichtet man die Liste der laufenden Projekte in der Förderlinie „Post-Quanten-Kryptografie" des BMBF [5], so stößt man auf sieben bewilligte Projekte, welche seit Ende 2019 von unterschiedlichen Konsortien bearbeitet werden. Die Konsortien sind in den Projekten dabei derart gestaltet, dass sowohl Universitäten bzw. außeruniversitäre Forschungsinstitutionen als auch Industrieunternehmen kollaborativ an den Projektinhalten arbeiten. Dies soll einen direkten Erkenntnistransfer in die Wirtschaft ermöglichen.

So bearbeiten verschiedene Projekte die Herausforderungen, die damit einhergehen, Post-Quanten-Kryptografie in eingebetteten Systemen zu erproben. Dabei sind die limitierten Ressourcen derartiger Systeme ein zentraler Untersuchungsaspekt. Ebenso wird hier

analysiert, wie sich Implementierungen in Software und in Hardware in der Praxis verhalten und wie Konzepte des Hardware-Software-Codesigns (kurz HW/SW-Codesign) zu effizienten Implementierungen führen können.

Auf Protokollebene werden Fragen zu einer vollständig quantenresistenten, abwärtskompatiblen Public-Key-Infrastruktur und die Entwicklung von quantenresistenten virtuellen privaten Netzen (VPNs) untersucht. Letztere machen es neben dem Einsatz von Post-Quanten-Kryptografie ebenso nötig, bestimmte Netzprotokolle auf die neuartigen Anforderungen der Post-Quanten-Kryptografie, wie beispielsweise einer Größe von Schlüsselängen im Bereich mehrerer Megabyte, anzupassen. Da dies nur im internationalen Konsens möglich ist, laufen parallel entsprechende Standardisierungsbemühungen, z. B. in verschiedenen Arbeitsgruppen der Internet Engineering Task Force (IETF) und der Internet Research Task Force (IRTF), etwa zu den Protokollen TLS [13], [39] oder IPsec/IKE [38], [40].

Einzelne Projekte erproben den Einsatz von Post-Quanten-Kryptografie in konkreten Einsatzbereichen, wie der Medizintechnik, in Anlagensteuerungssystemen oder auch in Chipkarten.

Schließlich befasst sich ein konkretes Projekt mit der Erweiterung der kryptografischen Open-Source Bibliothek Botan [6], welche vom BSI als geeignet für den Hochsicherheitsbereich eingestuft und in bestimmten Einsatzszenarien auch empfohlen wird [7]. Neben schnellen Softwareimplementierungen werden auch Open-Source Hardwarerealisierungen für besonders zeitkritische oder stromsparende Anwendungen untersucht. Ein weiteres laufendes Projekt, gefördert vom Bundesministerium für Wirtschaft und Energie, untersucht den Einsatz quantenresistenter Verfahren für hoheitliche Dokumente.

Auch sind einzelne Projekte bereits erfolgreich abgeschlossen worden. Als Beispiel nennen wir den Einsatz quantenresistenter hashbasierter Signaturverfahren [25], [30] zur Absicherung von Software-Updates: Hierbei wird der eigentliche Übergang zu Post-Quanten Kryptografie zunächst für den Update-Prozess von Software realisiert. Die Prüfung der Korrektheit neuer Software kann so schon heute quantensicher erfolgen, wobei die eigentliche Software selbst noch nicht vollständig migriert sein muss [22]. Zu beachten ist allerdings, dass die klassische kryptografische Funktionalität hierbei perfekt vorwärtssicher (engl. ist das die sog. *perfect forward secrecy*) sein muss, sodass das Bekanntwerden von dort verarbeiteten Langzeitgeheimnissen keine Auswirkungen mehr auf die Sicherheit hat.

2.2. Migrationsempfehlungen des BSI

Wir fassen in diesem Abschnitt in aller Kürze die aktuellen Migrationsempfehlungen des BSI [11] zusammen. Details hierzu finden sich im Beitrag "Standardisierung von Post-Quanten-Kryptografie und Empfehlungen des BSI" zum diesjährigen BSI-Kongress. Konkret werden in den Empfehlungen folgende Aspekte behandelt:

- Kryptoagilität
- Firmware- und Software-Updates
- Symmetrische Kryptografie
- Kurzfristige Maßnahmen
- Hybride Lösungen
- Anpassung von Protokollen
- Schlüsselaustausch und Schlüsseleinigung

Die Empfehlungen behandeln allesamt mögliche Erwägungen, die im Kontext eines (maßvollen) Risikomanagements beachtet werden sollten, um so einen Schritt-für-Schritt Übergang zu einer quantenresistenten Infrastruktur realisieren zu können.

2.3. Auswirkungen der Migrationsempfehlungen auf Software und Hardware

Die geforderte Kryptoagilität kann in Software erwartungsgemäß leichter erreicht werden als in Hardware. Aber auch hier wird untersucht, wie sich beispielsweise heutige Hardwarebeschleuniger für asymmetrische Kryptografie auch für neue Verfahren nutzen lassen.

2.3.1. Ansätze für Software

Bei der Software gibt es insbesondere bei Gitter-basierten Verfahren wenig Probleme bei der Migration. Die Rechenzeiten sind vergleichbar mit heute eingesetzten Verfahren. Der erhöhte Speicherbedarf muss allerdings durch ausreichende Platzreserven berücksichtigt werden. Auch Isogenie-basierte und Code-basierte Verfahren können auf leistungsstarken Prozessoren ähnlich wie Gitter-basierte Verfahren bereits heute eingesetzt werden. Daher ist die Integration entsprechender Post-Quanten Verfahren in aktuelle kryptografische Bibliotheken der hier momentan zu verfolgende Ansatz.

Ebenso ist die Anpassung der genutzten Kommunikationsprotokolle in Entwicklung. Hier ist oft ein (internationaler) Standardisierungsprozess zu durchlaufen, um interoperable Protokolle zu erhalten, die die Eigenheiten quantenresistenter kryptografischer Algorithmen, wie z.B. sehr große Schlüssellängen, mit einbeziehen und auch den hybriden Einsatz dieser neuartigen Verfahren mit klassischen, etablierten Verfahren ermöglichen.

Immer möglich ist das Anpassen der aktuell empfohlenen Sicherheitsparameter [10] auf Größen, die nicht in Reichweite erster kryptanalytisch relevanter Quantencomputer fallen. Derartige Schritte können aber von Natur aus nur vorübergehend sein und sind auch nicht ohne weiteres auf Hardware-Implementierungen übertragbar, auf die wir im Folgenden eingehen.

2.3.2. Ansätze für Hardware

In Hardware gibt es die Möglichkeit, eigene Hardwarebeschleuniger oder Coprozessoren für die Algorithmen zu realisieren. Hardwarebeschleuniger erfordern eine weitgehende Standardisierung der Verfahren und Parameter. Durch Implementierung eines Hardwarebeschleunigers ist man aber auch auf ein Verfahren festgelegt, was die Kryptoagilität stark einschränkt.

Ferner ist zu beachten, dass eine sichere und effiziente Implementierung von Post-Quanten Algorithmen in elektronischen Geräten, wie Smartcards, Mobilfunkgeräten und IoT Geräten, aufgrund limitierter Ressourcen und harten Performance-Voraussetzungen eine große Herausforderung darstellt. Je nach Anwendung müssen verschiedene Optimierungsziele erreicht werden: hohe Sicherheit, schnelle Reaktionszeiten und/oder geringe Verlustleistung. Insbesondere das Generieren von Gauß- oder binomialverteilten Zufallsvariablen, die Verwendung von Hashfunktionen und die Arithmetik im Polynomring gehören zu den ressourcenintensiven Operationen der Post-Quanten Kryptografie.

Deshalb bietet sich hier HW/SW-Codesign für Post-Quanten-Kryptografie an, indem man entweder eigene Coprozessoren konstruiert oder aber neue Instruktionen in bestehende Prozessoren integriert. Dies hat den Vorteil einer gewissen Flexibilität. Häufig benutzte Operationen werden in Hardware realisiert und andere in Software. Die Hardware-Blöcke können auch meist flexibel durch Software konfiguriert und angesteuert werden.

Während Hardware-Designs besonders effizient sind, benötigen Software-Lösungen meist wenig Chipfläche und gewährleisten eine hohe Flexibilität. Um die Vorteile der jeweiligen Ansätze zu kombinieren, wurde in [18] eines der ersten Hardware/Software Co-Designs für Post-Quanten-Kryptografie vorgestellt. Die Autoren implementierten die standardisierte IEEE-1363.1 Version von NTRU. In diesem Design werden die Software-Operationen auf einem ARM Cortex-A53 ausgeführt und NTRU durch einen Co-Prozessor für die Polynommultiplikation beschleunigt. Im letzten Jahr führte die Verbreitung von RISC-V, einer freien und offenen Befehlssatzarchitektur, zu einer neuen Ära in der Entwicklung von Hardware für Post-Quanten-Kryptografie. Während die ersten RISC-V basierten Post-Quanten Designs lose gekoppelte Hardware-Beschleuniger entworfen haben [19], entwickelten die Autoren in [20], [21] Befehlssatz- und Prozessorerweiterungen, z. B. Instruktionen für eine effiziente Polynommultiplikation mittels der sogenannten Number Theoretic Transform (NTT).

2.4. Ausblick

Es ist zu erwarten, dass es in den nächsten Jahren zu einem festen, international standardisieren Satz an neuen Post-Quanten Verfahren kommen wird [31]. Ebenso schreitet die Arbeit an der Anpassung einschlägiger Internetprotokolle voran. Die Erkenntnisse im Bereich der angewandten Forschung, die wir in den laufenden Anwendungsprojekten schon heute sammeln, werden diese internationalen Entwicklungen durch einschlägige nationale Fortschritte in unterschiedlichen Anwendungsbereichen ergänzen.

Die Herausforderung der Anpassung unserer digitalen Infrastruktur bleibt allerdings. Der Einsatz sicherer Software-Updates ist hierbei wohl ein erster, praktikabler Schritt, der etwas Zeit verschafft: So müssen die eigentlichen Anwendungsprogramme, wenn die Update Prozesse entsprechend quantenresistent realisiert sind, noch nicht migriert sein, um trotzdem ein nach heutigem Kenntnisstand quantenresistentes Gesamtsystem zu erhalten, sofern die

eingesetzte Kryptografie perfekt vorwärtssicher ist bzw. keine Langzeitgeheimnisse verarbeitet werden.

Anders sieht es mit hoch evaluierten Hardware-Komponenten (wie Smartcards) aus, die heute beispielsweise zur Absicherung kryptografischer Langzeitgeheimnisse genutzt werden. Deren Migration auf Seitenkanal-, Fehlerattacken- und Quantencomputer-resistente Lösungen scheint noch einiges an angewandter Forschung zu erfordern.

3. Quantenkryptografie

Eine Alternative zur Post-Quanten-Kryptografie ist die sogenannte Quantenkryptografie. Während die Sicherheit der Post-Quanten-Kryptografie auf der angenommenen Schwierigkeit bestimmter mathematischer Probleme beruht, verspricht – in Abgrenzung hierzu – die Quantenkryptografie Sicherheit auf Basis physikalischer Prinzipien zu erreichen. Besondere Aufmerksamkeit erfährt die Quantum Key Distribution (QKD), die quantencomputerresistenten Schlüsselaustausch ermöglichen soll. Im Folgenden wollen wir zunächst eine kurze Erklärung der groben Funktionsweise von QKD geben. Danach stellen wir dar, welche Chancen und Einschränkungen die Quantenkommunikation birgt, besonders im Vergleich mit Post-Quanten-Kryptografie. Schließlich stellen wir eine Auswahl von Projekten im Bereich QKD vor und geben eine Einschätzung zum derzeitigen Einsatz von QKD als quantencomputerresistente Technologie zur Schlüsseleinigung.

3.1. Funktionsweise von Quantum Key Distribution

Die klassischen Public-Key-Verfahren, die durch einen potenziellen Quantencomputer gebrochen würden, werden heute in der Regel dazu benutzt, um zwischen zwei Parteien – üblicherweise mit Alice und Bob bezeichnet – einen geheimen Schlüssel zu vereinbaren. Quantum Key Distribution liefert solche Schlüsseleinigungsprotokolle. Mittlerweile wurde eine Vielzahl verschiedener QKD-Protokolle mit unterschiedlichen praktischen Anforderungen entwickelt, denen aber eine ähnliche Idee zugrunde liegt, siehe beispielsweise [2], [16].

Bei QKD-Protokollen sind Alice und Bob durch einen Quantenkanal verbunden, über den sie Quantenzustände senden können. Praktisch handelt es sich dabei meist um einen optischen Kanal, über den Photonen gesendet werden, deren Polarisation sich wie ein Quantenzustand verhält. Quantenzustände haben die aus klassischer Sicht seltsame Eigenschaft, dass sich das Ergebnis einer Messung eines Quantenzustandes in der Regel nicht deterministisch vorhersagen lässt, sondern jedem Ergebnis nur eine bestimmte Wahrscheinlichkeit zugeordnet ist. Ferner wird durch die Messung eines Quantensystems der Zustand des Systems selbst beeinflusst. Letzteres macht sich QKD für die Vereinbarung geheimer Schlüssel zunutze: Wenn Alice einen Quantenzustand zu Bob schickt, so ändert sich durch Abhören der gesendete Quantenzustand, so dass die Anwesenheit eines Abhörenden von Alice und Bob statistisch detektiert werden kann. Um einen einfachen Man-in-the-Middle-Angriff zu verhindern, müssen Alice und Bob jedoch neben dem Quantenkanal auch über

einen klassischen authentifizierten Kanal verbunden sein, also zu Beginn des Protokolls schon über einen gemeinsamen geheimen Schlüssel verfügen. Unter dieser Voraussetzung werden in QKD-Protokollen durch Versenden von Quantenzuständen weitere Schlüssel vereinbart.

3.2. Chancen und Beschränkungen von QKD

Auch wenn viele mathematische Probleme, die der Sicherheit der Post-Quanten-Kryptografie zugrunde liegen, gut untersucht sind, ist es nicht ausgeschlossen, dass diese Verfahren durch algorithmische Fortschritte zukünftig gebrochen werden können. Die Quantenkryptografie stellt sich als mögliche Alternative dar. Als Vorteil von QKD wird häufig angeführt, dass die zugrundeliegenden Protokolle informationstheoretische Sicherheit bieten. Damit beruhen sie nicht auf der Annahme, dass bestimmte mathematische Probleme nicht effizient gelöst werden können. Jedoch bringt QKD selbst einige Einschränkungen und noch zu klärende grundlegende Forschungsfragen mit.

