
CISPA
DISPLAY
DE

VORWORT

Das Streben nach wissenschaftlicher Exzellenz gehört zur DNA des CISPA Helmholtz-Zentrum für Informationssicherheit. Diese mit unseren Forschungsleistungen auch zu erreichen, liegt in den Händen der mittlerweile 39 Faculty, den leitenden Wissenschaftler:innen am CISPA, sowie einem stetig wachsenden Kreis an Doktorand:innen, Post-Doktorand:innen und Forschungsgruppenleiter:innen. Organisiert in sechs Forschungsbereichen, decken die Forschenden eine große Bandbreite von Themenfeldern zwischen vertrauenswürdiger Informationsverarbeitung, Kryptographie, sicheren vernetzten und mobilen Systemen und vertrauenswürdiger künstlicher Intelligenz ab. Ihnen ideale Arbeitsbedingungen für ihre Forschung zu bieten, ist das Ziel des Zentrums. Ein Blick auf die CSRankings – ein metrikbasiertes Ranking der besten Informatik-Institutionen der Welt – zeigt, dass dies von Erfolg gekrönt ist: Das CISPA belegt dort im Bereich Computer Security regelmäßig, wie auch im abgelaufenen Wissenschaftsjahr, den ersten Platz.

Hochspezialisiertes Wissen innerhalb der Community

Wissenschaftliches Arbeiten stellt eine hochspezialisierte Tätigkeit dar. Die Forschenden am CISPA kombinieren innovative anwendungsorientierte Forschung mit hochmoderner Grundlagenforschung. Geteilt und diskutiert werden die Ergebnisse der Forschung vor allem innerhalb der Wissenschaftscommunity, insbesondere auf den großen, jährlich stattfindenden Fachkonferenzen der jeweiligen Forschungsbereiche, wie etwa dem USENIX Security Symposium, der ACM Conference on Computer and Communications Security (CCS) oder der International Conference on Machine Learning (ICML). Um auf einer dieser Konferenzen vortragen zu können, müssen die Forschenden im Vorhinein längere wissenschaftliche Aufsätze einreichen, die dann von einer Fachjury begutachtet werden. Diese Aufsätze laufen auch unter dem englischen Begriff „Paper“. Das Besondere an der IT-Forschung ist, dass die Aufsätze üblicherweise über die Konferenzen publiziert werden und nicht etwa in Fachzeitschriften, wie in anderen Wissenschaftsdisziplinen üblich.

Wissenstransfer in die Öffentlichkeit

Damit die über die Konferenzen publizierten Forschungsergebnisse auch einer breiten Öffentlichkeit zugänglich werden, veröffentlicht die Abteilung Corporate Communications des CISPA sogenannte Forschungstexte. Jeweils zu einem wissenschaftlichen Aufsatz werden kurze, allgemein verständliche

Mit der Publikation CISPA Display bekommt die Wissenschaftskommunikation am Zentrum ein neues Format. Ziel ist einen Teil der exzellenten Forschung am CISPA einer breiten Öffentlichkeit zugänglich zu machen.

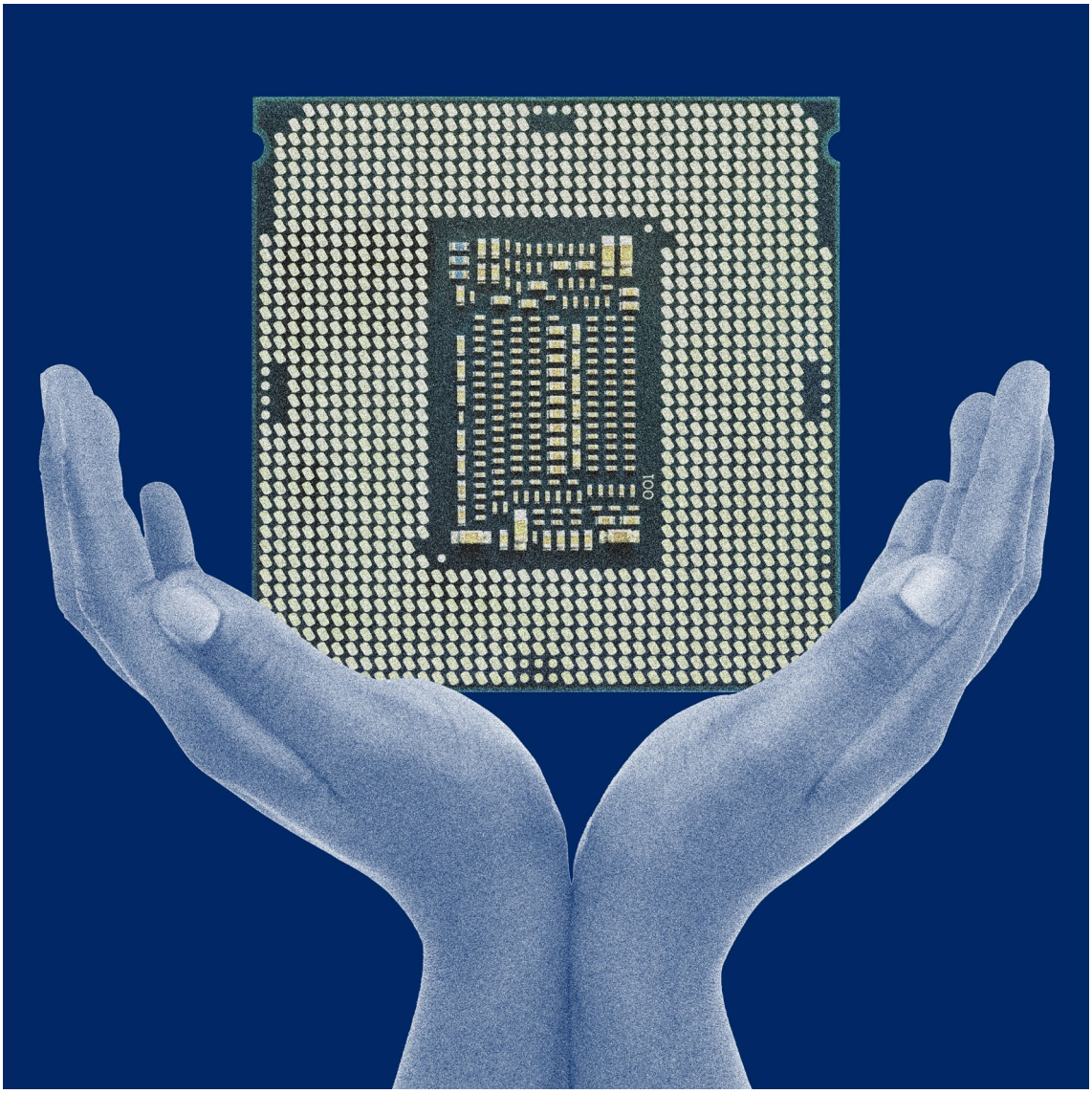
Zusammenfassungen geschrieben und auf unserer Website publiziert. Sie basieren auf den wissenschaftlichen Aufsätzen selbst, sowie Interviews mit den Autor:innen. Da wissenschaftliche Aufsätze in der Regel von mehreren Autor:innen verfasst werden, finden diese Interviews meist mit dem/der Erstautor:in oder bei Kooperationen zwischen Institutionen mit den Beteiligten aus dem CISPA statt. Zu jedem dieser Texte entwerfen unsere Kommunikationsdesignerinnen zudem grafische Illustrationen, die die Kernaussage oder das Themenfeld der wissenschaftlichen Aufsätze visualisieren und damit eine ganz eigene Übersetzungsarbeit leisten. So gelingt über die Texte und Grafiken ein Wissenstransfer der Forschungsergebnisse in die breite Gesellschaft hinein.