Sicherheitskriterien und -beweise. Als ein anerkanntes Sicherheitskriterium für QKD wird heute die sogenannte ε-Sicherheit [36] gesehen. Sie hat damit ein auf "accessible information" basierendes Kriterium abgelöst, das sich als ungeeignet herausgestellt hat [27]. Beim ε-Kriterium stellt sich jedoch die Frage nach der operationellen Bedeutung, um die theoretischen Sicherheitseigenschaften des vereinbarten Schlüssels zu verstehen und eine geeignete Größe für den Sicherheitsparameter festzulegen. Ferner ist es notwendig, Sicherheitsbeweise basierend auf diesem Kriterium auch für praktisch relevante QKD-Protokolle unter realen Bedingungen wie endlichen Schlüssellängen und unter Berücksichtigung des allgemeinsten Angriffsmodells zu entwickeln.

Ende-zu-Ende-Sicherheit. Die Reichweite von optischen Leitungen liegt bei etwa 100km, da die Signalverluste in Abhängigkeit von der Distanz exponentiell steigen. Da allgemeine Quantenzustände nicht wie klassische Signale exakt kopiert werden können, können herkömmliche Signalverstärker nicht verwendet werden. Dadurch kann QKD zurzeit über diese Distanz hinaus keine Ende-zu-Ende-Sicherheit liefern. Eine mögliche Lösung sind Quantenrepeater, deren praktisch relevante Entwicklung aber in den nächsten Jahren nicht erwartet werden kann. Eine Alternative dazu ist die Nutzung von Freistrahllasern unter Einbeziehung von Satelliten, durch die größere Reichweiten erreicht werden können.

Informationstheoretische Sicherheit versus Computational Security. Sobald ein Schlüssel durch ein QKD-Protokoll vereinbart wurde, muss festgelegt werden, wie er genutzt werden darf. Häufig wird hier die Verschlüsselung mit One-Time-Pad (im Folgenden OTP) als informationstheoretisch sicheres Verfahren angeführt. Eine Alternative ist die Verschlüsselung mit einem etablierten Verfahren wie dem AES, das computational security liefert. Aus Sicht des BSI ist die Nutzung des OTP alleine zur Verschlüsselung nicht geeignet, denkbar ist allerdings die OTP-Verschlüsselung von vorher AES-verschlüsselten Daten. Das OTP bietet keinen Integritätsschutz und selbst kleine Schiefen in den per QKD erzeugten Schlüsseln könnten für Angriffe genutzt werden. Darüber hinaus ist es heute noch nicht möglich,

Kanäle mit hohen Datenraten schnell mit QKD-Schlüsseln zur Nutzung des OTP zu versorgen.

Authentisierung. Neben dem Quantenkanal wird bei jedem QKD-Protokoll eine klassische Authentisierung benötigt. Die dabei verwendeten Schlüssel müssen vorab verteilt werden und werden nach jeder Schlüsseleinigung mit Teilen des neu vereinbarten Geheimnisses verknüpft. Bei informationstheoretisch sicheren Ansätzen wird dabei vor allem die sogenannte Wegman-Carter-Authentisierung [41] verwendet. Jedoch verschlechtert sich mit jeder Schlüsseleinigung der Sicherheitsparameter ε. Nach einiger Zeit muss der Schlüssel also extern erneuert werden. Dies verkompliziert den Betrieb von QKD-Netzen und ist ein altbekanntes Problem, siehe beispielsweise [3].

Netzwerkaspekte. Es muss noch geklärt werden, wie sich Quantenkommunikation in bestehende Infrastrukturen einfügen lässt. Das BSI hat dazu zusammen mit dem Projekt Q.Link.X einen ersten Workshop organisiert [42]. Es bleiben aber weiterhin viele Fragen, etwa zum Schlüsselmanagement, offen. Die Vorstellung, dass Quantenrepeater später problemlos Trusted-Nodes ersetzen können, ist ebenfalls nach heutiger Kenntnis falsch. Quantenrepeater werden vermutlich nur mit Prepare-and-Measure QKD-Devices funktionieren, die Photonen verwenden – und auch dann vielleicht nicht für alle QKD-Protokolle. Weiterhin ist nicht klar, dass Geräte verschiedener Hersteller nahtlos zusammenarbeiten können werden. Dies birgt die Gefahr der Herstellerabhängigkeit.

Seitenkanäle. Auch wenn man annimmt, dass QKD theoretisch sicher ist, wurde mittlerweile eine Vielzahl von Seitenkanalangriffen gefunden [17]. Diese müssen bei praktischen Implementierungen von QKD berücksichtigt werden. Um die Sicherheit praktischer QKD-Produkte zu gewährleisten, ist es wichtig, die Forschung zu möglichen Seitenkanalangriffen zu intensivieren und weiter voranzutreiben.

Quantenzufallszahlen. Ein wesentlicher Bestandteil von QKD-Protokollen ist, dass Zufallszahlen mit hoher Güte zur Verfügung stehen müssen. Oft wird dabei die Verwendung von Quantenzufallszahlengeneratoren (QRNGs) vorgeschlagen. Das BSI hat in Zusammenarbeit mit dem Fraunhofer IOF zwei Workshops zur Bewertung von QRNGs veranstaltet. Nach Auffassung des BSI handelt es sich bei QRNGs um einen speziellen Typ von physikalischen Zufallszahlengeneratoren, der nicht unbedingt herkömmlichen physikalischen Generatoren überlegen ist. Sicherlich falsch sind Aussagen der Art „QRNGs liefern Zufallszahlen auf Basis von Naturgesetzen und sind daher sicher“. Das BSI beabsichtigt in seiner Methodologie zur Bewertung von Zufallszahlengeneratoren (AIS 20/31) eine Einordnung in die vorhandenen Funktionalitätsklassen vorzunehmen.

Digitale Souveränität. Um vertrauenswürdige deutsche oder europäische Netze aufzubauen, bedarf es vertrauenswürdiger Komponenten. Dies gilt insbesondere dann, wenn Regierungsnetze geschützt werden sollen. Eine zu frühe Einführung von QKD kann dazu führen, dass Produkte bei Herstellern außerhalb der EU beschafft werden und diese Hersteller einen Vorsprung erlangen: Projekte mit dem Ziel, digitale Souveränität im Bereich

der Quantentechnologien herzustellen, könnten also dazu führen, dass sich im Sicherheitsbereich keine europäischen Marktführer entwickeln. Optimal wäre ein deutscher Hersteller, der QKD-Geräte für nationale Hochsicherheitsanwendungen anbietet.

3.3. Aktuelle Forschung und Entwicklung zur Quantenkommunikation

Über die letzten 20 Jahre wurden die Bemühungen um Forschung und Entwicklung im Bereich QKD intensiviert. Insbesondere China konnte früh Erfolge demonstrieren. In China und vielen anderen Ländern entstehen Teststrecken zur QKD. Auch in Europa entstanden in den letzten Jahren große Projekten im Bereich der Quantentechnologien und speziell der Quantenkommunikation. Eine Auswahl dieser Projekte wollen wir kurz vorstellen**.**

3.3.1. Europäische Projekte

Als Absichtserklärung zur Förderung der Quantenkommunikation innerhalb der EU unterschrieben 2019 einige EU-Mitgliedsstaaten eine Quantum Declaration [33] mit folgendem Ziel:

„*Plan to work together to establish a cooperation framework – EuroQCI – for exploring within the next 12 months, the possibility of developing and deploying in the Union, within the next 10 years, a certified secure end-to-end quantum communication infrastructure (QCI) composed of space-based and terrestrial-based solutions, enabling information and data to be transmitted and stored ultra-securely and capable of linking critical public communication assets all over the Union.“*

Mittlerweile wurde die Declaration von 25 Mitgliedsstaaten unterzeichnet (Stand: Dezember 2020). Erste Ausschreibungen für Studien zur Gestaltung eines europäischen Quantennetzwerkes laufen gerade.

Daneben hat die EU das Quantum Technologies Flagship [35] zur Förderung von Quantentechnologien aufgelegt. Es wurde 2018 begonnen und soll über zehn Jahre mit einem Gesamtvolumen von einer Milliarde Euro laufen. Es umfasst europäische Forschungsaktivitäten in allen Bereichen der Quantentechnologien wie Quantensimulationen, Quantencomputing, Quantensensoren und Quantenkommunikation.

3.3.2. Deutsche Projekte

In Deutschland werden momentan auf verschiedenen Ebenen Quantenprojekte gefördert und Innovationspotenziale untersucht [1]. Wir berichten nur über zwei dieser Projekte, an denen das BSI beratend beteiligt ist.

Bei QuNET [34] handelt es sich um eine vom BMBF geförderte Initiative zur Erforschung von Quantennetzwerken im Umfang von 165 Millionen Euro, wobei verschiedene Technologien untersucht werden. Im Rahmen des Teilprojektes QuNET-alpha sollte am 1.12.2020 eine verschlüsselte Videoverbindung zwischen BMBF und BSI in Bonn hergestellt werden. Pandemiebedingt wurde diese Demonstration auf 2021 verschoben. Sie soll

hybrid gestaltet sein, d.h. die Schlüsseleinigung verwendet einen Post-Quanten-Algorithmus und QKD, um zwei Schlüssel zu vereinbaren, aus denen dann ein gemeinsamer Schlüssel abgeleitet wird. Die Bundesregierung hat in der Antwort auf eine kleine Anfrage [12] ausführlich zu QuNET Stellung genommen. QuNET bildet aus deutscher Sicht in gewisser Weise den Kern des deutschen Anteils an EuroQCI.

Wie bereits beschrieben werden Quantenrepeater [29] benötigt, um bei fasergebundener QKD Ende-zu-Ende-Sicherheit über größere Distanzen zu erreichen. Aus dem No-Cloning-Theorem der Quantenmechanik folgt, dass ein solcher Repeater nicht wie ein herkömmlicher Verstärker arbeiten kann. Vielmehr sind völlig neue Technologien zu entwickeln. Das vom BMBF geförderte Projekt Q.Link.X (für Quantum-Link-Extension) erforscht solche Repeater. Um größere Reichweiten, wie sie für ein landesweites sicheres Netz ohne Verwendung von Trusted Nodes benötigt werden, zu erreichen, sind Repeater erforderlich. Mit Repeatern im Sinne der Quantenkommunikation sind komplexere Protokolle und Quantenspeicher verbunden, die eine Reichweitensteigerung ermöglichen sollen. Eine Marktreife ist kurzfristig nicht zu erwarten. Die zu entwickelnden Komponenten können potenziell allerdings auch in nicht-kryptografischen Anwendungen genutzt werden, beispielsweise bei der optischen Koppelung von Quantencomputern.

Angesichts der zunehmenden Bedeutung der Quantentechnologien in Deutschland hat sich der "Deutsche Industrieverbund für Quantensicherheit" [15] gegründet.

3.3.3. Aktivitäten des BSI

Wegen der Dringlichkeit durch die Bedrohung durch Quantencomputer und der großen Aufmerksamkeit und Förderung, die QKD zurzeit erfährt, ist es wichtig, die Technologie auf ihre Sicherheit zu untersuchen. Dies gilt sowohl in theoretischer Hinsicht als auch für praktische Systeme. Deshalb engagiert sich das BSI und begleitet die Entwicklungen durch eigene Aktivitäten.

QKD ist angesichts der noch hohen Kosten eher als Lösung für Anwendungen mit hohen Sicherheitsanforderungen zu sehen. Daher ist eine Evaluierung nach einer international anerkannten Evaluationsmethodologie wie den Common Criteria (CC) [14] erforderlich. Eine Zulassung würde noch weitere Untersuchungen erfordern. Wegen der hohen Anforderungen wird für eine CC-Evaluierung mindestens EAL 4+ (mit Schwachstellenanalyse AVA_VAN.5 und Lifecycle-Aspekten aus ALC) erforderlich sein. Das BSI hat die Telekom Security mit der Erstellung eines Protection Profile (PP) beauftragt. Die Erstellung der Prüfkriterien erfolgt in Zusammenarbeit mit ETSI. Das gewählte EAL stellt dabei eine Mindestanforderung dar. Aus Sicht des BSI wird QKD (auch wegen der damit verbundenen Kosten) zunächst in Bereichen mit hohen Sicherheitsanforderungen eingesetzt werden. Damit ist es unabdingbar, Seitenkanalanalysen und Quellcodeuntersuchungen durchzuführen. Ebenso spielt der Lebenszyklus der Geräte eine große Rolle. Eine erste Version des PP wird momentan im Rahmen der ETSI diskutiert.

Wie bereits erwähnt ist es nicht nur erforderlich, die Implementierungssicherheit eines QKD-Gerätes zu bewerten, auch die theoretische Sicherheit spielt eine wesentliche Rolle. Das BSI plant daher eine Studie zur theoretischen Sicherheit von QKD, die aber auch praktische Aspekte berücksichtigen soll. Auf längere Sicht plant das BSI, Angaben zu QKD in seine Technischen Richtlinien aufzunehmen.

3.4. Zusammenfassung und Einschätzung

Durch wissenschaftliche Fortschritte und Warnungen der NSA aus dem Jahr 2015 hat das Thema Quantencomputerresistenz an Aktualität gewonnen (siehe auch [32]). Die Auswirkungen und zu ergreifende Maßnahmen wurden vom BSI bereits in zahlreichen Beiträgen (etwa [8], [9], [11]) thematisiert. Betroffen sind viele international gängige kryptografische Verfahren. Zurzeit arbeiten große Firmen wie IBM und Google mit Hochdruck an der Entwicklung von Quantencomputern und haben sich große Fortschritte in den nächsten zehn Jahren zum Ziel gesetzt [26]. Deshalb gilt es, Anwendungen und Systeme, die aufgrund der verwendeten kryptografischen Verfahren anfällig für Quantencomputerangriffe sein könnten, zu identifizieren und anzupassen. Insbesondere bei langfristigem Schutzbedarf müssen das Risikopotential und die Kryptoagilität genauer untersucht werden.

Neben der Post-Quanten-Kryptografie gibt es als auf unabhängigen Prinzipien basierende Alternative die Quantenkommunikation, die derzeit in vielen Projekten untersucht wird. Quantenkommunikation bietet eine interessante Ergänzung zur traditionellen Kryptografie, die auf einem anderen Prinzip beruht. Beim Einsatz von QKD sind jedoch Einschränkungen und offene Fragen zu beachten, die teilweise hier behandelt wurden. So wird beispielsweise für QKD im Gegensatz zu Post-Quanten-Verfahren hochspezialisierte Hardware benötigt, die gegen zahlreiche bekannte Seitenkanalangriffe resistent sein muss, fasergebundene QKD bietet zurzeit keine Ende-zu-Ende-Sicherheit über größere Distanzen, und es werden vorverteilte und wieder aufzufrischende Authentisierungsschlüssel benötigt. Eine alleinige Verwendung von QKD zur Schlüsseleinigung kommt heute unter anderem wegen der beschriebenen Einschränkungen und theoretischen Grundsatzfragen noch nicht in Frage. Hybride Ansätze in Kombination mit klassischen und Post-Quanten-Verfahren sind denkbar.

Unabhängig von den dargestellten Einschränkungen von QKD sind die Forschungsaktivitäten in diesem Bereich begrüßenswert, da sich auch andere mögliche Anwendungen der Quantenkommunikation ergeben, die nicht im Bereich der Kryptografie liegen. Beispielsweise wird die Koppelung von Quantencomputern unter dem Stichwort Distributed Quantum Computing diskutiert, zu der insbesondere Repeatertechnologien einen Beitrag leisten können.

Literaturhinweise

[1] Innovationspotenziale der Quantentechnologien der zweiten Generation. acatech IMPULS (2020), https://www.acatech.de/publikation/innovationspotenziale-der-quantentechnologien/

[2] C. H. Bennett, G. Brassard: Quantum cryptography: Public key distribution and coin tossing, Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, volume 175, page 8 (1984)

[3] M. Ben-Or, M. Horodecki, D. W. Leung, D. Mayers, J. Oppenheim: The Universal Composable Security of Quantum Key Distribution, arXiv:quant-ph/0409078v1 (2004)

[4] Bundesministerium des Innern, für Bau und Heimat, Allgemeine Verwaltungsvorschrift zum materiellen Geheimschutz, Verschlusssachenanweisung – VSA (2018)

[5] Bundesministerium für Bildung und Forschung: Post-Quanten-Kryptografie (2019), https://www.forschung-it-sicherheit-kommunikationssysteme.de/foerderung/bekanntmachungen/pqk

[6] Randombit: Botan; Crypto and TLS for Modern C++, https://botan.randombit.net/, Letzter Zugriff 02. Oktober 2020

[7] Bundesamt für Sicherheit in der Informationstechnik: Sichere Implementierung einer allgemeinen Kryptobibliothek (2017)

[8] H. Hagemeier, S. Kousidis, M. Lochter, S. Maßberg: Quantencomputer als Herausforderung für die Informationssicherheit. In 15. Deutscher IT- Sicherheitskongress (Mai 2017)

[9] H. Hagemeier, S. Kousidis, M. Lochter: Informationssicherheit im Quantenzeitalter - ein Update. In 16. Deutscher IT-Sicherheitskongress (Mai 2019)

[10] Bundesamt für Sicherheit in der Informationstechnik: Kryptographische Verfahren: Empfehlungen und Schlüssellängen, Technische Richtlinie BSI TR-02102-1, Bonn (2020)

[11] Bundesamt für Sicherheit in der Informationstechnik: Migration zu Post-Quanten-Kryptografie. Tech. Rep., Bonn (2020), https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Krypto/Post-Quanten-Kryptografie.html

[12] Antwort auf die kleine Anfrage: Hochsicheres Quantennetzwerk QuNET (2020), https://dip21.bundestag.de/dip21/btd/19/183/1918355.pdf

[13] M. Campagna, E. Crockett: Hybrid Post-Quantum Key Encapsulation Methods (PQ KEM) for Transport Layer Security 1.2 (TLS), Internet-Draft draft-campagna-tls-bike-sike-hybrid-01, Internet Engineering Task Force (May 2019), https://datatracker.ietf.org/doc/html/draft-campagna-tls-bike-sike-hybrid-01, in Arbeit

[14] Common Criteria for Information Technology Security Evaluation, www.commoncriteriaportal.org

[15] Deutscher Industrieverbund für Quantensicherheit, www.divqsec.de

[16] A. Ekert: Quantum cryptography based on Bell's theorem, Physical Review Letters. 67 (6), pp. 661-663 (1991).

[17] ETSI White Paper No. 27: Implementation Security of Quantum Cryptography, (2018)

[18] T. Fritzmann, T. Schamberger, C. Frisch, K. Braun, G. Maringer, J. Sepúlveda: Efficient Hardware/Software Co-design for NTRU. In IFIP/IEEE International Conference on Very Large Scale Integration-System on a Chip (pp. 257-280). Springer, Cham (2018)

[19] T. Fritzmann, U. Sharif, D. Müller-Gritschneder, C. Reinbrecht, U. Schlichtmann, J. Sepulveda: Towards Reliable and Secure Post-Quantum Co-Processors based on RISC-V, In 2019 Design, Automation & Test in Europe Conference & Exhibition (DATE) (pp. 1148-1153). IEEE (2019)

[20] T. Fritzmann, G. Sigl, J. Sepúlveda: Extending the RISC-V instruction set for hardware acceleration of the post-quantum scheme LAC, In 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE) (pp. 1420-1425). IEEE (2020)

[21] T. Fritzmann, G. Sigl, J. Sepúlveda: RISQ-V: Tightly Coupled RISC-V Accelerators for Post-Quantum Cryptography, IACR Transactions on Cryptographic Hardware and Embedded Systems 2020(4): 239-280 (2020)

[22] S.L. Gazdag, M. Friedl, D. Loebenberger: Post-Quantum Software Updates: A Case Study on Code Signing with Hash-Based Signatures, In: K. David, K. Geihs, M. Lange, G. Stumme (eds.) INFORMATIK 2019: Konferenzbeiträge der 49. Jahrestagung der Gesellschaft für Informatik. vol. P-294, pp. 459–472. Köllen Druck+Verlag GmbH, Bonn (2019), https://doi.org/10.18420/inf2019_63

[23] C. Gidney, M. Ekerå: How to Factor 2048 bit RSA Integers in 8 Hours Using 20 Million Noisy Qubits. arXiv preprint arXiv:1905.09749 (2019)

[24] L.K. Grover: A Fast Quantum Mechanical Algorithm for Database Search, In: Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, p. 212–219, STOC '96, Association for Computing Machinery, New York, NY, USA (1996), https://doi.org/10.1145/237814.237866

[25] A. Huelsing, D. Butin, S. Gazdag, J. Rijneveld, A. Mohaisen: XMSS: eXtended Merkle Signature Scheme, RFC 8391 (Informational) (May 2018), https://doi.org/10.17487/RFC8391, https://www.rfc-editor.org/rfc/rfc8391.txt

[26] IBM's Roadmap For Scaling Quantum Technology (2020), https://www.ibm.com/blogs/research/2020/09/ibm-quantum-roadmap/

[27] R. König, R. Renner, A. Bariska, U. Maurer: Locking of accessible information and implications for the security of quantum cryptography, Phys. Rev. Lett. 98, 140502 (2007)

[28] M. Lochter: Digitale Souveränität trotz Quantencomputer, OMNISECURE - the world of smart ID solutions, Berlin (2020)

[29] P. v. Loock, W. Alt, C. Becher, O. Benson, H. Boche, C. Deppe, J. Eschner, S. Höfling, D. Meschede, P. Michler, F. Schmidt, H. Weinfurter: Extending Quantum Links: Modules for Fiber- and Memory-Based Quantum Repeaters, Advanced Quantum Technologies, 1900141 (2020)

[30] D. McGrew, M. Curcio, S. Fluhrer: Leighton-Micali Hash-Based Signatures. RFC8554 (Informational) (2019), https://doi.org/10.17487/RFC8554, https://www.rfc-editor.org/rfc/rfc8554.txt

[31] NIST: Post-Quantum Cryptography round 3 submissions (2020), https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Round-3-Submissions

[32] National Security Agency: Information Assurance Directorate MFQ U/OO/815099-15 (2016)

[33] The future is quantum: EU countries plan ultra-secure communication network (2019), https://ec.europa.eu/digital-single-market/en/news/future-quantum-eu-countries-plan-ultra-secure-communication-network

[34] QuNET-alpha, https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/qunet-alpha

[35] Quantum Technologies Flagship, https://qt.eu/

[36] R. Renner, R. König: Universally composable privacy amplication against quantum adversaries, In Theory of Cryptography, Proceedings of TCC 2005 , LNCS, Vol. 3378 (2005)

[37] P.W. Shor: Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer, SIAM Review 41(2), 303–332 (1999)

[38] V. Smyslov: Intermediate Exchange in the IKEv2 Protocol, Internet-Draft draft-ietf-ipsecme-ikev2-intermediate-04, Internet Engineering Task Force (2020), https://datatracker.ietf.org/doc/html/draft-ietf-ipsecme-ikev2-intermediate-04

[39] D. Steblia, S. Fluhrer, S. Gueron: Hybrid key exchange in TLS 1.3. Internet-Draft draft-ietf-tls-hybrid-design-00, Internet Engineering Task Force (Apr 2020), https://datatracker.ietf.org/doc/html/draft-ietf-tls-hybrid-design-00, in Arbeit

[40] C. Tjhai, M. Tomlinson, grbartle@cisco.com, S. Fluhrer, D.V. Geest, O. Garcia-Morchon, V. Smyslov: Multiple Key Exchanges in IKEv2, Internet-Draft draft-ietf-ipsecme-ikev2-multiple-ke-00, Internet Engineering Task Force (Jan 2020), https://datatracker.ietf.org/doc/html/draft-ietf-ipsecme-ikev2-multiple-ke-00, in Arbeit

[41] M. Wegman, J. Carter: New hash functions and their use in authentication and set equality, Journal of Computer and System Sciences, Volume 22, Issue 3 (1981)

[42] Physikalische und Industrielle Anforderungen an zukünftige Quantennetzwerke, https://photonicnet.de/veranstaltungen/veranstaltung/physikalische-und-industrielle-anforderungen-an-zukuenftige-quantennetzwerke-766/

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Standardisierung von Post-Quanten-Kryptografie und Empfehlungen des BSI

Dr. Heike Hagemeier[1], Dr. Stavros Kousidis[1], Dr. Thomas Wunderer[1]

Kurzfassung:
In diesem Beitrag geben wir einen Überblick über den aktuellen Stand der Standardisierung von Post-Quanten-Kryptografie mit Hinblick auf die Empfehlungen des BSI. Wir fassen den aktuellen Stand des NIST Post-Quanten Standardisierungsprozesses zusammen und gehen dabei näher auf die Verfahren ein, die derzeit vom BSI empfohlen werden. Anschließend beleuchten wir die aktuellen Vorschläge, wie kryptografische Protokolle wie IKE oder TLS angepasst werden können, um eine hybride Schlüsseleinigung zu ermöglichen, und gehen dann auf Authentisierung und digitale Zertifikate ein. Auch dafür werden bereits verschiedene Ansätze für die Migration zu Post-Quanten-Kryptografie diskutiert.

Stichworte: hybride Schlüsseleinigung, NIST-Prozess, Post-Quanten-Kryptografie, Public-Key-Infrastrukturen, Schlüsseleinigungsverfahren, Standardisierung, TR-02102-1

1. Einleitung

Post-Quanten-Kryptografie (oder auch quantencomputerresistente Public-Key-Kryptografie) gibt es im Grunde genommen schon seit den 1970er Jahren, nämlich spätestens seit der Entwicklung des McEliece Verschlüsselungsverfahren [1] sowie der Merkle-Signaturen [2]. Nur existierte damals weder der Begriff noch die Notwendigkeit für Post-Quanten-Kryptografie. Dies änderte sich mit der Entwicklung von Quantenalgorithmen [3], die die für eine Vielzahl von asymmetrischen Verfahren (Public-Key-Kryptografie) grundlegenden mathematischen Probleme (Faktorisierung und Diskreter Logarithmus Problem) effizient lösen können. Natürlich unter der Voraussetzung, dass ein Quantencomputer existiert, auf dem diese Algorithmen implementiert werden können. Eine aktuelle Studie des BSI zum Entwicklungsstand Quantencomputer [4] zeigt, dass aktuell eine enorme Anstrengung nötig wäre, um eine kryptografisch relevante Skalierung vorzunehmen. Gleichzeitig aber wird deutlich, dass die Entwicklung durch starke Industrieakteure und große Forschungsprogramme an Fahrt gewonnen hat, und somit der Bedarf nach einer quantencomputerresistenten Public-Key-Kryptografie immer dringender wird.

In den letzten Jahren hat die Post-Quanten-Kryptografie daher erheblich an Bedeutung gewonnen. Bereits im August 2015 hat die US-amerikanische National Security Agency (NSA) vor Quantencomputern gewarnt und die Migration zu quantencomputerresistenten Verfahren eingeleitet. Als Begründung hat die NSA Fortschritte in Physik und Technologie angegeben, welche die Entwicklung eines kryptografisch relevanten Quantencomputers ermöglichen könnten. Konkrete quantencomputerresistente Verfahren hat die NSA dabei nicht benannt, sondern auf künftige Standards des National Institute of Standards and Technology (NIST) verwiesen. In Folge dessen arbeiten NIST sowie andere

[1] Bundesamt für Sicherheit in der Informationstechnik, Bonn

Standardisierungsgremien intensiv an der Standardisierung von Post-Quanten-Kryptografie.

Mit der Standardisierung neuer Algorithmen ist es allerdings nicht getan. Einerseits passen die Algorithmen nicht ohne weiteres in das Gerüst bestehender kryptografischer Protokolle wie beispielsweise das Transport Layer Security (TLS) Protokoll. Andererseits wird den neuen Algorithmen allgemein noch nicht das gleiche Vertrauen entgegen gebracht wie den klassischen Verfahren, da sie bezüglich einiger Aspekte (beispielsweise der Möglichkeit von Seitenkanalangriffen) noch nicht intensiv genug untersucht sind und die Erfahrung bei der Implementierung dieser Verfahren ausgebaut werden muss. Nach Ansicht vieler Experten sollte Post-Quanten-Kryptografie - zumindest in einer Übergangszeit - nicht alleine eingesetzt werden, sondern nur "hybrid", d.h. in Kombination mit einem klassischen Verfahren. Dafür müssen die kryptografischen Protokolle entsprechend geändert bzw. ergänzt werden. Zusätzlich müssen beispielsweise auch Public-Key-Infrastrukturen angepasst werden. Auch hier stellt sich die Frage, ob eine Signatur mit einem quantencomputerresistenten Verfahren ausreichend ist oder ob Zertifikate benötigt werden, die mehrere Signaturverfahren umfassen.

Diese und andere Migrationsprobleme werden unabhängig von der Auswahl konkreter Algorithmen von Standardisierungsgremien wie IETF oder ITU-T angegangen. Grundsätzlich können Migrationsprobleme abgemildert werden, wenn von Beginn an auf ein flexibles ("kryptoagiles") Design geachtet wird. Das Ziel der "Kryptoagilität" sollte daher bei allen Anpassungen ein wichtiges Kriterium sein. Diese und andere Empfehlungen zur "Migration auf Post-Quanten-Kryptografie" hat das BSI im April 2020 veröffentlicht [5].

In diesem Beitrag beschreiben wir zunächst den Stand der Standardisierung von Post-Quanten-Kryptografie. Danach gehen wir auf die Aktivitäten zur Anpassung kryptografischer Protokolle bezüglich einer hybriden Schlüsseleinigung ein. (Hybride) Authentisierung und mögliche Lösungen für "Post-Quantum-Public-Key-Infrastrukturen" diskutieren wir im vierten Kapitel. Der Fokus liegt dabei immer auf den Empfehlungen des BSI, die wir abschließend zusammenfassen.