Ein neues Druckprodukt für das CISPA

Aus der Bedeutung der Paper-Texte und den dazugehörigen Grafiken für die Wissenschaftskommunikation am CISPA ist die Idee entstanden, ihnen eine eigene Publikation zu widmen. Das war die Geburtsstunde von CISPA Display. Die Publikation vereint alle im Jahr 2023 publizierten Forschungstexte. Vorgestellt werden insgesamt 18 wissenschaftliche Aufsätze von CISPA-Forschenden. Die Texte zeigen die große Vielfalt der Forschungsthemen, die am CISPA verfolgt werden. Die thematische Spannbreite reicht von Schlüsselmanagement bei Kryptowährungen über Satellitensicherheit und Authentifizierungsmechanismen in Messengerdiensten bis hin zu Verfahren zum Schutz vor Deepfakes. Mit CISPA Display wird das Portfolio unserer regelmäßigen Druckpublikationen nach dem bereits seit 2022 existierenden CISPA Zine um ein weiteres Produkt erweitert. Wir wünschen viel Spaß beim Lesen!

INDEX

INDEX



© Janine Wichmann-Paulus

In seinem Paper „Automatic Detection of Speculative Execution Combinations“ stellt PhD-Student und CISPA-Forscher Xaver Fabian einen neuen Ansatz vor, mit dem automatisch Schwachstellen im Prozessor gefunden werden können – auch wenn diese erst durch die Kombination mehrerer spekulativer Mechanismen entstehen. Die Conference on Computer and Communications Security (CCS) hat Fabian dafür mit einem Distinguished Paper Award ausgezeichnet.

Neuer Ansatz verbessert automatisierte Schwachstellensuche in Prozessoren

Xaver Fabian

Der Prozessor wird oft als Herz des Computers bezeichnet, dabei ist er viel eher dessen Gehirn. Denn der Prozessor steuert und interpretiert Befehle, koordiniert Abläufe und sorgt dafür, dass Aufgaben an die entsprechenden Stellen weitergeleitet werden. Moderne Prozessoren tun das in enormer Geschwindigkeit und erledigen dabei viele Aufgaben parallel. Das schaffen sie unter anderem durch einen Trick, der ihnen hilft, die Prozessor-Ressourcen sinnvoll zu nutzen: die spekulative Ausführung. Fabian erklärt, was das bedeutet: „In Phasen, in denen der Prozessor nicht voll ausgelastet ist, versucht er mithilfe verschiedener Mechanismen vorauszusehen, welcher Programmschritt als nächstes folgen könnte. Er führt dann die dafür nötigen Berechnungen aus und speichert die Ergebnisse in einen Zwischenspeicher. Werden die Daten doch nicht gebraucht, verwirft er sie. Allerdings hat sich gezeigt, dass die verworfenen Daten eine Spur im Cache-Speicher hinterlassen und dort unter Umständen von Angreifer:innen ausgelesen werden können.“ Die erste Schwachstelle, die Angriffe dieser Art zulässt, ist 2018 unter dem Namen Spectre bekannt geworden und hat hohe Wellen geschlagen. Forschende finden seither regelmäßig neue Spectre-Varianten.

»Eine Möglichkeit, eine Sicherheitsgarantie abgeben zu können, ist ein formales Modell zu entwickeln.«

Methoden zum Schließen von Spectre-Lücken

Nachdem die ersten Spectre-Lücken bekannt wurden, wurden ad hoc Maßnahmen entwickelt, um sie zu schließen. Aber wer sagt eigentlich, dass diese Maßnahmen dann auch wirklich immer funktionieren? „Eine Möglichkeit, eine Sicherheitsgarantie dafür abgeben zu können, ist ein formales Modell zu entwickeln. Damit wird eine mathematische Analyse möglich und so können wir die Wirksamkeit der Maßnahmen auch belegen“, erklärt Fabian. Forschende erstellen solche Modelle mithilfe von logisch-mathematischen Methoden und speziellen formalen Sprachen. „Bislang haben sich viele Forschende dabei nur auf bestimmte Varianten von Spectre-Lücken und damit nur auf bestimmte Spekulationsmechanismen im Prozessor konzentriert und diese isoliert betrachtet. Ehrlich gesagt, weil auch so schon alles kompliziert genug ist. In Wahrheit laufen im Prozessor aber ja mehrere Spekulationsmechanismen gleichzeitig ab. Wir haben versucht, ein Modell zu schaffen, das es erlaubt, diese Mechanismen zu kombinieren. Dahin zu kommen, war nicht einfach. Allein 200 Seiten mathematischer Beweisführung liegen irgendwo auf meinem Schreibtisch.“

Zunächst hat Fabian zwei bislang unbeachtete Spekulationsmechanismen in einer formalen Sprache ausgedrückt und sie somit der mathematischen Beweisführung zugänglich gemacht. Dann hat er diese sogenannten Semantiken mit einer bereits existierenden für einen anderen Mechanismus kombiniert. „Die Grundlage für meine Arbeit bietet die Vorarbeit von IMDEA-Forscher Dr. Marco Guarnieri und anderen. Zum einen habe ich eine von ihnen entwickelte Assembler-Sprache zur Formalisierung genutzt. Zum anderen konnte ich die beiden von mir neu entwickelten Semantiken direkt in das bereits bestehende Tool namens SPECTECTOR einbauen und so testen.“ Und seine Arbeit hat sich ausgezahlt.

»Wir machen hier absolute Grundlagenarbeit.«

Grundstein für zukünftige komplexe Analysen

Für eine umfangreiche Testung, wie robust Prozessoren gegen Spectre-Attacken sind, reicht SPECTECTOR dennoch nicht aus. „Das Modell ist noch ziemlich vereinfacht. Moderne Prozessoren sind sehr komplex und können ganz viel. Die Modellierungsansätze in der Forschung hängen da noch ziemlich hinterher." Fabians Arbeit bildet aber den Grundstein für eine weitaus komplexere Analyse als sie bislang möglich war. „Wir machen hier absolute Grundlagenarbeit. Und die braucht es auch. IT-Sicherheit wurde lange zu wenig Beachtung geschenkt. Aber ich habe den Eindruck, dass sich da in den vergangenen Jahren etwas verändert hat." Über seinen Distinguished Paper Award, mit dem herausragende Forschungsarbeiten ausgezeichnet werden, freut sich der gebürtige Schleswig-Holsteiner sehr. „Ich war ehrlich gesagt total überrascht. Es ist toll, dass meine Arbeit in der Community geschätzt wird."

Fabian, Xaver; Guarnieri, Marco; Patrignani, Marco (2022) Automatic Detection of Speculative Execution Combinations. In: CCS 2022, 7-11 Nov 2022, Los Angeles, CA, USA. Conference: CCS ACM Conference on Computer and Communications Security

© Lea Mosbach

Visuelle digitale Zertifikate könnten ein sicherer und datenschutzfreundlicher Ansatz sein, um offizielle Dokumente wie zum Beispiel den Führerschein digital verfügbar zu machen, sagt CISPA-Forscher Dañiel Gerhardt. „Aber nur, wenn sie auch korrekt auf ihre Richtigkeit geprüft werden." Großflächig eingesetzt wurden solche Zertifikate während der Corona-Pandemie in Form der Impfzertifikate, die EU-weit gültig waren. Im Paper „Investigating Verification Behavior and Perceptions of Visual Digital Certificates" haben Gerhardt und sein CISPA-Kollege Alexander Ponticello am Beispiel der Impfzertifikate untersucht, warum Menschen oftmals bei der Verifizierung digitaler Zertifikate Fehler machen und wie das in künftigen Einsatzfällen vermieden werden kann. Ihre Arbeit haben die beiden Forscher auf dem renommierten USENIX Security Symposium 2023 vorgestellt.