2. Aktuelle Entwicklungen in der Standardisierung von Post-Quanten-Kryptografie

Wir berichten vom aktuellen Stand des NIST-Prozesses zur Standardisierung von Post-Quanten-Kryptografie [6] und gehen anschließend auf Verfahren zur Schlüsseleinigung und digitale Signaturverfahren ein, die bereits in die Technische Richtlinie des BSI TR-02102-1 [7] aufgenommen wurden.

2.1. Der NIST Post-Quanten Standardisierungsprozess

Das NIST ist als US-amerikanische Behörde für Standardisierungsprozesse zuständig. Es hat unter anderem Wettbewerbe durchgeführt, welche die weltweit anerkannten Algorithmen AES und SHA-3 hervorgebracht haben. Als Reaktion auf die Ankündigung der NSA hat NIST im November 2016 einen Prozess gestartet, an dessen Ende eine Auswahl von quantencomputerresistenten kryptografischen Verfahren zur Verfügung

stehen soll [6]. Dieser Prozess wird in mehreren Runden durchgeführt. Bis zur Einreichungsfrist im November 2017 wurden insgesamt 82 Verfahren zum Prozess eingereicht, von welchen 69 die Mindestkriterien erfüllten und von NIST als Kandidaten in die erste Runde des Prozesses aufgenommen wurden. Im Januar 2019 hat NIST, basierend auf den Kommentaren der öffentlichen Forschungsgemeinschaft und der NIST-internen Analyse, 26 dieser Kandidaten für die zweite Runde des Prozesses ausgewählt. Im Juli 2020 hat NIST dann die Verfahren bekannt gegeben, welche in die dritte Runde übernommen werden. NIST hat die Kandidaten der dritten Runde in "Finalisten" und "Alternativen" eingeteilt. Voraussichtlich wird am Ende der dritten Runde eine Auswahl der Finalisten standardisiert. Möglicherweise wird es im Anschluss daran noch eine vierte Runde geben, an deren Ende eine Auswahl der alternativen Kandidaten standardisiert werden kann. Die Finalisten der dritten Runde sind die vier asymmetrischen Verschlüsselungs- bzw. Schlüsseleinigungsverfahren Classic McEliece [8], CRYSTALS-KYBER [9], NTRU [10] und SABER [11] sowie die drei Signaturverfahren CRYSTALS-DILITHIUM [12], FALCON [13] und Rainbow [14]. Bei den acht alternativen Verfahren handelt es sich um BIKE [15], FrodoKEM [16], HQC [17], NTRU Prime [18], SIKE [19], GeMSS [20], Picnic [21] und SPHINCS+ [22].

2.2. Schlüsseleinigung

Im Rahmen einer Schlüsseleinigung verständigen sich zwei Parteien auf einen gemeinsamen kryptografischen Schlüssel. Dieser Schlüssel wird in der Regel anschließend für eine symmetrisch verschlüsselte Kommunikation verwendet. Das Schlüsseleinigungsverfahren, mit dem dieser Schlüssel vereinbart wird, ist jedoch asymmetrisch.

Schlüsseleinigung ist ein wesentliches Element zum Schutz der Vertraulichkeit von Daten. Gerade bei Daten, die über einen langen Zeitraum vertraulich gehalten werden müssen, ist ein zeitnaher Einsatz quantencomputerresistenter Verfahren zu empfehlen. Wird nämlich beispielsweise heute ein Schlüsselaustausch, der nicht quantencomputerresistent ist, von einem Angreifer oder einer Angreiferin beobachtet, so kann diese(r) möglicherweise in Zukunft, wenn kryptografisch relevante Quantencomputer verfügbar sind, den gemeinsamen Schlüssel berechnen und die damit verschlüsselten Daten entschlüsseln und lesen. Dieses Szenario ist auch unter dem Namen „store now, decrypt later“ („jetzt speichern, später entschlüsseln“) bekannt. Der Einsatz eines quantencomputerresistenten Schlüsseleinigungsverfahrens ist somit also auch schon ratsam, bevor kryptografisch relevante Quantencomputer existieren. Ab wann quantencomputerresistente Schlüsseleinigungsverfahren zum Schutz vertraulicher Daten eingesetzt werden sollten, ist eine Frage der Zeitspanne, über welche die Daten vertraulich gehalten werden müssen, und des Risikomanagements.

2.2.1. Classic McEliece

Das McEliece-Kryptosystem ist ein aus der Kodierungstheorie entstandenes asymmetrisches Verschlüsselungsverfahren und wurde 1978 von Robert McEliece vorgestellt [1]. Dessen Sicherheit basiert auf den Annahmen, dass die verwendeten binären Goppa Codes nicht unterscheidbar von zufälligen linearen Codes sind und die letztgenannten

sowohl auf Digitalrechnern als auch mit Hilfe von Quantencomputern nur mit exponentiellem Aufwand dekodiert werden können. Abgesehen von einer Anpassung der ursprünglich von McEliece vorgeschlagenen Parameter (diese konnten vor dem Hintergrund moderner Rechenleistung erst in 2008 von Bernstein, Lange und Peters in ca. 260 Operationen angegriffen werden [23]) ist es nach über 40 Jahren Forschung nicht gelungen, eine strukturelle Schwäche des McEliece-Kryptosystems bei Verwendung von binären Goppa Codes zu finden. Somit gilt das McEliece-Kryptosystem als eines der ältesten ungebrochenen quantencomputerresistenten Verfahren überhaupt.

Ein großer Nachteil besteht im Platzbedarf des öffentlichen Schlüssels (im Megabyte-Bereich für Hochsicherheitsanwendungen). Andererseits jedoch sind die Chiffrate sehr klein (ca. 200 Bytes) und die Ver- und Entschlüsselung ist wesentlich effizienter als RSA- oder EC-basierte asymmetrische Verschlüsselungsverfahren. Neuere Ansätze haben mehr Struktur in die verwendete Code-Klasse eingeführt, um den Platzbedarf des öffentlichen Schlüssels deutlich zu verringern, z. B. [24]. Diese zusätzlichen Strukturen haben jedoch zu erfolgreichen Angriffen auf manche Ansätze geführt [25].

Im NIST-Prozess gehört die Einreichung Classic McEliece [ABC+20] aktuell zu den Finalisten in Runde 3. Classic McEliece beschreibt einen Key Encapsulation Mechanism (KEM) auf Basis des zu McEliece äquivalenten codebasierten Niederreiter-Kryptosystem von 1986 [26]. Das BSI hat Classic McEliece mit entsprechenden Sicherheitsparametern in die Empfehlungen der TR-02102-1 [7] aufgenommen.

2.2.2. FrodoKEM

FrodoKEM ist ein gitterbasiertes Schlüsseleinigungsverfahren, dessen Sicherheit auf der Annahme beruht, dass das sogenannte Learning With Errors (LWE) Problem für klassische Computer und Quantencomputer schwer zu lösen ist. Das LWE Problem wurde 2005 von Oded Regev eingeführt [27] und dient seitdem als Grundlage vieler gitterbasierter kryptografischer Verfahren. Trotz intensiver kryptoanalytischer Forschung konnte das LWE Problem bisher weder mit klassischen, noch mit Quantencomputern effizient gelöst werden und man kann zeigen, dass typische LWE Instanzen mindestens so schwer zu lösen sind wie die schwierigsten Instanzen bestimmter Gitter-Probleme.

Im Gegensatz zu vielen anderen gitterbasierten Verfahren im NIST-Prozess haben die FrodoKEM zugrundeliegenden Gitter keine zusätzliche algebraische Struktur. Auch wenn nicht bekannt ist, ob solche zusätzlichen Strukturen für Angriffe ausgenutzt werden können, eliminiert FrodoKEM somit dieses Risiko. Dafür ist FrodoKEM im Vergleich zu manch anderen gitterbasierten Schlüsseleinigungsverfahren etwas ineffizienter. Weitere Informationen zu FrodoKEM finden sich auch im BSI-Magazin [28].

FrodoKEM ist aktuell in der dritten Runde des NIST-Prozesses, allerdings unter den alternativen Kandidaten. NIST begründet diese Entscheidung damit, dass FrodoKEM zwar potenzielle Sicherheitsvorteile gegenüber anderen gitterbasierten Verfahren hat, dies aber auch mit einer schlechteren Performance einhergeht. Vor allem für den Hochsicherheitsbereich mit hohem Schutzbedarf sind die potenziellen Sicherheitsvorteile

schwerer zu gewichten als die Nachteile bei der Performance. Somit wurde FrodoKEM in die TR-02102-1 [7] aufgenommen.

2.3. Digitale Signaturen

Bei einem digitalen Signaturverfahren wird eine Nachricht mit einem Wert versehen, der die Überprüfung der Authentizität bzw. der nichtabstreitbaren Urheberschaft der Nachricht und deren Integrität erlaubt. Digitale Signaturverfahren gehören zu den asymmetrischen Kryptosystemen.

Im Gegensatz zur asymmetrischen Verschlüsselung und Schlüsseleinigung sind digitale Signaturverfahren nicht primär vom "Store now, decrypt later"-Szenario betroffen. Die Überlegung ist, dass man den Gültigkeitszeitraum der Signaturen in der Regel kontrollieren und somit diese mittels eines quantencomputerresistenten Signaturverfahrens erneuern bzw. austauschen kann. Diese Überlegung - für sich allein betrachtet - ist bei der Authentisierung einer Schlüsseleinigung unstrittig. D.h. in diesem Szenario wird ein quantencomputerresistentes Signaturverfahren erst dann benötigt, wenn Quantencomputer verfügbar werden. Jedoch ist eine reibungslose Migration der bestehenden digitalen Infrastrukturen aufwendig und nimmt einen gewissen Zeitraum in Anspruch. Daher ist auch im oben genannten Szenario und vor allem bei längerfristig gültigen Signaturen, z.B. im Rahmen von Public-Key-Infrastrukturen, Handeln geboten.

2.3.3. Hashbasierte Signaturverfahren

Das BSI empfiehlt seit Längerem Merkle-Signaturen als quantencomputerresistente Signaturverfahren in seiner Technischen Richtlinie TR-02101-1 [7]. Merkle-Signaturen sind hashbasierte Signaturverfahren, die mit Hilfe von Hashbäumen und Einmal-Signaturverfahren konstruiert werden und 1979 von Ralph Merkle eingeführt wurden [2]. Ihre Sicherheitseigenschaften sind sehr gut verstanden und sie gelten in ihrer aktuellen Form (LMS [29], XMSS [30]) als ausgereifte quantencomputerresistente Signaturverfahren. Ein entscheidender Nachteil ist jedoch die Zustandsbehaftung dieser Verfahren, d.h. dass der Signaturersteller exakt nachhalten muss, welche Einmal-Signaturschlüssel bereits verwendet wurden. Jeder Fehler bei diesem Vorgehen hat den Sicherheitsverlust zur Folge und es werden somit hohe Anforderungen an die Implementierung und Nutzung gestellt. Zudem ist die Anzahl der möglichen Signaturen beschränkt. Bei der Schlüsselgenerierung muss zwischen Signaturgröße und der Anzahl von erstellbaren Signaturen abgewogen werden. Daher eignen sich Merkle-Signaturen, neben symmetrischen Verfahren, vor allem zur zukunftssicheren Gestaltung von Software-Updateverfahren, bei denen die Zustandsbehaftung eine lösbare Anforderung stellt und die maximale Anzahl an benötigten Signaturen gut abgeschätzt werden kann. Dementsprechend werden Merkle-Signaturen in der Technischen Richtlinie TR-03140 im Rahmen des Satellitendatensicherheitsgesetz (SatDSiG) als zukunftssichere Digitale Signaturverfahren "for update-able crypto module by signature methods" empfohlen [31], Abschnitt 5.5.2.1. Die zustandsbehafteten hashbasierten Signaturverfahren LMS und XMSS wurden von der IETF bereits als RFC8554 [32] bzw. RFC8391 [33] standardisiert. NIST hat diese Standards als Special Publication 800-208 [34] übernommen. Parallel haben hashbasierte Signaturverfahren in Form von LMS Einzug in die Cryptographic Message

Syntax (CMS) [35] und die Concise Binary Object Representation (COSE) [36] gehalten. Beides sind grundlegende Datenformate, die z.B. bei S/MIME bzw. IoT verwendet werden.

3. Schlüsseleinigung und Anpassung kryptografischer Protokolle

Zurzeit steht man bei der Migration zu Post-Quanten-Kryptografie vor dem Problem, dass viele der neuen Verfahren bezüglich einiger Aspekte (beispielsweise Möglichkeiten für Seitenkanalangriffe) noch nicht so gut untersucht sind wie die derzeit eingesetzten Verfahren. Andererseits muss der Umstieg auf quantencomputerresistente Verfahren rechtzeitig erfolgen. Daher hat sich allgemein die Idee durchgesetzt, Post-Quanten Kryptografie nicht isoliert einzusetzen, sondern nur in Kombination mit etablierten Verfahren. In diesem Abschnitt wollen wir näher auf "hybride Schlüsseleinigung" eingehen, warum dieser Ansatz aus unserer Sicht sinnvoll ist und wie er sich in bestehende Protokolle integrieren lässt.

3.1. Hybride Schlüsseleinigung

Die Idee einer hybriden Schlüsseleinigung ist einfach: Man führe einen "klassischen" Schlüsselaustausch durch, danach noch eine Schlüsseleinigung mit einem quantencomputerresistenten Verfahren und kombiniere die so erhaltenen gemeinsamen Geheimnisse (sogenannte *shared secrets*) in einer geeigneten Weise, um einen geheimen Schlüssel für die Verschlüsselung der Nutzdaten zu erhalten. Schwierigkeiten ergeben sich, wenn man diese Idee in einem bestehenden Protokoll (beispielsweise dem Internet Key Exchange (IKE) Protokoll) umsetzen möchte. Dazu mehr in den folgenden Abschnitten. Außerdem stellt sich die Frage, wie konkret die Ableitung des geheimen Schlüssels aus den shared secrets erfolgen soll. Darauf soll hier eingegangen werden:

Eine Schlüsselableitungsfunktion bzw. Key Derivation Function (KDF) ist eine Funktion, mit der aus einem oder mehreren shared secrets kryptografisches Schlüsselmaterial abgeleitet wird, beispielsweise für die Verschlüsselung von Nutzdaten. Solch eine KDF kann im Kontext einer hybriden Schlüsseleinigung beispielsweise dazu genutzt werden, um aus den resultierenden shared secrets der einzelnen Schlüsselaustausche einen gemeinsamen kryptografischen Schlüssel abzuleiten. Gegebenenfalls kann zusätzlich zu den gemeinsamen shared secrets auch noch ein gemeinsamer vorverteilter Schlüssel (ein sogenannter preshared key) als Input eingehen. In seiner technischen Richtlinie TR-02101-1 [7] empfiehlt das BSI als Schlüsselableitungsfunktion die Key Derivation through Extraction-then-Expansion nach [37]. In der zweiten Revision [38] dieses Dokuments wird der hybride Fall berücksichtigt.

3.2. Multiple Key Exchange in IKEv2

Das Internet Key Exchange Protocol Version 2 (IKEv2) [39] wird zur Aushandlung des Schlüsselmaterials und zur Authentisierung der Kommunikationspartner für IPsec-Verbindungen verwendet. Die Schlüsselaushandlung in IKE basiert wesentlich auf einem klassischen Diffie-Hellman Schlüsseltausch (über elliptischen Kurven, (EC)DH), so dass quantensichere Alternativen dringend benötigt werden.

In dem Internet-Draft [40] wird ein Ansatz für eine hybride Schlüsseleinigung vorgeschlagen, die sehr flexibel ist. Dieser Ansatz nutzt einen sogenannten Intermediate Exchange, der in dem Internet-Draft [41] beschrieben ist. Dabei wird ein weiteres Paar von Nachrichten (IKE_INTERMEDIATE) zwischen den initialen Nachrichten (IKE_SA_INIT) und den Nachrichten zur Authentisierung (IKE_AUTH) ausgetauscht. In [40] werden sieben neue Transform Types definiert, die mit in der initialen Nachricht enthalten sein können. Jeder dieser Transform Types enthält eine Liste mit unterstützten (quantensicheren) Schlüsseleinigungsverfahren. So können bis zu sieben zusätzliche Schlüsselaushandlungen durchgeführt werden. Für jede dieser Schlüsselaushandlungen ist allerdings jeweils ein weiterer Intermediate Exchange notwendig. Zudem wird Transform Type 4, der bisher für die Aushandlung der für den Diffie-Hellman Schlüsseltausch verwendeten Gruppe diente (und konsequenter Weise Diffie_Hellman_Group heißt), in KE_Method (KE = Key Exchange) umbenannt. Die Liste für die Auswahl der möglichen Verfahren, die über Transform Type 4 ausgehandelt werden können, ist dieselbe, aus der auch die Verfahren für die weiteren Schlüsseleinigungen ausgesucht werden. Das bedeutet, dass für die initiale Schlüsseleinigung nicht zwingend ein klassischer Diffie-Hellman Austausch durchgeführt werden muss, sondern bereits hier ein quantencomputerresistentes Verfahren gewählt werden kann.

Ein Problem dabei kann sein, dass die öffentlichen Schlüssel einiger Verfahren wesentlich größer sind als die bisher verwendeten und nicht in die Key Exchange Payload der initialen IKE Nachricht (IKE_SA_INIT) passen. Zudem sieht IKEv2 keine Möglichkeit vor, diese initialen Nachrichten zu fragmentieren, was bei dem (üblichen) Transport über UDP zu Fragmentierung auf IP-Ebene und damit bei einigen Netzwerkknoten zum Verlust einzelner Pakete und somit zum Scheitern des Aushandelns der Sicherheitsbeziehung führen kann. Eine Übertragung von großen Schlüsseln ist aber über die Intermediate-Nachrichten möglich, da es für diese einen IKE-spezifischen Mechanismus zur Fragmentierung gibt, siehe [42]. Auch hier ist aber die Größe der Schlüssel begrenzt und zwar durch die Größe eines IKE Encrypted_Payload.

3.3. TLS Hybrid

Das Transport Layer Security (TLS) Protokoll wird zur sicheren Übertragung von Daten im Internet verwendet, die aktuellste Version ist 1.3 [43]. In einem sogenannten TLS-Handshake werden die dafür benötigten Schlüssel ausgehandelt und die Kommunikationspartner gegenseitig authentifiziert. Bisher werden auch für die Schlüsselaushandlung im TLS-Handshake "klassische" Verfahren wie RSA oder (EC)DH verwendet.

Im April 2020 hat die TLS Arbeitsgruppe der Internet Engineering Task Force (IETF) einen Draft für einen RFC veröffentlicht [44], in dem eine Lösung für hybride Schlüsseleinigung in TLS 1.3 vorgeschlagen wird. Der Draft beruht auf früheren Entwürfen von D. Stebila (University of Waterloo), S. Fluhrer (Cisco Systems) und S. Gueron (Amazon Web Services). Der Lösungsansatz in [44] sieht im Wesentlichen vor, neue Object Identifier (OIDs) für Kombinationen aus jeweils einem klassischen und einem Post-Quanten-Verfahren zu registrieren und diese über die „NamedGroup“ Erweiterung im Handshake auszuhandeln. Für jede Kombination, die der Client anbietet, sollte er allerdings auch schon die entsprechenden öffentlichen Schlüssel in seiner Client Hello

Nachricht verschicken. Dieser Ansatz führt einerseits zu einer Vielzahl von benötigten neuen OIDs und andererseits unter Umständen zu sehr großen Client Hello Nachrichten. Der intendierte Status des Draft-RFC ist "informational", d.h. Implementierungen müssen ihn nicht zwingend umsetzen, um konform zum aktuellen TLS 1.3 Standard zu bleiben. Es ist jedoch möglich, dass informational RFCs zu einem späteren Zeitpunkt in einen Standards Track RFC übergehen.

4. Authentisierung und die Zukunft von Public-Key-Infrastrukturen

Beim Einsatz von digitalen Signaturverfahren zur Authentisierung von Nachrichten und Gestaltung von Public-Key-Infrastrukturen steht man vor ähnlichen Problemen wie bei der Schlüsseleinigung in Sicherheitsprotokollen. Auch hier liegt die Idee nahe, eine Kombination von quantencomputerresistenten Signaturverfahren mit etablierten klassischen Verfahren wie ECDSA in Form von "hybriden Zertifikaten" einzusetzen. Für Public-Key-Infrastrukturen wird derzeit alternativ zu Stichtagsumstellungen oder parallelen PKIn an Migrationspfaden zur Einführung von quantencomputerresistenten Signaturverfahren gearbeitet. Auf diese Aspekte und speziell die beiden Zertifikatsformate X.509 und OpenPGP wollen wir in diesem Abschnitt näher eingehen.

4.1. X.509

Im Oktober 2019 wurde von ITU-T ein Update des X.509v3 Standards [45] veröffentlicht. Darin wird erstmals das Problem adressiert, dass neue Signaturverfahren ohne eine Stichtagsumstellung in Zertifikate bzw. Public-Key-Infrastrukturen eingebracht werden müssen. Die ITU-T kommt zu dem Schluss: "it is unlikely that it is possible to change cryptographic algorithms simultaneously for all entities within a PKI or PMI". Um eine Migration von alten zu neuen Verfahren zu ermöglichen, werden Zertifikatserweiterungen (subjectAltPublicKeyInfo, altSignatureAlgorithm und altSignatureValue) spezifiziert, so dass ein X.509-Zertifikat einen "alternativen" öffentlichen Schlüssel enthalten kann [45], § 7.22. Unter den in diesem Zusammenhang beschriebenen Regeln für die Erstellung und Validierung der Zertifikate sowie den Implikationen für CRLs und AVLs stechen drei Aspekte hervor. Erstens, aus Kompatibilitätsgründen wird empfohlen, diese Erweiterungen als "non-critical" zu markieren [45], §9.8.2., §9.8.3, damit auch Anwendungen, die diese Erweiterungen nicht kennen, entsprechende Zertifikate als gültig prüfen können. Zweitens, der beschriebene Ansatz stellt keine "hybride Lösung" dar, da bei Vorhandensein nur die alternativen Werte für Schlüssel, Algorithmus und Signatur verwendet bzw. geprüft werden sollen. Drittens, die Anpassungen im Zertifikatsaufbau sind nur als Übergangslösung gedacht, bis der Migrationsprozess zu quantencomputerresistenten Signaturverfahren abgeschlossen ist. Hierzu schreibt die ITU: "After the migration period, it is expected that new public-key certificates be issued without these extensions and with the new set of cryptographic algorithms and the digital signature in the base part of the public-key certificate." [45], §7.22.

4.2. OpenPGP

Der aktuell gültige OpenPGP Standard aus dem Jahr 2007 [46] befindet sich derzeit in einer Aktualisierung [47], die in 2015 begonnen hat und in der noch keine Ansätze zur

Migration auf Post-Quanten-Kryptografie zu erkennen sind. Allerdings gab es beim OpenPGP Summit in 2018 einen Vorstoß in Richtung Post-Quanten-Kryptografie [48].

4.3. Mixed PKI

Neben den beschriebenen Änderungen am Zertifikatsaufbau gibt es erste Ansätze, quantencomputerresistente Signaturverfahren innerhalb des X.509 Standards zu definieren. Insbesondere für die bereits standardisierten hashbasierten Signaturverfahren LMS und XMSS gab es hier einen Vorstoß in Form eines IETF RFC Drafts [49], der jedoch im September 2019 abgelaufen ist. Eine Ansicht, die auch in dem genannten RFC vertreten wird, ist, dass sich diese Signaturverfahren aufgrund ihrer Zustandsbehaftung am ehesten für die Gestaltung von langlebigen Wurzelzertifikaten und weniger für Endnutzerzertifikate eignen, und somit zum Aufbau einer gemischten PKI dienen können.

5. Empfehlungen des BSI

Grundsätzlich geben die technischen Richtlinien TR-02102 des BSI Empfehlungen zu kryptografischen Verfahren und Protokollen (TLS, IKE/IPsec, SSH). Schon seit Längerem werden in der TR-02102-1 [7] hashbasierte Merkle-Signaturen empfohlen. Anfang 2020 wurden die zwei quantencomputerresistenten Schlüsseleinigungsverfahren Classic McEliece und FrodoKEM in die TR-02102-1 aufgenommen.

Da die Post-Quanten-Kryptografie einen Umbruch in der Kryptografie bedeutet und einer gesonderten Aufmerksamkeit bedarf, hat das BSI im März 2020 erste Handlungsempfehlungen veröffentlicht [5], wie eine Migration zu Post-Quanten-Kryptografie heute schon begonnen werden kann. Diese wurden im August 2020 noch einmal aktualisiert [50], nachdem NIST die Kandidaten für die dritte Runde bekannt gegeben hatte. Das BSI hält weiter an der Entscheidung fest, FrodoKEM in der TR-02102-1 zu empfehlen, auch wenn dieses Verfahren von NIST nur zu den alternativen Kandidaten zugeordnet wurde.

Neben allgemeinen Designprinzipien wie Kryptoagilität und hybriden Lösungen, insbesondere bei der Anpassung kryptografischer Protokolle, werden konkrete Empfehlungen zu kurzfristigen Schutzmaßnahmen, Schlüssellängen für symmetrische Verschlüsselung, hashbasierten Signaturverfahren und quantencomputerresistenter Schlüsseleinigung gegeben. Diese Handlungsempfehlungen werden auch in Zukunft kontinuierlich weiterentwickelt werden. Nach Abschluss des NIST-Prozesses werden zudem voraussichtlich weitere quantencomputerresistente Verfahren in die TR-02102-1 aufgenommen werden.

Literaturhinweise

[1] R. McEliece: „A Public-Key Cryptosystem Based on Algebraic Coding Theory“, Deep Space Network Progress Report, Band 42, Nr. 44, 1978, S. 114–116.

[2] R. Merkle: „Secrecy, Authentication, and Public Key Systems", Stanford University Information Systems Laboratory Technical Report 1979-1, 1979.

[3] P. Shor: "Algorithms for quantum computation: Discrete logarithms and factoring", Proceedings, 35th Annual Symposium on Foundations of Computer Science, 1994, IEEE Computer Society Press, pp. 124-134, http://math.mit.edu/~shor/papers/algsfqc-dlf.pdf.

[4] Bundesamt für Sicherheit in der Informationstechnik: Studie "Entwicklungsstand Quantencomputer", www.bsi.bund.de/qcstudie.

[5] Bundesamt für Sicherheit in der Informationstechnik: Pressemeldung "Post-Quanten-Kryptografie: BSI veröffentlicht Handlungsempfehlungen", März 2020, https://www.bsi.bund.de/DE/Service-Navi/Presse/Alle-Meldungen-News/Meldungen/Post-Quanten-Kryptografie_260320.

[6] National Institute of Standards and Technology (NIST): "Post-Quantum Cryptography", https://csrc.nist.gov/Projects/Post-Quantum-Cryptography.

[7] Bundesamt für Sicherheit in der Informationstechnik: TR-02102-1 "Kryptographische Verfahren: Empfehlungen und Schlüssellängen", 2020-01, https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr02102/tr02102_node.

[8] M. R. Albrecht, D. J. Bernstein, T. Chou, C. Cid, J. Gilcher, T. Lange, V. Maram, I. von Maurich, R. Misoczki, R. Niederhagen, K. G. Paterson, E. Persichetti, C. Peters, P. Schwabe, N. Sendrier, J. Szefer, C. J. Tjhai, M. Tomlinson, W. Wang: "Classic McEliece", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[9] P. Schwabe, R. Avanzi, J. Bos, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, J. M. Schanck, G. Seiler, D. Stehle: "CRYSTALS-KYBER", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[10] C. Chen, O. Danba, J. Hoffstein, A. Hulsing, J. Rijneveld, J. M. Schanck, P. Schwabe, W. Whyte, Z. Zhang, T. Saito, T. Yamakawa, K. Xagawa: "NTRU", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions

[11] J.-P. D'Anvers, A. Karmakar, S. S. Roy, F. Vercauteren, J. M. Bermudo Mera, M. Van Beirendonck, A. Basso: "SABER", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[12] V. Lyubashevsky, L. Ducas, E. Kiltz, T. Lepoint, P. Schwabe, G. Seiler, D. Stehle, S. Bai: "CRYSTALS-DILITHIUM", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[13] T. Prest, P.-A. Fouque, J. Hoffstein, P. Kirchner, V. Lyubashevsky, T. Pornin, T. Ricosset, G.Seiler, W. Whyte, Z. Zhang: "FALCON", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[14] J. Ding, M.-S. Chen, A. Petzoldt, D. Schmidt, B.-Y. Yang, M. Kannwischer, J. Patarin: "Rainbow", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[15] N. Aragon, P. Barreto, S. Bettaieb, L. Bidoux, O. Blazy, J.-C. Deneuville, P. Gaborit, S. Gueron, T. Guneysu, C. A. Melchor, R. Misoczki, E. Persichetti, N. Sendrier, J.-P. Tillich, G. Zemor, V. Vasseur, S. Ghosh: "BIKE", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[16] M. Naehrig, E. Alkim, J. Bos, L. Ducas, K. Easterbrook, B. LaMacchia, P. Longa, I. Mironov, V. Nikolaenko, C. Peikert, A. Raghunathan, D. Stebila: "FrodoKEM", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[17] C. A. Melchor, N. Aragon, S. Bettaieb, L. Bidoux, O. Blazy, J.-C. Deneuville, P. Gaborit, E. Persichetti, G. Zémor, J. Bos: "HQC", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[18] D. J. Bernstein, B. B. Brumley, M.-S. Chen, C. Chuengsatiansup, T. Lange, A. Marotzke, B.-Y. Peng, N. Tuveri, C. van Vredendaal, B.-Y. Yang: "NTRU Prime", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[19] D. Jao, R. Azarderakhsh, M. Campagna, C. Costello, L. De Feo, B. Hess, A. Jalali, B. Koziel, B. LaMacchia, P. Longa, M. Naehrig, J. Renes, V. Soukharev, D. Urbanik, G. Pereira, K. Karabina, A. Hutchinson: "SIKE", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[20] A. Casanova, J.-C. Faugere, G. Macario-Rat, J. Patarin, L. Perret, J. Ryckeghem: "GeMSS", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[21] G. Zaverucha, M. Chase, D. Derler, S. Goldfeder, C. Orlandi, S. Ramacher, C. Rechberger, D. Slamanig, J. Katz, X. Wang, V. Kolesnikov, D. Kales: "Picnic", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[22] A. Hülsing, D. J. Bernstein, C. Dobraunig, M. Eichlseder, S. Fluhrer, S.-L. Gazdag, P. Kampanakis, S. Kolbl, T. Lange, M. M. Lauridsen, F. Mendel, R. Niederhagen, C. Rechberger, J. Rijneveld, P. Schwabe, J.-P. Aumasson, B. Westerbaan, W. Beullens: "SPHINCS+", National Institute of Standards and Technology, 2020, abrufbar unter https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[23] D. Bernstein, T. Lange, C. Peters: „Attacking and Defending the McEliece Cryptosystem“, Proceedings of the 2nd International Workshop on Post-Quantum Cryptography, 2008, S. 31-46.