Warum visuelle digitale Zertifikate bislang nur theoretisch sicher sind

Dañiel Gerhardt

Um Informationen visuell codiert weiterzugeben, kommen schon seit langer Zeit Bar- und QR-Codes zum Einsatz. Im Alltag begegnen sie uns zum Beispiel auf Produkten im Supermarkt, auf Paketen oder Konzert-Tickets. „Die Datenmenge, die in dieser Form codiert werden kann, ist allerdings begrenzt. Oft steckt hinter den Codes deshalb auch nur der Link auf eine externe Quelle, wie eine Website oder ähnliches“, erklärt Gerhardt. Auch die Covid-Impfzertifikate, die während der Corona-Pandemie die „Eintrittskarte“ zu Restaurants und anderen öffentlichen Orten waren, trugen einen QR-Code. Dahinter steckte aber nicht nur ein Link, sondern kryptografisch signierte Daten, die den Impfstatus einer Person nachweisen konnten. Dazu schickten Impfstellen persönliche Daten wie Namen und Geburtsdatum einer Person, ein persönliches Erkennungsmerkmal sowie das Impfdatum und die Info über den verwendeten Impfstoff ans Robert-Koch-Institut (RKI). Das RKI wiederum versah diese Informationen mit einer digitalen Signatur und stellte ein entsprechendes Zertifikat aus. Das konnten sich die geimpften Personen dann unter Vorlage ihres Impf- und Personalausweises in Apotheken oder bei Ärzt:innen abholen. Das Zertifikat gab es digital oder auf Papier. Der QR-Code auf dem Zertifikat ließ sich mit Apps wie der Corona-Warn-App einscannen und so mit sich tragen. Das RKI löschte die erhobenen Daten nach Erstellen der Signatur wieder. „Dass die Daten in Deutschland nicht zentral, sondern nur lokal gespeichert wurden, machte das Verfahren sehr datenschutzfreundlich. Zudem ist es sicherer gegen Fälschungen, nachhaltig und kosteneffizient, da keine Behörde fälschungssichere Ausdrucke herstellen muss “, erklärt Gerhardt.

Der Faktor Mensch bei der Kontrolle

In der Praxis wurde der 25-Jährige allerdings auf ein Problem aufmerksam: „Wenn ich zum Beispiel in ein Restaurant gehen wollte, habe ich einige Male erlebt, dass Mitarbeitende, statt den QR-Code zu scannen und meinen Ausweis zu prüfen, nur einen Blick auf den Code in einer App geworfen und mich dann reingelassen haben. Das ist natürlich keine sinnvolle Prüfung. Andere benutzten zwar eine zur Verifizierung nötige Scan-App und scannten auch den QR-Code, kontrollierten aber zum Beispiel meinen Ausweis nicht.“ Da ein funktionierender

Prüfprozess der Zertifikate für deren Sicherheit entscheidend ist, hat sich Gerhardt in einer qualitativen Interview-Studie damit auseinandergesetzt, warum in der Praxis so viele Fehler passierten. Er hat 17 Menschen, die auf ihrer Arbeit mit der Prüfung der Zertifikate betraut waren, bei der Verifizierung von Covid-Zertifikaten beobachtet und später befragt. „Wir wollten so vor allem zwei Fragen klären: Wie verifizieren diese Personen die Zertifikate und warum machen sie es so? Außerdem haben wir untersucht, wie viel die prüfenden Personen über den Verifikationsprozess und seinen Ablauf wissen." Gerhardts qualitative Studie soll Ansätze liefern, um das Verhalten von Nutzer:innen besser zu verstehen und so die theoretischen Sicherheitsvorteile visueller digitaler Zertifikate auch in die echte Welt zu überführen.

Die Ergebnisse haben den CISPA-Forscher überrascht: „Die Studienteilnehmer:innen haben die Zertifikate ganz unterschiedlich geprüft. So viele Varianten hatte ich nicht erwartet." Einige Befragte haben bei der Prüfung alle nötigen Schritte korrekt ausgeführt: Sie scannten das Zertifikat mit einer entsprechenden Prüf-App ein, gleichten die in ihrer App angezeigten Daten mit dem Personalausweis der Person ab und prüften auch, ob deren Foto auch wirklich die Person zeigt, die vor ihnen stand. Andere Befragte führten ebenfalls all diese Schritte aus und darüber hinaus noch einige unnötige. „So versuchten einige Menschen, sich aufgrund von Äußerlichkeiten ein Bild von ihrem Gegenüber und dessen Vertrauenswürdigkeit zu machen. Andere wurden grundsätzlich misstrauisch, wenn ihnen ein Screenshot präsentiert wurde. Dabei ist das nicht wirklich ein Hinweis, dass etwas schiefläuft." Misstrauen zeigten einige Befragte auch, wenn ihnen das Zertifikat in einer App präsentiert wurde, die sie nicht kannten. Andere Studien-Teilnehmer:innen verließen sich bei der Beurteilung eines Impfzertifikats lieber auf das eigene Gefühl als auf die korrekte technische Überprüfung und scannten die Zertifikate nur von Zeit zu Zeit. Andere gaben an, grundsätzlich nur zu schauen, ob ein QR-Code vorhanden ist und scannten die Zertifikate gar nicht.

Bessere Ausstattung und Aufklärung nötig

„Die Mehrheit der Studienteilnehmer:innen wusste nicht viel über den Verifikationsprozess und wie er technisch abläuft. Das hatte aber nicht unbedingt zur Folge, dass sie dabei Fehler machten", sagt Gerhardt. „Nur umgekehrt war es so, dass wer den Prozess gut verstanden hat, keine Fehler machte." Im Geschäftsleben sind laut dem Forscher oft andere Faktoren viel entscheidender dafür, wie der Prüfprozess abläuft: zum Beispiel wie zeitaufwendig er ist. „Außerdem haben uns einige Teilnehmer:innen berichtet, dass ihnen der Arbeitgeber kein Gerät zum Scannen zur Verfügung gestellt hat und