[24] R. Misoczki, P. Barreto: „Compact McEliece keys from Goppa codes“. Selected Areas in Cryptography (SAC 2009), August 2009.

[25] J.-C. Faugère, A. Otmani, L. Perret, F. de Portzamparc, J.-P. Tillich: „Structural cryptanalysis of McEliece schemes with compact keys“, Designs, Codes and Cryptography, Volume 79, Issue1, 2016, S. 87–112.

[26] H. Niederreiter: „Knapsack-type cryptosystems and algebraic coding theory", Problems of Control and Information Theory, Problemy Upravlenija i Teorii Informacii, 15, S. 159–166, 1986.

[27] O. Regev: "On lattices, learning with errors, random linear codes, and cryptography", Proceedings of the 37th Annual {ACM} Symposium on Theory of Computing, S. 84--93, 2005.

[28] H. Hagemeier: "Frodo ist die 'neue Hoffnung'", in BSI-Magazin 2020/01, S. 12-14, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Magazin/BSI-Magazin_2020_01.

[29] T. Leighton, S. Micali: "Large provably fast and secure digital signature schemes from secure hash functions", U.S. Patent 5,432,852, Juli 1995.

[30] J. Buchmann, E. Dahmen, A. Huelsing: „XMSS – A Practical Forward Secure Signature Scheme Based on Minimal Security Assumptions", Lecture Notes in Computer Science: Post-Quantum Cryptography, 2011.

[31] Bundesamt für Sicherheit in der Informationstechnik: TR-03140 "Technical Guideline SatDSiG", https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr03140/TR-03140_node.

[32] D. McGrew, M. Curcio, S. Fluhrer: „Leighton-Micali Hash-Based Signatures“, IETF RFC 8554, April 2019, https://tools.ietf.org/html/rfc8554.

[33] A. Huelsing, D. Butin, S. Gazdag, J. Rijneveld, A. Mohaisen: „XMSS: eXtended Merkle Signature Scheme“, IETF RFC 8391, Mai 2018, https://tools.ietf.org/html/rfc8391.

[34] National Institute of Standards and Technology: Special Publication 800-208 “Recommendation for Stateful Hash-Based Signature Schemes”, Oktober 2020, https://csrc.nist.gov/publications/detail/sp/800-208/final.

[35] R. Housley: "Use of the HSS/LMS Hash-Based Signature Algorithm in the Cryptographic Message Syntax (CMS)", IETF RFC 8708, Februar 2020, https://tools.ietf.org/html/rfc8708

[36] R. Housley: "Use of the HSS/LMS Hash-Based Signature Algorithm with CBOR Object Signing and Encryption (COSE)", IETF RFC 8778, April 2020, https://tools.ietf.org/html/rfc8778.

[37] Nationa Institute of Standards and Technology (NIST): “Recommendation for Key-Derivation Methods in Key-Establishment Schemes”, Special Publication 800-56C, Rev.1, April 2018.

[38] Nationa Institute of Standards and Technology (NIST): “Recommendation for Key-Derivation Methods in Key-Establishment Schemes”, Special Publication 800-56C, Rev.2, August 2020.

[39] C. Kaufman, P. Hoffman, Y. Nir, P. Eronen, T. Kivinen: „Internet Key Exchange Protocol Version 2 (IKEv2), IETF RFC 7296, Oktober 2014, https://tools.ietf.org/html/rfc7296.

[40] C. Tjhai, M. Tomlinson, G. Bartlett, S. Fluhrer, D. Van Geest, O. Garcia-Morchon, V. Smyslov: "Multiple Key Exchanges in IKEv2", Juli 2020, https://tools.ietf.org/html/draft-ietf-ipsecme-ikev2-multiple-ke-01.

[41] V. Smyslov: "Intermediate Exchange in the IKEv2 Protocol", September 2020, https://tools.ietf.org/html/draft-ietf-ipsecme-ikev2-intermediate-05.

[42] V. Smyslov: "Internet Key Exchange Protocol Version 2 (IKEv2) Message Fragmentation", RFC 7383, November 2014.

[43] E. Rescorla: "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, August 2018.

[44] D. Stebila, S. Fluhrer, S. Gueron: "Hybrid key exchange in TLS 1.3", April 2020, https://tools.ietf.org/html/draft-ietf-tls-hybrid-design-01.

[45] International Telecommunication Union (ITU): “ITU-T Recommendation X.509”, Oktober 2019.

[46] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, R. Thayer: "OpenPGP Message Format", IETF RFC 4880, November 2007.

[47] W. Koch, B. Carlson, R. Tse, D. Atkins, D. Gillmor: "OpenPGP Message Format", IETF RFC Draft 4880bis, August 2020, https://tools.ietf.org/html/draft-ietf-openpgp-rfc4880bis-10.

[48] P. Zimmermann: "Improve OpenPGP", Plenary Session, OpenPGP Summit, Oktober 2018, https://wiki.gnupg.org/Summit2018PlenaryPhil.

[49] D. Van Geest, S. Fluhrer: "Algorithm Identifiers for HSS and XMSS for Use in the Internet X.509 Public Key Infrastructure", IETF RFC Draft, März 2019, https://tools.ietf.org/html/draft-vangeest-x509-hash-sigs-03.

[50] Bundesamt für Sicherheit in der Informationstechnik: "Migration zu Post-Quanten-Kryptografie – Handlungsempfehlungen des BSI", August 2020, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Krypto/Post-Quanten-Kryptografie.pdf

Zurück zum Inhaltsverzeichnis

Zurück zum Inhaltsverzeichnis

Zur Integration von Post-Quantum Verfahren in bestehende Softwareprodukte

Alexander Zeier[1], Prof. Dr. Alexander Wiesmaier[1], Prof. Dr. Andreas Heinemann[1]

Kurzfassung:
Aktuell werden PQC-Algorithmen standardisiert, um der aufziehenden Gefahr für konventionelle asymmetrische Algorithmen durch Quantencomputer zu begegnen. Diese neuen Algorithmen müssen dann in bestehende Protokolle, Applikationen und Infrastrukturen eingebunden werden. Dabei ist mit Integrationsproblemen zu rechnen, die einerseits durch Inkompatibilitäten mit existierenden Standards und Implementierungen begründet sind, andererseits aber auch durch fehlendes Wissen der Softwareentwickler über die Handhabung von PQC-Algorithmen zustande kommen. Um Inkompatibilitäten beispielhaft aufzuzeigen, integrieren wir zwei unterschiedliche PQC-Algorithmen in zwei verschiedene bestehende Softwareprodukte (InboxPager E-Mail Client und TLS Implementierung der Bouncy Castle Bibliothek). Hierbei setzen wir auf die hoch-abstrahierende Krypto-Bibliothek eUCRITE, die Entwicklern das Detailwissen über die korrekte Verwendung klassischer und PQC-Algorithmen abnimmt und damit bereits einige potenzielle Implementierungsfehler vermeidet. Die dabei zutage getretenen Probleme bestätigen teilweise bereits bekannte Inkompatibilitäten, beinhalten aber auch neue, bisher nicht angesprochene Schwierigkeiten.

Stichworte: API, eUCRITE-API, Kryptoagilität, Post-Quantum Verfahren

1. Einführung

Quantencomputer sind Gegenstand der laufenden Forschung. Bei ausreichender Leistung, d.h. wenn Shor's Algorithmus [23] auf einem Quantencomputer mit ausreichender Qubitlänge ausgeführt werden kann, wird man in der Lage sein, die derzeit verwendeten asymmetrischen Algorithmen wie RSA, DSA, ECDSA und ECDH zu brechen [9]. Der Bedarf an Post-Quanten-Kryptografie (PQC), insbesondere asymmetrischen Verfahren[2], ist offensichtlich, da potenziell unsicher werdende asymmetrische Verfahren in vielen ausgerollten hybriden Kryptosystemen zu finden sind.

Dieser Beitrag stellt die im Rahmen des Forschungsprojektes *Use-A-PQClib* [13] entwickelte eUCRITE-API [12] vor, die als Designziele eine gute Benutzbarkeit und Verständlichkeit für einen Entwickler sowie eine hohe Abstraktion von technischen Parametern wie beispielsweise Schlüssellängen von Krypto-Algorithmen aufweist.

Der Hauptteil beinhaltet darauf aufbauend einen Erfahrungsbericht bei der Integration der PQC-Verfahren McEliece [19] und SPHINCS+ [6] auf Basis der eUCRITE-API in zwei bestehende Softwareprodukte. Zum einen die Integration in InboxPager[3], einen E-Mail Client für das Android Betriebssystem, zum anderen die Integration in die TLS

[1] Hochschule Darmstadt, FB Informatik, Haardtring 100, 64295 Darmstadt

[2] Auch symmetrische Verfahren wie DES oder AES sind durch den Algorithmus von Grover ([11]) bedroht, jedoch könne längere Schlüssel hier die Gefahr lindern.

[3] https://github.com/itprojects/InboxPager (besucht am 29.12.2020)

Implementierung von Bouncy Castle[4]. Bouncy Castle ist eine weit verbreitete Krypto-Bibliothek für Java und C#, welche neben den grundlegenden Krypto-Operationen wie Verschlüsseln und Signieren auch eine TLS Implementierung zur Verfügung stellt. Hierbei konzentriert sich die Integration auf den Austausch der klassischen asymmetrischen Krypto-Verfahren gegen die oben genannten PQC-Verfahren der jeweiligen hybriden Kryptosysteme.

Damit soll aufgezeigt werden, mit welchen Herausforderungen und technischen Problemen Entwickler bei der Integration von PQC-Verfahren rechnen müssen. Des Weiteren werden die Vor- und Nachteile sowie die Implikationen einer hohen API-Abstraktion vorgestellt.

Abschließend wird ein Fazit gezogen und ein Ausblick auf weitere Schritte und noch zu adressierende Fragestellungen bei der Umstellung von klassischen auf PQC-Verfahren gegeben.

2. Verwandte Arbeiten

Der Dagstuhl Report [4] berichtet über *Biggest Failures in IT Security* und empfiehlt (unter anderem) Entwickler bei der Implementierung von Sicherheitsmechanismen stärker zu unterstützen. Dazu gehört zum einen die Bereitstellung von Werkzeugen und Methoden, die es dem Entwickler leichter machen guten Code zu schreiben (*to do the good/right thing*, Seite 20). Zum anderen gilt es, die verschiedenen Kenntnisse und Vorlieben der verschiedenen Entwickler- bzw. Benutzergruppen zu beachten.

Ott und Peikert [21] präsentieren zahlreiche Forschungsfragen zur Kryptoagilität [18] und Migration nach PQC und decken dabei auf hohem Diskussionsniveau ein weites Feld an Themen ab, darunter Implementierungsaspekte. Dabei geht es den Autoren nicht nur um die Umsetzung der in mathematischen Formeln ausgedrückten PQC-Algorithmen auf verschiedensten Plattformen und in unterschiedlichen Programmiersprachen. Es ist ebenfalls von enormer Wichtigkeit, die implementierten Algorithmen so in vorhandene Systeme einzubringen, dass Kontinuität und Interoperabilität während der Migrationsphase erhalten bleiben.

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) gibt Handlungsempfehlungen zur *Migration zu Post-Quanten-Kryptografie* [24] und empfiehlt (zumindest während der Migrationsphase) den Einsatz hybrider Lösungen, d.h. der Kombination klassischer und PQC-Algorithmen, und die entsprechende Anpassung kryptografischer Protokolle. Dabei soll die Umsetzung dem Prinzip der Kryptoagilität folgen, um auch zukünftige Empfehlungen und Standards entsprechend umsetzen zu können. Als besonders dringend gilt der Umstieg auf (hybride) PQC-Verfahren bei Schlüsseleinigungsverfahren zum Schutz langfristiger Geheimnisse.

Campagna u. a. [8] nennen, neben der Beschreibung des State-of-the-Art in PQC und Anpassungsempfehlungen, um die Standards X.509, IKEv2, TLS 1.2, S/MIME und SSH2 PQC-bereit zu machen, wichtige Anwendungsfelder und -fälle für Kryptografie.

[4] https://www.bouncycastle.org (besucht am 29.12.2020)

Je nach bereits vorhandener Kryptoagilität beschränken sich die Empfehlungen auf die einfache Einführung neuer OIDs bzw. Cipher Suites oder beinhalten die mehr oder weniger tiefgreifende Anpassung vorhandener Standards. Die Gefahren von Quantencomputer-Angriffen auf typische Anwendungsfelder (Verschlüsselung, Authentisierung etc.) werden vorgestellt und in dieser Hinsicht besondere Verwundbarkeiten verschiedener Industriezweige diskutiert.