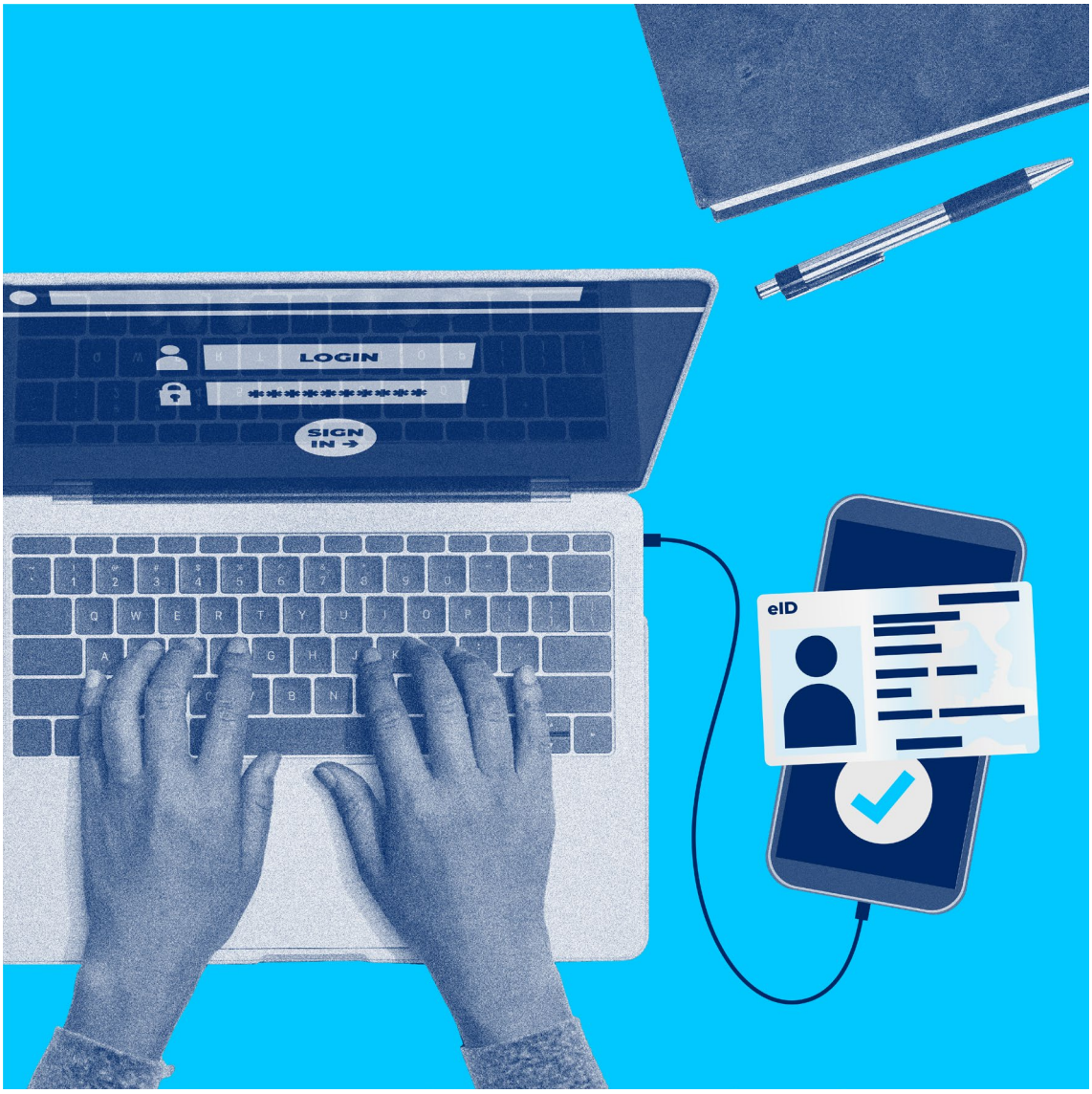
sie dazu nicht ihr privates Smartphone nutzen wollten. Andere wussten wiederum nicht, dass sie die Prüf-App einfach im App-Store herunterladen können. Sie dachten, die App stehe nur offiziellen Stellen zur Verfügung."
Zu all diesen Missverständnissen kommt es laut Gerhardt auch deshalb, weil viele Befragte keine Informationen von offiziellen Stellen erhielten, wie sie die Zertifikate prüfen sollen. Eine bessere Kommunikation und einheitliche Aufklärung für Personen, die den Verifikationsprozess durchführen sollen, ist laut Gerhardt eine wichtige Voraussetzung, um die Technik künftig sicher einsetzen zu können. „Der Gesetzgeber könnte das mit einer gesetzlichen Verpflichtung zur Einhaltung der Vorschriften unterstützen." Wichtig sei auch die Prüfpersonen mit geeigneten Prüfgeräten und Software auszustatten. „Außerdem müssen sie wissen, wie sie damit umgehen sollen, wenn Zertifikate der Prüfung nicht standhalten."

Klareres Design für weniger Missverständnisse

Ein letzter wichtiger Punkt ist laut Gerhardt das Design der Nachweis-Apps. Einige Befragte haben sich durch bestimmte Angaben oder auch durch die Farbgebung der App täuschen lassen. Zum Beispiel haben einige ein Zertifikat als sicher eingestuft, sobald der QR-Code blau umrandet war. Andere verleitete die zusätzlich zum Zertifikat gemachte Angabe „3 von 3" zur Zahl der Impfvorgänge dazu, den Prüfprozess nicht korrekt durchzuführen. Beim Design künftiger Apps sollte laut Gerhardt deshalb darauf geachtet werden, dass nicht unbedacht falsche Signale gesendet werden. „Wenn wir durch solche und ähnliche Maßnahmen die Verifizierung verbessern und die visuellen Zertifikate richtig implementiert werden, gibt es dafür einige sinnvolle Einsatzgebiete. Neben dem digitalen Führerschein zum Beispiel auch E-Rezepte. Sie könnten von Ärzten und Ärztinnen digital signiert und sicher ausgestellt werden", sagt Gerhardt.

Der gebürtige Baden-Württemberger freut sich, dass sein Paper auf dem renommierten USENIX Security Symposium angenommen wurde. „Das Thema war schon Teil meiner Bachelorarbeit. Zusammen mit den CISPA-Forschenden Alexander Ponticello, Adrian Dabrowski und Katharina Krombholz habe ich es für die Konferenz zu einem vollständigen Paper ausgearbeitet", sagt Gerhardt. Der 25-Jährige ist an der Graduate School of Computer Science der Universität des Saarlandes eingeschrieben und will nach der Vorbereitungsphase in ein von Krombholz betreutes PhD-Studium starten. „Ich finde die Themen im Bereich Usable Security sehr spannend und fand die Zusammenarbeit mit den Forschenden super."

Gerhardt, Dañiel; Ponticello, Alexander; Dabrowski, Adrian; Krombholz, Katharina (2023) Investigating Verification Behavior and Perceptions of Visual Digital Certificates. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium



© Janine Wichmann-Paulus

Für den Login bei Webservices ist die 2-Faktor-Authentifizierung zum Standard geworden. Während viele Nutzer:innen eine Kombination aus Passwort und Handy-Code nutzen, gilt als sicherste Variante bisher der FIDO2-Standard, für den jedoch eine zusätzliche Hardware-Komponente nötig ist. CISPA-Forscher Fabian Schwarz und seine Kollegen aus den Teams von CISPA-Faculty Prof. Dr. Christian Rossow und CISPA-Faculty Dr. Lucjan Hanzlik haben nun mit FeIDo ein neues Verfahren entwickelt, das ohne spezielle Nutzer-Hardware auskommt. Das dazugehörige Paper „FeIDo: Recoverable FIDO2 Tokens Using Electronic IDs“ haben sie im November 2022 auf der renommierten ACM Conference on Computer and Communications Security (CCS) vorgestellt.

Entwicklung eines Open-Source-Prototyps für die 2-Faktor-Authentifizierung

Fabian Schwarz

Es ist eine einfache Erkenntnis: Ohne Log-In stehen viele Bereiche des World Wide Web und insbesondere eine Vielzahl von Diensten, seien es Messenger, Informationsangebote oder Online-Banking, den Nutzer:innen nicht zur Verfügung. Mit jedem neuen Account geben Nutzer:innen aber Daten aus der Hand und müssen sich neue Passwörter merken. Dabei ist allgemein bekannt, dass Passwörter eine eher unsichere Variante des Logins sind, weshalb eine Vielzahl neuer Verfahren in Anwendung und Erprobung ist. An diesem Punkt setzten die Überlegungen von Fabian Schwarz an. „Wir wollten den Login-Prozess in Webservices für Nutzer:innen so einfach und gleichzeitig so sicher wie möglich machen“, erzählt er. Das Ziel von Schwarz und seinen Kollegen war es, bisherige Standards einer breiten Masse verfügbar und sicher nutzbar zu machen. In ihrem Fokus stand dabei der FIDO2-Standard zur 2-Faktor-Authentifizierung, der von der internationalen FIDO-Allianz entwickelt wurde. FIDO ist die Abkürzung für „Fast Identity Online“.

Die Besonderheit des FIDO2-Standards ist, dass er für die Authentifizierung auf zusätzliche Hardware-Komponenten zurückgreift. Das kann etwa ein sogenannter Security-Token in Form eines USB-Sticks sein, der darüber hinaus auch noch mit einem Fingerabdruckscanner gesichert sein kann, aber auch ein Smartphone mit neuesten Sicherheitsstandards. FIDO2 greift auf den W3C-Web-Authentication-Standard (WebAuthn) und das Client-to-Authenticate-Protocol (CTAP) der FIDO-Allianz zurück. Die Authentifizierung erfolgt mit einem privaten und einem öffentlichen Schlüssel, welche vom Security-Token generiert werden. Während die privaten Schlüssel nie den Security-Token verlassen, werden die öffentlichen Schlüssel auf den jeweiligen Servern der genutzten Webservices hinterlegt. Nutzer:innen verwenden die privaten Schlüssel um eine Authentifizierung zu beantragen, welche von den Webservices durch Nutzung der öffentlichen Schlüssel sicher überprüft und zugeordnet werden können.