Crockett, Paquin und Stebila [10] stellen ihre Integration von PQC-Verfahren in die Protokolle TLS 1.2, TLS 1.3 und SSH2 vor, und berichten von den dabei zu bewältigenden Herausforderungen. Ein Beispiel für eine solche Herausforderung ist die zertifikatsbasierte Übermittlung mehrerer Schlüssel für unterschiedliche Verfahren, die in hybriden Umsetzungen benötigt werden. Ein anderes Beispiel sind Limitierungen für die Größe von Nachrichten zum Schlüsseltausch oder Signaturen, die teilweise implementierungsbedingt sind, aber teilweise auch auf die Spezifikationen in den Standards zurückzuführen sind.

Herath Mudiyanselage [14] untersucht, unter anderem, Eigenschaften von hybriden Signaturverfahren und deren Integration in verschiedene Implementierungen von X.509, TLS und S/MIME. Als wichtige Eigenschaften hybrider Signaturverfahren werden Unfälschbarkeit unter Signaturorakeln (EUF-CMA) und Nicht-Separierbarkeit[5] identifiziert. Die Integration eines zweiten Signaturschemas wurde über Erweiterungsmechanismen (X.509 extensions), nachgelagerte Authentifizierung (TLS) oder parallele/verschachtelte Signaturen (S/MIME) erreicht. In vielen Konstellationen führt die Einführung von Zusatzinformationen über ca. 40 KiB zu Kompatibilitätsproblemen.

Die vorliegende Arbeit eruiert an zwei real umgesetzten PQC-Migrationen die dabei auftretenden Probleme und konkretisiert so die in den obigen Arbeiten aufgeführten theoretischen Überlegungen bzw. ergänzt die dort vorgestellten praktischen Herausforderungen.

3. Verwendete Softwarekomponenten und Produkte

Im Rahmen dieser Arbeit wurden Post-Quantum Verfahren in zwei Software-Produkte integriert. Hierbei wurde zum einen ein E-Mail-Client ausgewählt, da es sich hierbei mit über 300 Milliarden E-Mails pro Tag[6] um eine der wesentlichen Anwendungen im Internet handelt. Zum anderen wurde mit TLS 1.2 ein Sicherheits-Protokoll gewählt, welches der aktuell gebräuchliche Standard zur Absicherung von HTTP im Internet ist. Zur Bereitstellung der benötigten PQC-Algorithmen wurde auf die Krypto-Bibliothek eUCRITE-API zurückgegriffen. Da es sich dabei um eine Java-API handelt, sind die beiden betrachteten Implementierungen ebenfalls in Java geschrieben. Die API sowie die Softwareprodukte werden im Folgenden kurz vorgestellt.

[5] Eine hybride Signatur kann vom Angreifer nicht in eine Einzelsignatur umgeformt werden.

[6] https://de.statista.com/statistik/daten/studie/252278/umfrage/prognose-zur-zahl-der-taeglich-versendeter-e-mails-weltweit/ (besucht am 29.12.2020)

3.1. eUCRITE-API

Die eUCRITE-API [28, 12] ist eine kryptografische Bibliothek mit Fokus auf einfache Benutzbarkeit, die auch Laien die sichere Verwendung kryptografischer Verfahren ermöglicht. Dazu werden sogenannte *Templates* eingesetzt, um die Auswahl der Algorithmen und Parameter für den Entwickler transparent zu gestalten. Die Auswahl wird durch die einfache Angabe des gewünschten Sicherheitsniveaus umgesetzt. Hier hat ein Entwickler die Wahl zwischen den Niveaus `LOW`, `MEDIUM`, `HIGH`. Das Sicherheitsniveau `HIGH` würde intern dann beispielsweise zur Wahl der PQC Verfahren McEliece und SPHINCS+ führen.

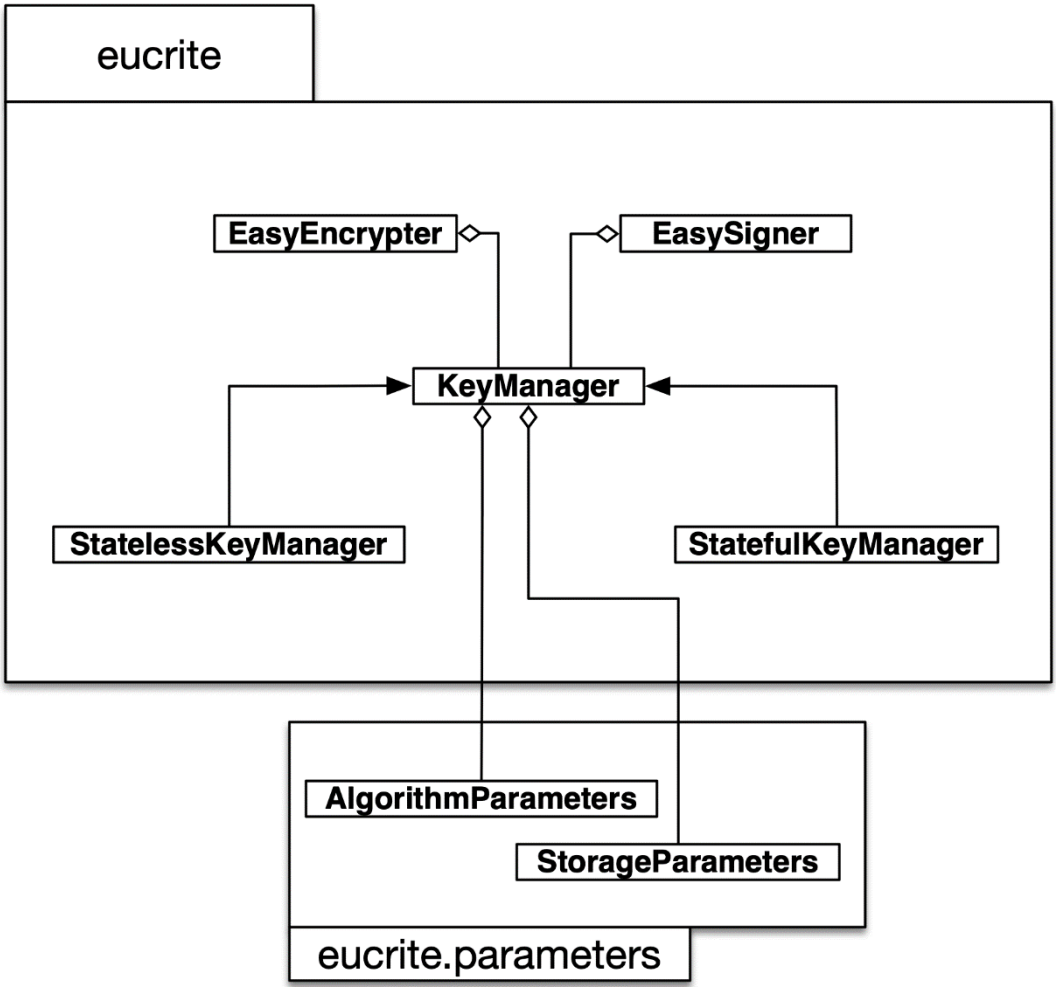


Abb. 1: Klassendiagramm der eUCRITE-API

Abbildung 1 zeigt das Klassendiagramm der eUCRITE-API. Die Klasse `EasyEncrypter` enthält Funktionen zur Ver- und Entschlüsselung, die Klasse `EasySigner` zur Signierung und Verifizierung von Daten. Beide Klassen nutzen einen `KeyManager` zur Speicherung des Schlüsselmaterials. Je nach Algorithmus wird ein `StatelessKeyManager` oder `StatefulKeyManager` verwendet. Bei der Erzeugung eines neuen Schlüsselpaares werden durch `AlgorithmParameters` der verwendete

Algorithmus sowie dessen Parameter festgelegt. `StorageParameters` geben den Speicherort des Schlüsselmaterials an.

Die Usability der eUCRITE-API wurde in einer Reihe von vorgelagerten Nutzerstudien untersucht. Bei der zu diesem Zeitpunkt letzten Studie [15] wurde ein API-Usability Score nach Acar [2] von 70,5 (aus maximal 100) erzielt. Im direkten Vergleich hat die API *tink*[7] einen Score von 48,23 erreicht. In einer früheren Studie wurde insbesondere die Benutzbarkeit von zustandsbehafteten Verfahren in der eUCRITE-API untersucht ([28]).

```
1 Signature sig = Signature.getInstance("SHA256withRSA", "BC");
2 sig.initSign(privateKey);
3 byte[] toBeSigned = "Hallo␣Welt!".getBytes();
4 sig.update(toBeSigned, 0, toBeSigned.length);
5 byte[] signature = sig.sign();
```

List. 1: Signieren von Daten (klassisch) mit BouncyCastle

Das Codebeispiel in Listing 1 zeigt – vereinfacht – Methodenaufrufe für die Signierung mit einem klassischen Verfahren (SHA256 und RSA) mithilfe der Implementierung von Bouncy Castle, welcher als Provider über die Java Cryptography Extension (JCE) eingebunden wird.

```
1 Signature sig = Signature.getInstance("SHA3-512WITHSPHINCS256", "BCPQC");
```

List. 2: Neue Zeile 1: Signieren von Daten (PQC) mit BouncyCastle

Um auf ein PQC Verfahren zu wechseln, müsste hier die erste Zeile mit dem Codebeispiel aus Listing 2 ersetzt werden, die Zugriff auf PQC-Verfahren von Bouncy Castle erlaubt. Beide Codebeispiele illustrieren, dass ein Entwickler eine Reihe von technischen Parametern kennen und korrekt anwenden muss, z.B. den Namen und die Bitlänge des zu verwendenden Hash-Algorithmus.

```
1 AlgorithmParameters algorithmParameters =
2       AlgorithmParameters.Template.Signature.Security_Level.HIGH.getParameters();
3 KeystoreParameters keystoreParameters = new KeystoreParameters(keyStoreFile, "password");
4 EasySigner signer = EasySigner.withNewKey(algorithmParameters, keystoreParameters);
5 byte[] signature = signer.sign("Hallo␣Welt!");
```

List. 3: Signieren von Daten mit der eUCRITE-API

In eUCRITE kann hingegen das bereits erwähnte Template verwendet werden (siehe Codebeispiel in Listing 3). Die Wahl des Sicherheitslevels `HIGH` für Signaturen führt intern z. Zt. zur Auswahl von SHA3-512 und SPHINCS+.

In einem späteren Schritt soll eUCRITE durch einen Update Mechanismus stets ein zur Laufzeit als sicher geltendes Verfahren auswählen, d.h.: sollte SPHINCS+ in Zukunft gebrochen werden, würde hier ein anderer Algorithmus zum Einsatz kommen.

[7] https://github.com/google/tink (besucht am 29.12.2020)

3.2. InboxPager (Android E-Mail Client)

Als erster Untersuchungsgegenstand zur Integration von PQC-Verfahren auf Basis der eUCRITE-API wurde ein E-Mail Client ausgewählt, da hier sowohl asymmetrische Verschlüsselung als auch Signaturen zum Einsatz kommen. Die Auswahl des Clients erfolgte auf Basis der folgende Kriterien:

1. Klassische Kryptografie-Verfahren müssen bereits integriert sein
2. die kryptografischen Funktionen sind gut getrennt vom restlichen Code
3. die Anwendung ist nicht zu umfangreich
4. die Anwendung ist in Java geschrieben.

Durch eine erste Recherche standen die folgenden drei E-Mail Clients zur Auswahl: K9 [26], InboxPager [16] und FairMail [1]. Anhand der oben genannten Kriterien fiel die Wahl auf InboxPager (Version 4.5).

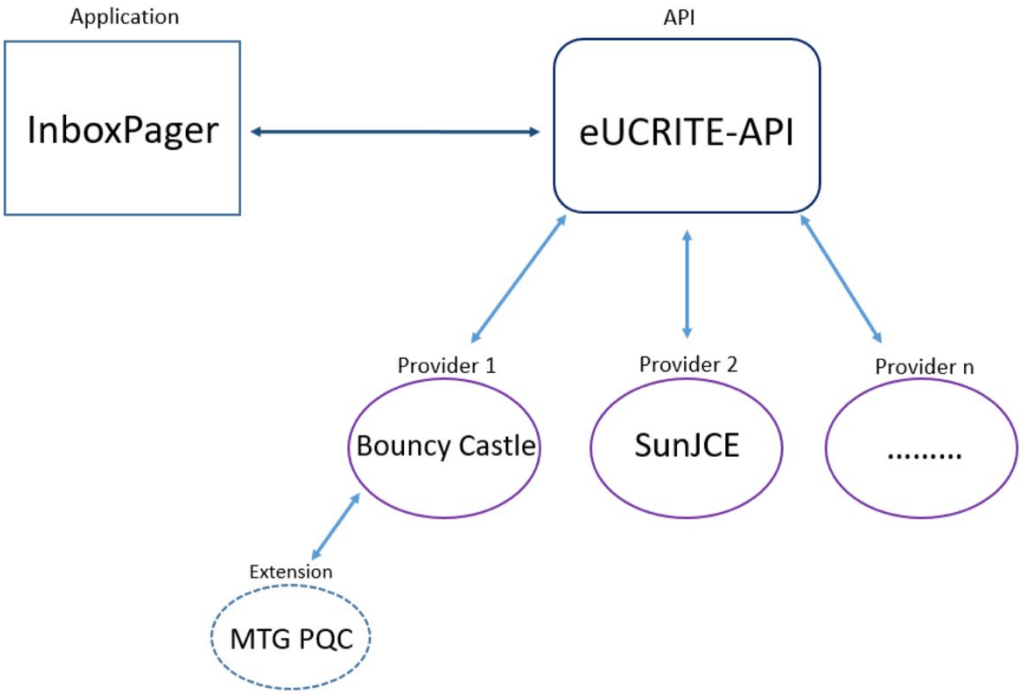


Abb. 2: Abhängigkeiten der kryptografischen Bibliotheken, Quelle [27]

Abbildung 2 zeigt die Abhängigkeiten der verwendeten kryptografischen Bibliotheken. In der überarbeiteten InboxPager Implementierung werden die Funktionen zur Verschlüsselung und Signierung der E-Mails über die eUCRITE-API implementiert. Intern greift diese, abhängig vom Krypto-Verfahren, auf verschiedene Provider zurück. Die beiden PQC Verfahren Classic McEliece und SPHINCS+ werden über die MTG PQC Extension[8] des Bouncy Castle Providers angesprochen.

Für weitere Details wird auf die Bachelorarbeit von Yousofzai [27] verwiesen, in deren Rahmen wesentliche Teile der Integration entstanden.

[8] Siehe Abschnitt 4.2

3.3. TLS Implementierung in Bouncy Castle

Aufgrund der zum Durchführungszeitpunkt immer noch sehr weiten Verbreitung von TLS 1.2 (vgl. [10]) sowie der Verfügbarkeit einer TLS 1.2 Implementierung in Bouncy Castle (Version 1.66) wurde die Integration der eUCRITE-API in diese TLS Implementierung untersucht. Wesentliche Teile der Integration sind im Rahmen der Bachelorarbeit von Merz [20] entstanden.

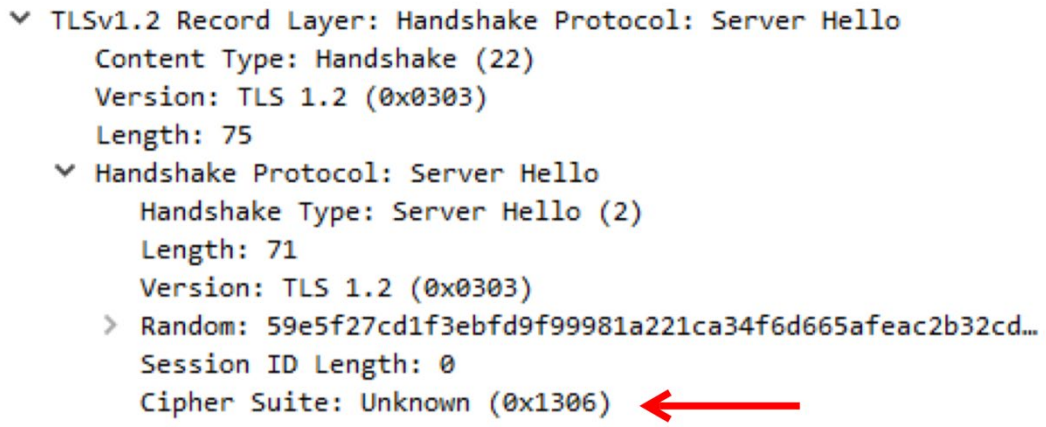


Abb. 3: TLS `Server Hello` mit neuer, für Wireshark unbekannter Cipher Suite, Quelle [22]

Wie in Kapitel 1. erwähnt, lag der Fokus dieser Arbeit auf dem Austausch der asymmetrischen Krypto-Verfahren in einem hybriden Kryptosystem wie TLS. Da zum aktuellen Zeitpunkt seitens der IANA noch keine passende TLS Cipher Suite mit PQC-Verfahren definiert ist [5], wurde für die Implementierung der Wert `0x1306` verwendet, der noch nicht von der IANA vergeben wurde[9] (siehe Listing 4).

```
1 public static final int TLS_CME_SPX_WITH_AES_25_CBC_SHA512 = 0x1306;
```

List. 4: Verwendete Cipher Suite Definition

Der PQC-basierte TLS Handshake lief erfolgreich zwischen zwei Knoten ab, die beide die neue Cipher Suite verarbeiten konnten. Die Analyse des Datenverkehrs mithilfe von Wireshark meldet erwartungsgemäß an dieser Stelle eine unbekannte Cipher Suite (siehe Abbildung 3).

4. Erkenntnisse

Im Folgenden wird auf die wesentlichen Hürden eingegangen, die bei der Integration der PQC-Verfahren in hybride Krytosysteme auftraten. Diese lassen sich in konzeptionelle/organisatorische (Abschnitt 4.1) und technische Aspekte (Abschnitt 4.2) unterscheiden.

4.1. Konzeptionelle/organisatorische Aspekte

Die eUCRITE-API bietet bei der Wahl der Krypto-Verfahren eine höhere Abstraktion als herkömmliche Bibliotheken. Ein Entwickler spezifiziert zur Compile-Zeit nur noch

9 Status *unassigned*

sein gewünschtes Sicherheitslevel (`LOW`, `MEDIUM`, `HIGH`). Intern wählt eUCRITE dann die passenden Krypto-Verfahren aus. Beispielsweise könnte für `LOW` noch RSA ausreichend sein, während `HIGH` bereits auf das Quantencomputer-resistente McEliece-Verfahren zurückgreift. Diese genau spezifizierte Angabe zur Wahl des Krypto-Verfahrens wird dann z.B. im TLS- Handshake verwendet. Hier offenbart sich eine semantische Unschärfe, sollten eUCRITE- basierte Kommunikationspartner zu unterschiedlichen Compile-Zeiten unterschiedliche Verfahren für ein Sicherheitslevel ausgewählt haben. Ein Sicherheitslevel `HIGH` im Jahr 2021 könnte zur Wahl eines anderes Verfahrens führen als ein Sicherheitslevel `HIGH` im Jahr 2031. Es muss also eine Möglichkeit geschaffen werden (z.B. über einen Update Mechanismus der eUCRITE-API oder durch Versionsnummern der eUCRITE-API) dies zu erkennen und programmtechnisch zu behandeln.

Zum jetzigen Zeitpunkt fehlen Spezifikationen für Cipher Suites, die PQC-Verfahren unterstützen. Es gibt erste Ansätze[10], die jedoch nicht alle Kombinationen von Verfahren abdecken. So haben wir `TLS_CME_SPX_WITH_AES_256_CBC_SHA512` für unsere Umsetzung gewählt. In Folge einer Standardisierung könnte dieser Wert ungültig werden.

4.2. Technische Aspekte

Die Analyse des Netzwerkverkehrs unserer prototypischen Implementierungen zeigt, dass bei unserem TLS Handshake, basierend auf McEliece, ca. 1.5 MB an Nutzdaten bei der Übermittlung des Zertifikats anfallen. Diese Beobachtung deckt sich mit anderen Arbeiten [25], [17], [7] und könnte sich negativ auf die User-Experience beim *Surfen* im Internet auswirken, da der TLS Handshake zu viel Zeit konsumiert, insbesondere bei sog. *lossy networks* verschärft sich dieses Problem [22].

Unsere Wahl der Programmiersprache Java für die eUCRITE-API und die E-Mail Anwendung InboxPager sowie die Bibliothek Bouncy Castle zog die folgenden technischen Anpassungen nach sich: Die Implementierung der PQC-Verfahren wurde uns in Form einer Erweiterung der Bouncy-Castle-Bibliothek durch unseren Projektpartner, der MTG AG aus Darmstadt [3], bereitgestellt, die aus Gründen der Performance über das Java Native Interface die eigentlichen Algorithmen nah an der Ziel-Hardware in der Programmiersprache C umsetzt und auf Prozessorarchitektur-Optimierungen setzt. Dies limitiert die Portabilität unserer Lösungen.

Weiter ist es aufgrund der Größe des McEliece Schlüsselmaterials notwendig, die Größe des JVM Thread Stack zu erhöhen, da das Schlüsselmaterial über den Stack an die native C-Anbindung (via Java Native Interface) übergeben wird.

Schließlich sei erwähnt, dass das Android OS bereits mit einer Bouncy-Castle-Bibliothek als Java Cryptographic Service Provider ausgeliefert wird. Dieser Provider muss zunächst deaktiviert werden, da es ansonsten zur Auswahl des falschen Providers

[10] z.B. https://tools.ietf.org/html/draft-campagna-tls-bike-sike-hybrid-01 (besucht am 29.12.2020)

kommt. Weiter wird Google für Android in Zukunft obligatorisch auf Conscrypt[11] als Implementierung für die klassischen Krypto-Algorithmen setzen, was für die Integration von PQC-Verfahren Anpassungen beim Zusammenspiel mehrerer Provider erfordern wird.

5. Fazit und Ausblick

5.1. Fazit

Die vorliegende Arbeit berichtet über praktische Erkenntnisse bei der Integration von PQC-Verfahren in bestehende Softwareprodukte. Die Integration erfordert zum Teil detaillierte Kenntnisse bzw. Änderungen an der Laufzeitumgebung, wie der beobachtete Thread-Stack-Fehler zeigt. Um die Interoperabilität zwischen (beliebigen) Anwendungen sicherstellen zu können, bedarf es weiterer Standardisierung, z.B. bezüglich der zu verwendenden Konstanten für Cipher Suites. Der Punkt Endbenutzerfreundlichkeit bedarf ebenfalls der Aufmerksamkeit. So müssen z.B. die PQC-Algorithmen hardwarenah ausgeführt werden, um eine akzeptable Wartezeit zu erreichen. Obwohl die Integration von PQC in TLS-1.2 technisch erfolgreich war, ist es anzuraten hier auf modernere Protokolle (z.B. TLS-1.3 oder Google QUIC) zu setzen, da diese den Overhead zum Aufbau einer Sitzung, und damit die Ausführung von PQC-Verfahren deutlich reduzieren werden.

5.2. Ausblick

Weiterführend soll das korrekte Zusammenwirken der hier vorgestellten Abstraktionsmechanismen über System-, und Zeitgrenzen untersucht werden. Hierfür wird, insbesondere für die zeitliche Komponente, der erwähnte Update-Mechanismus von Bedeutung sein. Neben der Durchführung von weiteren Integrationen und Tests sollen die erzielten Ergebnisse in das Design und die Funktionalität der eUCRITE-API zurückfließen.

Danksagung

Dieser Beitrag wurde im Rahmen der Innovationsförderung des Landes Hessen aus Mitteln der LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz, Förderlinie 3: KMU-Verbundvorhaben unter HA-Projekt-Nr.: 633/18-56 gefördert.

[11] https://conscrypt.org (besucht am 29.12.2020)

Literaturhinweise

[1] Marcel Bokhorst (M66B). *FairEmail - Open source, privacy friendly email app for Android*. https://email.faircode.eu (besucht am 28.12.2020).

[2] Yasemin Acar u. a. „*Comparing the Usability of Cryptographic APIs*“. In: 2017 IEEE Symposium on Security and Privacy (SP). 2017, S. 154–171.

[3] MTG AG. MTG AG - IT-Security für kritische Infrastrukturen. https://www.mtg.de/de/start/index.html (besucht am 28.12.2020).

[4] F. Armknecht u. a., Hrsg. *Biggest Failures in Security*. Bd. 9. Dagstuhl Reports 11. Dagstuhl Publishing, 2019, S. 1–23.

[5] Internet Assigned Numbers Authority. *Transport Layer Security (TLS) Parameters*. https://www.iana.org/assignments/tls-parameters/tls-parameters.xhtml (besucht am 28.12.2020).

[6] Daniel J Bernstein u. a. „*SPHINCS: practical stateless hash-based signatures*“. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer. 2015, S. 368–397.

[7] Kevin Bürstinghaus-Steinbach u.a. *Post-Quantum TLS on Embedded Systems*. Techn. Ber. 308. 2020. https://eprint.iacr.org/2020/308 (besucht am 03.04.2020).

[8] Matthew Campagna u. a. „*Quantum Safe Cryptography and Security: An introduction, benefits, enablers and challenges*“. In: European Telecommunications Standards Institute ETSI White Paper.8 (Juni 2015), S. 1–64. https://www.etsi.org/images/files/ETSIWhitePapers/QuantumSafe-Whitepaper.pdf.

[9] Lily Chen u. a. *Report on Post-Quantum Cryptography*. US Department of Commerce, National Institute of Standards und Technology, 2016. doi: 10.6028/NIST.IR.8105.

[10] Eric Crockett, Christian Paquin und Douglas Stebila. „*Prototyping post-quantum and hybrid key exchange and authentication in TLS and SSH*“. en. In: NIST, 2019.

[11] Lov K Grover. „*Quantum mechanics helps in searching for a needle in a haystack*“. In: Physical review letters 79.2 (1997), S. 325.

[12] User-Centered Security Working Group (CS Dept. h_da). *eUCRITE API*. https://use-a-pqclib.h-da.io/eucrite-documentation/ (besucht am 28.12. 2020).

[13] User-Centered Security Working Group (CS Dept. h_da). *Projekt - Use-A-PQClib*. http://fbi.h-da.de/ucs (besucht am 28.12.2020).

[14] Udyani S Herath Mudiyanselage. „*Next-Generation Web Public-Key Infrastructure Technologies*“. PhD. Queensland University of Technology, 2019. doi: 10.5204/ thesis.eprints.128643. https://eprints.qut.edu.au/128643 (besucht am 02.04.2020).

[15] Rolf Huesmann u.a. „*Zur Benutzbarkeit und Verwendung von API- Dokumentationen*“. In: Mensch und Computer 2020 - Workshopband. Hrsg. von Christian Hansen, Andreas Nürnberger und Bernhard Preim. Bonn: Gesellschaft für Informatik e.V., 2020. doi: 10.18420/muc2020-ws119-002.

[16] GitHub user: itprojects. *An E-mail client for the Android platform. Imap, pop, and smtp via SSL/TLS, with AES/PGP support*. https://github.com/itprojects/InboxPager (besucht am 28.12.2020).

[17] Panos Kampanakis u. a. *The Viability of Post-quantum X.509 Certificates*. Techn. Ber. 063. 2018. http://eprint.iacr.org/2018/063 (besucht am 10.03. 2020).

[18] Tyson Macaulay und Richard Henderson. *Cryptographic Agility in Practice*. White Paper. InfoSec Global, 2019.

[19] Robert J McEliece. „*A Public-Key Cryptosystem Based on Algebraic Coding Theory*“. In: Coding Thv 4244 (1978), S. 114–116.

[20] Matthias Merz. „*Integration von Post Quantum Verfahren in Bouncy Castle als Grundlage für eine Evaluation der Usability der eUCRITE API*“. Bachelorarbeit. Hochschule Darmstadt, 2020.

[21] David Ott und Christopher Peikert. „*Identifying Research Challenges in Post Quantum Cryptography Migration and Cryptographic Agility*“. In: arXiv:1909.07353 [cs] (Sep. 2019). arXiv: 1909.07353. http://arxiv.org/abs/1909.07353 (besucht am 06.02.2020).

[22] Christian Paquin, Douglas Stebila und Goutam Tamvada. *Benchmarking Post-Quantum Cryptography in TLS*. Techn. Ber. 1447. 2019. http://eprint.iacr.org/2019/1447 (besucht am 18.09.2020).

[23] Peter W. Shor. „*Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer*“. In: SIAM J. Comput. 26.5 (Okt. 1997), S. 1484–1509. issn: 0097-5397.

[24] Bundesamt für Sicherheit in der Informationstechnik. *Migration zu Post-Quanten-Kryptografie*. 2020. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Krypto/Post-Quanten-Kryptografie.html (besucht am 05.10.2020).

[25] Dimitrios Sikeridis, Panos Kampanakis und Michael Devetsikiotis. „*Post-Quantum Authentication in TLS 1.3: A Performance Study*“. In: Proceedings 2020 Network and Distributed System Security Symposium. Internet Society, 2020.

[26] K9 Team. *K9 Mail*. https://k9mail.app (besucht am 28.12.2020).

[27] Basset Yousofzai. „*Integration von Post-Quanten-Kryptographie Verfahren in einen Android Open-Source Java E-Mail-Client*“. Bachelorarbeit. Hochschule Darmstadt, 2020.

[28] Alexander Zeier, Alexander Wiesmaier und Andreas Heinemann. „*API Usability of Stateful Signature Schemes*“. In: International Workshop on Security. Springer. 2019, S. 221–240.

Zurück zum Inhaltsverzeichnis