Nachteile bisheriger Verfahren

Mit dem FIDO2-Standard können Passwörter durch den Einsatz von Hardware-gestützten Security-Tokens

ergänzt werden. Langfristig wird mit den FIDO2-Tokens, wie zum Beispiel dem YubiKey der Firma Yubico, das Ziel verfolgt, eine vollständig passwortlose Authentifizierung zu ermöglichen. Obwohl dies laut Schwarz eine begrüßenswerte Entwicklung ist, gibt es seiner Ansicht nach jedoch auch eine Reihe von Nachteilen. So gibt es etwa den Kostenfaktor, weil sich Nutzer:innen neue Hardware-Komponenten beschaffen müssen beispielsweise in Form von Hardware Security-Tokens oder Smartphones mit neuesten Sicherheitsstandards. Darüber hinaus wirkt sich der hohe Sicherheitsstandard negativ auf die Benutzerfreundlichkeit aus. Denn wenn der Security-Token, wie zum Beispiel der USB-Stick, mit den gespeicherten Login-Daten verloren geht, ist auch ein Login nicht mehr möglich, womit der Zugang zu den eigenen Online-Accounts blockiert ist. Verfahren zum Wiederherstellen des Zugangs existieren zwar, haben jedoch meist Sicherheitslücken oder verursachen zusätzliche Nutzerkosten, etwa durch die Vorabregistrierung eines Backup-Tokens, so Schwarz. Die Herausforderung für die Entwicklung eines neuen Verfahrens war, diese Nachteile zu umgehen.

Von FIDO2 zu FeIDo

Der Ausgangspunkt von Schwarz und seinen Kollegen ist eine einfache, aber umso überzeugendere Idee: die Nutzung von Objekten, wie ein Personalausweis und ein Handy, die fast alle Bürger:innen zur Verfügung haben. „Wir haben uns angeschaut, wie man elektronische Personalausweise oder Reisepässe für diesen Anwendungsfalls nutzen kann, ohne dass sensible Nutzerdaten, die in den Pässen enthalten sind, an die Betreiber:innen der Websites gehen“, erklärt Schwarz. Sie wollten sich den Umstand zu nutzen machen, dass moderne Handys über die NFC-Technik, also die kontaktlose Datenübertragung mittels Radiowellen, auch digitale Identitätsnachweise, sogenannte eIDs, auslesen können. Gebraucht wird nur noch ein NFC-fähiges Smartphone, worunter fast alle handelsüblichen Apple- und Android-Handys fallen, aber eben keine extra Hardware mehr. „Benötigt wird dann nur noch eine kleine Zwischen-App, die den Leseprozess durchführt und Daten an unseren speziell abgesicherten Service übermittelt“, so Schwarz weiter. Genau dies haben die Forscher als Prototyp umgesetzt und diesen dann erfolgreich verschiedenen theoretischen Sicherheitsüberprüfungen unterzogen.

Anonymer Log-In als erweitertes Anwendungsfeld

Schwarz sieht im FeIDo-Verfahren aber noch weitere Vorteile, die aus dem Arbeiten mit den Daten aus den eIDs resultieren. Entscheidend ist dabei, dass im FeIDo-Verfahren diese Daten zwar ausgelesen, aber nicht weitergegeben werden. Dies unterscheidet FeIDo von anderen Verfahren, die ebenfalls biometrische Daten zur Authentifizierung nutzen. Damit werden auch neue

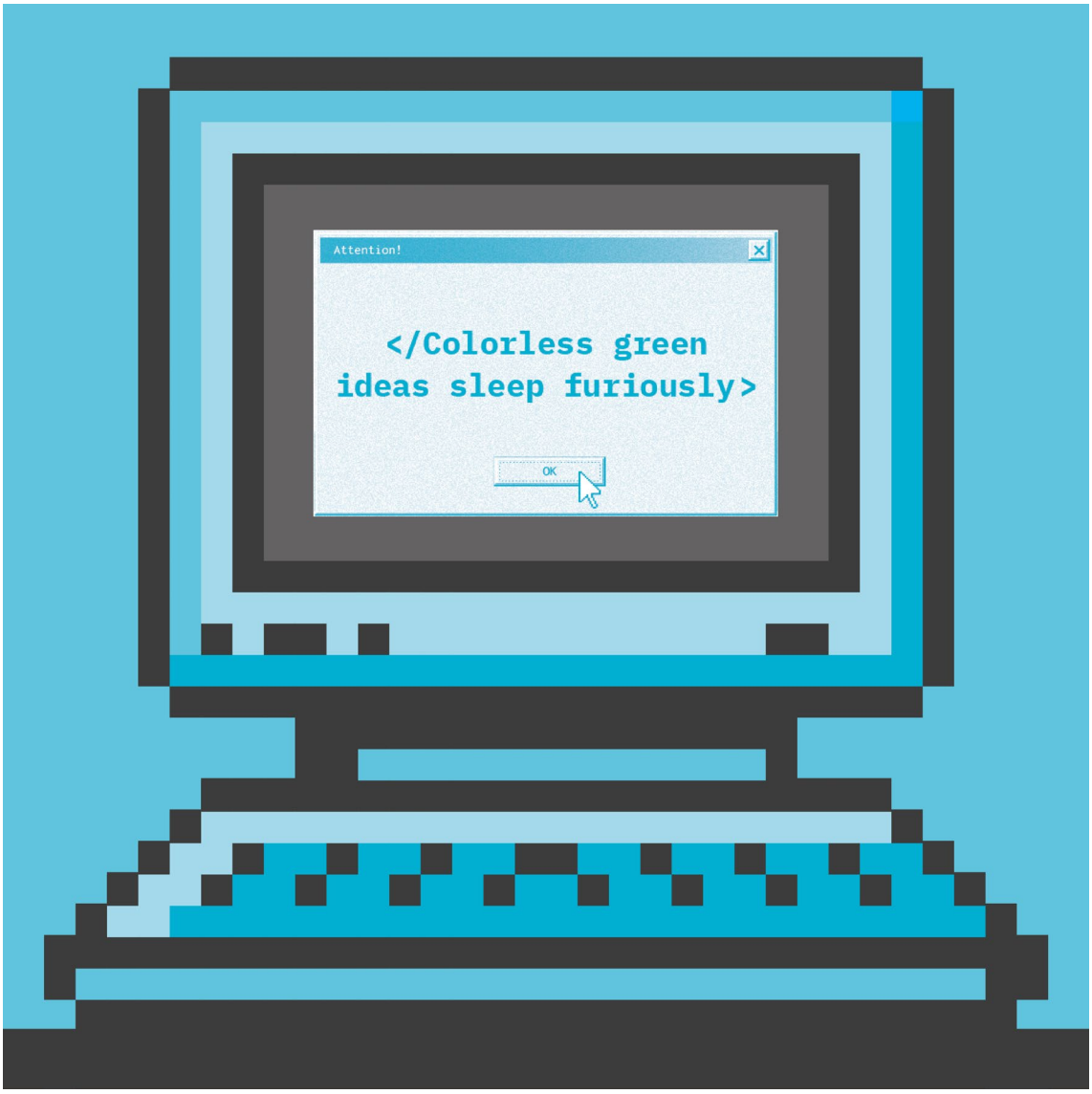
Anwendungsfelder für FeIDo denkbar, wie etwa das Überprüfen von Altersbeschränkungen beim Login bei speziell geschützten Webservices. „Wir können mit unserer App eine anonyme Anmeldung ermöglichen, bei der aber unser Dienst gleichzeitig einen Nachweis führt, dass der Nutzer volljährig ist“, erklärt Schwarz. Einschränkend fügt er hinzu, dass dafür jedoch Änderungen bei den Applikationen der Webservices nötig sind. „Dies wäre jedoch problemlos möglich“, so Schwarz weiter. Für Webservices ohne Zusätze wie Altersabfrage könnte das Log-In Verfahren des Prototyps der CISPA-Forscher direkt zum Einsatz kommen.

Gutes Feedback auf Konferenz CCS

Zum ersten Mal vorgestellt wurde das Paper auf der ACM Conference on Computer and Communications Security (CCS), die im November 2022 in Los Angeles stattfand. Das Interesse am Thema war dort sehr groß: „Es gab so viele Fragen, dass die Zeit gar nicht ausgereicht hat“, erzählt Schwarz. Auch in Reaktion darauf publizierten Schwarz und Kollegen ein Extended Paper. Schwarz selbst hat sich mittlerweile anderen Forschungsthemen zugewandt. Der Prototyp seiner Anwendung ist jedoch Open Source und damit frei verfügbar. „Das Ganze war von uns so aufgebaut, dass es möglichst frei als Community Project nutzbar ist“, erklärt er. Ziel ist, dass die Community solch einen Dienst kostenfrei zur Verfügung stellen kann. Schwarz würde sich freuen, wenn sich Kolleg:innen oder Unternehmen des Projektes annehmen und den Prototyp weiterentwickeln würden.

Schwarz, Fabian; Do, Khue; Heide, Gunnar Hanzlik, Lucjan; Rossow, Christian (2023) FeIDo: Recoverable FIDO2 Tokens Using Electronic IDs (Extended Version). *Technical Report. UNSPECIFIED.*

Forscher: ***Fabian Schwarz***
Autor: ***Felix Koltermann***



© Lea Mosbach

Die neue Spezifikationssprache ISLa kann nicht nur das automatisierte Testen von Software enorm verbessern. Sie könnte ein Meilenstein im Bereich der Softwaresicherheit und -zuverlässigkeit werden. Entwickelt wurde sie von CISPA-Forscher Dr. Dominic Steinhöfel. Er forscht im Team von CISPA-Faculty Prof. Dr. Andreas Zeller und hat ISLa in seinem Paper „Input Invariants" im November 2022 auf der renommierten European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE) erstmals vorgestellt. ISLa ist ein wichtiger Baustein für Zellers Forschung im Projekt „S3 – Semantics of Software Systems", in dem der Forscher zusammen mit seinem Team jede Software der Welt automatisch testbar machen will. Der ERC fördert Zeller und S3 mit einem ERC Advanced Grant in Höhe von 2,5 Millionen Euro.

Neue Spezifikationssprache revolutioniert automatisierte Softwaretests

Dominic Steinhöfel

Bei der Programmierung von Software passieren häufig Fehler. Damit die nicht zu ungewollten Abstürzen führen oder Sicherheitslücken eröffnen, testen Entwickler:innen Programme vor der Veröffentlichung oft mithilfe von sogenannten Fuzzern systematisch auf Schwachstellen. „Diese Tools produzieren massenhaft Zufallseingaben, um zu testen, wie sich Programme im Echtbetrieb schlagen. Eingaben zu produzieren, mit denen sich auch die tieferen Programmfunktionen testen lassen, ist allerdings schwierig“, erklärt Steinhöfel.

Woran liegt das? Die Datensprachen, die Computer sprechen und in denen mögliche Programm-Eingaben formuliert sind, ähneln in ihrem Grundaufbau den natürlichen Sprachen der Menschen. Und so spielt neben der Grammatik eben auch die Semantik, also die Bedeutung, eine Rolle. Den Unterschied hat der amerikanische Linguist Noam Chomsky in den 1950er-Jahren an einem Beispiel verdeutlicht: „Farblose grüne Ideen schlafen wütend.“ Die grammatikalische Struktur dieses Satzes ist einwandfrei, semantisch ist er allerdings inkorrekt. Sprich: Der Satzbau passt, auf der Bedeutungsebene ist das aber völliger Blödsinn.

Das Problem mit der Semantik-Ebene

Wenn ein Fuzzer also eine solche grammatikalisch zwar korrekte Eingabe produziert, sie aber keine für das zu testende Programm sinnvolle Aussage enthält, dann wird diese Eingabe direkt vom sogenannten Parser aussortiert. Der Parser ist ein Unter-Programm, das eingehende Eingaben in ihre Bestandteile zerlegt und prüft, ob sie für das Programm verständlich sind. Ist das der Fall, bereitet der Parser die Eingabe in ein zur Weiterverarbeitung geeignetes Format um. Wenn aber nicht, spuckt er eine Fehlermeldung aus und beschäftigt sich gar nicht weiter mit ihr. „Mit solchen Eingaben lässt sich also im Prinzip nur testen, wie gut der Parser ist, aber nicht, ob das Programm ansonsten stabil läuft“, sagt Steinhöfel. Es seien zwar durchaus Fuzzer im Einsatz, die klügere Eingaben produzieren, und so um den Parser herumkommen. „Oft geht es aber spätestens danach nicht mehr weiter, weil dann kompliziertere Eigenschaften auf der Semantik-Ebene folgen.“

Neue Spezifikationssprache als Antwort

Die von Steinhöfel entwickelte Spezifikationssprache ISLa kann hier zum Gamechanger werden. „Wir können mit ISLa Eingaben mit einer nie dagewesenen Präzision verstehen und die Programme damit tief und gründlich testen.“ Der Schlüssel liegt laut dem Forscher in einem sehr allgemeinen Formalismus, der Forschenden und Entwickler:innen nahezu jedes Programm zugänglich macht. „Allerdings brauchen wir dafür die Eingabebeschreibung. Die können wir manuell schreiben oder aus einem bestehenden Programm heraus lernen“, erklärt Steinhöfel. Das sei allerdings kompliziert und oft nur annäherungsweise möglich. „Es wird in der Praxis daher immer Programme geben, die zu groß oder zu kompliziert sind und sich nicht vollständig verstehen lassen. Aber wir können immer besser werden.“ ISLa ist dafür ein mächtiges Werkzeug und kann nicht nur Eingaben generieren, sondern diese auch prüfen, reparieren und mutieren. Zudem lassen sich damit laut Steinhöfel auch die Ausgaben des Programms beschreiben. „Wenn wir die Ein- und die Ausgaben beschreiben können, können wir das Verhalten des ganzen Programmes beschreiben. Und damit können wir wirklich viel tun: Wir können sagen, wie ein Programm sich verhalten soll, wir können analysieren, wie sich ein Programm verhält und wir können auch erzwingen, dass es sich verhält, wie wir es wollen. Kurzum: Wenn du die Ein- und Ausgaben kontrollierst, kontrollierst du die Programme.“ Auch Andreas Zeller, mit dem er eng zusammengearbeitet hat, unterstreicht, wie wichtig Steinhöfels Forschungsarbeit ist: „ISLa eröffnet ganze Welten für das Testen von Systemen.“

»Wir können mit ISLa Eingaben mit einer nie dagewesenen Präzision verstehen und die Programme damit tief und gründlich testen.«

Von der Theorie in die Praxis

In der näheren Zukunft wird Steinhöfel sich damit beschäftigen, auf dem durch ISLa bereitgestellten Fundament praxisnahe Ansätze zum Testen relevanter Softwaresysteme zu entwickeln. Unter anderem wird er sich dabei mit dem Lernen komplexer Ein- und Ausgabebeschreibungen beschäftigen und sich zustandsbasierten Programmen wie zum Beispiel Datenbanken und Servern zuwenden. Zudem will er testen, ob eine Kombination mit bereits etablierten Testansätzen möglich ist. Der gebürtige Pfälzer ist derzeit Postdoc am CISPA. Zuvor hat er an der TU Darmstadt studiert und promoviert. „Ohne genau zu wissen, wo das enden würde, arbeite ich schon seit 2021 auf ISLa hin." Und der Stolz auf das, was ihm damit gelungen ist, ist dem 34-Jährigen anzumerken.

Steinhöfel, Dominic; Zeller, Andreas (2022) Input Invariants. In: Technical Track, 2022. Conference: ESEC/FSE European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering (formerly listed as ESEC)

© Lea Mosbach

Humanitäre Hilfsprogramme werden in schwierigen, manchmal sogar feindlichen Umgebungen durchgeführt, in denen es in der Regel keine angemessene digitale Infrastruktur gibt. Zudem haben die Hilfs-Empfänger:innen nur wenig Handlungsspielraum, um auf die Einhaltung ihrer Rechte zu pochen. Umso wichtiger sind Verfahren zur Verteilung von Hilfsgütern, die keine negativen Konsequenzen für die Empfänger:innen mit sich bringen und die auch für große Empfängergruppen skalierbar sind. Gemeinsam mit dem Internationalen Komitee des Roten Kreuzes haben CISPA-Faculty Dr. Wouter Lueks und seine Kolleg:innen von der EPFL in Lausanne eine neue datenschutzgerechte digitale Lösung für die Verteilung humanitärer Hilfe entwickelt. Ihr Paper „Not Yet Another Digital ID: Privacy-Preserving Humanitarian Aid Distribution“ wurde beim IEEE Symposium on Security and Privacy 2023 (S&P) veröffentlicht und mit einem „Distinguished Paper Award“ ausgezeichnet.

Ein neues digitales System zur Verteilung humanitärer Hilfe vereint Datenschutz und Rechenschaftspflicht

Wouter Lueks

Im Jahr 2021 versorgte das Internationale Komitee vom Roten Kreuz (IKRK) 3.575.484 Menschen weltweit mit Nahrungsmittelhilfe. Hilfsorganisationen wie das IKRK müssen in der Lage sein, Opfern von Gewalt, Hungersnöten und Katastrophen unter schwierigen Bedingungen in Regionen mit begrenzter Internetanbindung zu helfen. Sie tun dies mit begrenzten finanziellen Ressourcen. Um zu gewährleisten, dass so viele Menschen wie möglich unterstützt werden können, muss der Verteilungsprozess effizient und nachvollziehbar gestaltet sein. Traditionell verlassen sich humanitäre Organisationen auf papiergestützte Systeme, um die Verteilung der Hilfe zu organisieren. Diese lassen sich jedoch nur schwer auf große Gruppen von Empfänger:innen ausweiten. Darüber hinaus erschweren sie die Durchführung von Audits, um aus Sicht der Geldgeber die zweckgemäße Verwendung der Hilfsgelder zu überprüfen. Aus diesen Gründen haben einige Organisationen zuletzt damit begonnen, digitale Lösungen auszuprobieren. Die meisten arbeiten mit so genannten Identitätsmanagementsystemen (IdM), wie sie etwa bei Reisepässen Verwendung finden. Die Verwendung von IdM-basierten Lösungen birgt jedoch erhebliche Risiken für die Privatsphäre der Empfänger:innen humanitärer Hilfe, da in zentralen Datenbanken gespeicherte Informationen gestohlen oder missbräuchlich genutzt werden könnten.

Probleme lösen, ohne Risiken zu schaffen

Um eine datenschutzgerechte Lösung für diese Probleme zu finden, wandte sich das IKRK an Dr. Wouter Lueks. „Zwei Dinge haben mein Interesse geweckt“, erklärt er: „Es geht um eine technische Herausforderung und um ein datenschutzsensibles Thema.“ Das aus der Anfrage entstandene Projekt war eine Kooperation zwischen Lueks’ ehemaligem Arbeitgeber, dem EPFL in Lausanne, und dem IKRK. „Die übliche Antwort wäre gewesen, biometrische Daten zu verwenden“, so Lueks weiter. Für das IKRK war die Verwendung von Fingerabdrücken

jedoch nicht die präferierte Lösung. Denn biometrische Daten sind extrem datenschutzsensibel, gerade weil sie unveränderlich sind. Und auch die Sicherung dieser Daten ist schwierig. An dieser Stelle kamen Lueks' Forschungsinteressen ins Spiel. „Normalerweise entwickeln wir ein System, um ein Problem zu lösen. Mich interessieren jedoch die Risiken, die mit der Lösung des Problems erst entstehen können. Während einige Risiken mit dem Problem selbst verbunden sind, ergeben sich andere aus der Art und Weise, wie man die Lösung gestaltet."

Nutzt man Fingerabdrücke, um eine gerechte Verteilung humanitärer Hilfe zu gewährleisten, entstehen gleichzeitig bedeutende Risiken. Staatliche oder nichtstaatliche Akteure könnten sich etwa Zugang zu den biometrischen Informationen in der zugrunde liegenden zentralen Datenbank verschaffen und sie für Repressionen nutzen. Solche Risiken sind auf eine Entscheidung bei der Projektentwicklung zurückzuführen. Um so etwas zu vermeiden und eine maßgeschneiderte Lösung zu entwickeln, stimmten sich Lueks und das IKRK eng ab. So fanden im Laufe eines Jahres zwei Workshops und regelmäßige Treffen statt. Es entstand eine Liste von Lösungsanforderungen, die verschiedene Bereiche von den Einsatzbedingungen bis hin zu Sicherheits- und Datenschutzfaktoren umfassen und die Einhaltung der ethischen Standards des IKRK gewährleisten sollen.

Ein tokenbasiertes System zur Hilfeverteilung

Um die genannten Anforderungen zu erfüllen, schlugen Lueks und seine Kolleg:innen einen tokenbasierten Ansatz vor. Die wichtigste Entscheidung des CISPA-Forschers war es, die für das Verfahren nötigen Informationen zu dezentralisieren und über digitale Token verfügbar zu machen. Dies bedeutet, dass alle gesammelten Informationen nur in einem, bei den Empfänger:innen verbleibenden Token, gespeichert werden. Als Token können eine Smartcard oder ein Smartphone verwendet werden. Smartcards haben den Vorteil, dass sie preiswert sind und sich für Hilfsmaßnahmen mit großen Empfängergruppen in Regionen mit mangelnder Infrastruktur eignen. Das tokenbasierte System folgt dem bestehenden Workflow bei der Verteilung humanitärer Hilfe. Ist das System einmal eingerichtet, was etwa außerhalb der Zielregion und vor Beginn eines Einsatzes geschehen kann, müssen keine Updates mehr übertragen werden und auch eine Internetverbindung ist nicht mehr erforderlich.

Die Token funktionieren offline, das heißt die Chipkarte kommuniziert bei Vorlage lokal mit den Registrierungs- und Verteilungsstationen. „Eine der größten Herausforderungen bei der Entwicklung war es sicherzustellen, dass nur berechtigte Personen Hilfe erhalten und die Prüfprotokolle nicht gefälscht werden können, während gleichzeitig so wenig Informationen wie möglich über die

Empfänger:innen preisgegeben werden“, erklärt Lueks. Aufgrund seines tokenbasierten Ansatzes können die Verteilerstation und die Auditor:innen überprüfen, ob berechtigte Empfänger:innen erschienen sind, ohne dass Informationen über die Empfänger:innen selbst preisgegeben werden. Auf diese Weise wird der Schutz der Privatsphäre gewährleistet und gleichzeitig die Durchführung von Audits durch die Geldgeber:innen ermöglicht.

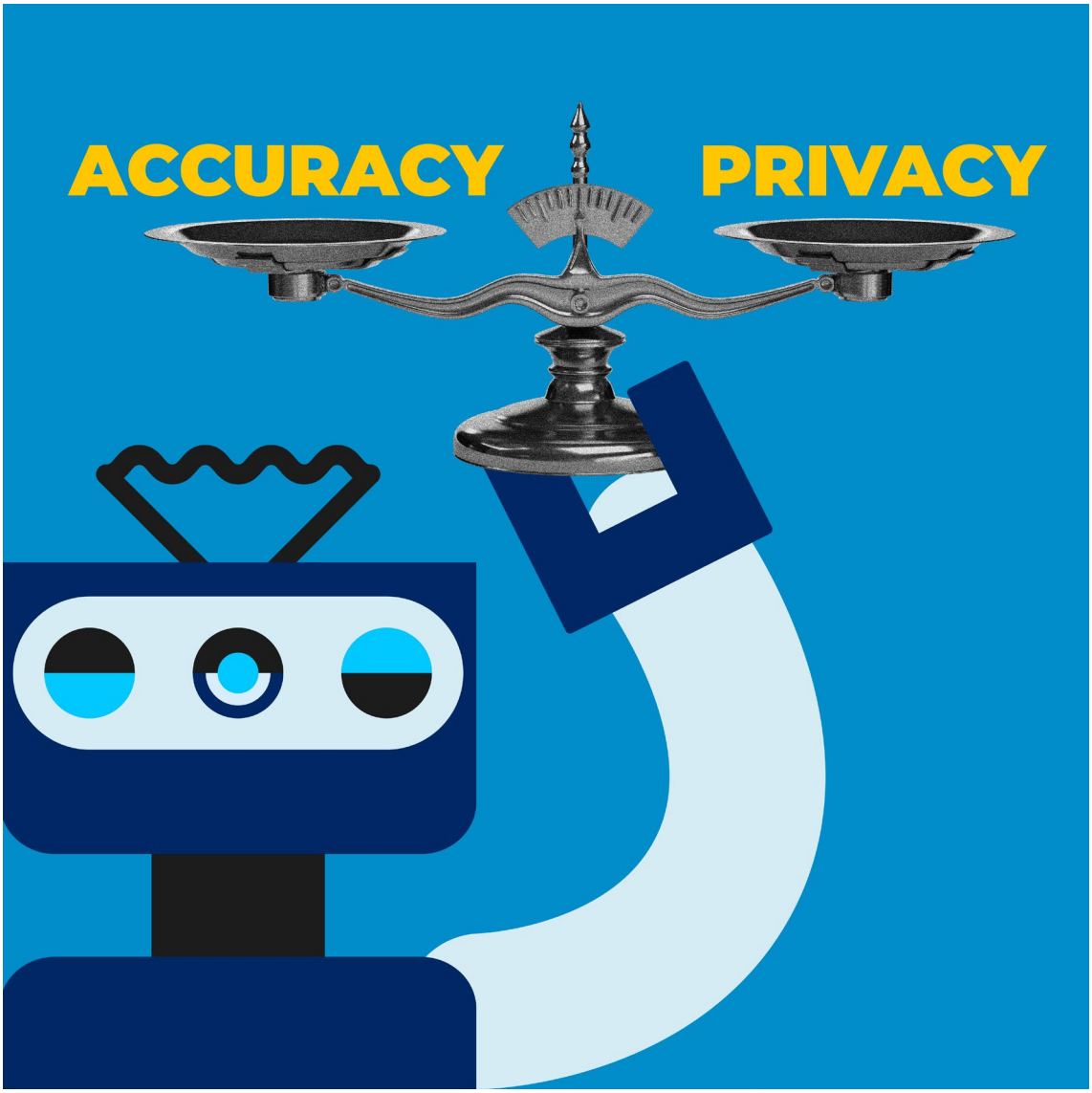
Mehr Effizienz erhöht die Fähigkeit zu helfen

Ein auf dezentrale, digitale Prozesse gestütztes System zur Verteilung von Hilfsgütern kann dazu beitragen, die Effizienz humanitärer Hilfe zu steigern. Da jede vor Ort durchgeführte Handlung gleichzeitig einen Kostenfaktor darstellt, benötigen NGOs effizientere Registrierungs- und Verteilprozesse, um mehr Menschen in Not helfen zu können. Bisher bestand das Problem papierbasierter und auch vieler digitaler Lösungen im mangelnden Datenschutz sowie Schwierigkeiten beim Durchführen von Audits. Das von Lueks entwickelte Verfahren schließt diese Lücke. „Wenn man herausfindet, was das eigentliche Problem ist, stößt man oft auf neue Herausforderungen, was diese Arbeit sehr befriedigend macht“, erklärt er.

Lueks’ Ansatz, sich auf das Problem selbst und nicht auf die Lösung zu konzentrieren, passt sehr gut zum Do-No-Harm-Ansatz, der in der humanitären Hilfe breite Anwendung findet. Er zielt darauf ab, unbeabsichtigte negative und positive Auswirkungen von humanitärer Hilfe vor Beginn eines Einsatzes zu ermitteln. Lueks hat gezeigt, dass dieser Ansatz auch für die Entwicklung neuer digitaler Lösungen genutzt werden kann. In Zukunft möchte er weiter mit NGOs zusammenarbeiten: „Für mich sind sie ein interessanter Partner, weil sie sich für die Gesellschaft einsetzen.“ Das passt hervorragend zu dem, was ihn antreibt: Technologie für einen guten Zweck zu nutzen.

Wang, Boya; Lueks, Wouter; Sukaitis, Justinas; Graf Narbel, Vincent; Troncoso, Carmela (2023) Not Yet Another Digital ID: Privacy-Preserving Humanitarian Aid Distribution. In: 44th IEEE Symposium on Security and Privacy, May 22-25, 2023, San Francisco, CA, USA. Conference: SP IEEE Symposium on Security and Privacy

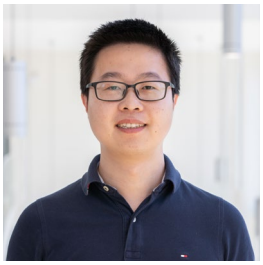
Forscher: ***Wouter Lueks***
Autor: ***Felix Koltermann***



© Lea Mosbach

Differential Privacy (DP) gilt als Gamechanger zum Schutz der Privatsphäre bei der Auswertung von Daten. Das Verfahren kommt auch schon seit einigen Jahren zum Einsatz. So hat beispielsweise das U.S. Census Bureau die letzte Volksbefragung in den USA im Jahr 2020 unter dem Einsatz von Differential Privacy durchgeführt. Auch der Tech-Riese Apple, der sich Datenschutz besonders auf die Fahnen schreibt, greift darauf zurück, um die Daten seiner Nutzer:-innen privatsphärekonform auswerten zu können. CISPA-Forschungs-gruppenleiter Zhikun Zhang hat sich in verschiedenen Papern mit der Optimierung von DP beschäftigt und diese auf dem USENIX Security Symposium präsentiert. Er erklärt, was Differential Privacy ist, welche Probleme es damit bislang noch gibt und wie er mit seiner Forschung dazu beiträgt, dass das Verfahren besser wird.

Der neue Goldstandard: Differential Privacy weitergedacht



Zhikun Zhang

Dass Daten in unserer durchdigitalisierten Welt längst zum handelbaren Gut geworden sind, ist eine Binsenweisheit. Längst nicht allen Menschen bewusst ist allerdings, wie sehr Datensammlung und -analyse der Gesellschaft heute schon dienen und in Zukunft noch dienen könnten. Ein paar Beispiele: Die Analyse von medizinischen Daten wie Blutwerte, Sauerstoffsättigung, MRT-Aufnahmen oder Röntgenbilder mithilfe von künstlicher Intelligenz (KI) wird laut Expert:innen unsere Gesundheitsversorgung in den kommenden Jahren auf ein ganz neues Level heben. Die KI kann riesige Datenmengen kombinieren und analysieren. Auch autonomes Fahren ist nicht denkbar, ohne die Auswertung immenser Mengen von Sensordaten, die überall am und im Auto gesammelt werden. Ganz zu schweigen von längst verbreiteten Bequemlichkeiten wie der Anzeige, wann im Schwimmbad wenig Gedränge zu erwarten ist, oder wo der nächste Stau droht. All das ist nur möglich durch die Auswertung riesiger Datenmengen.

An diesen Beispielen ist aber auch leicht abzulesen, wo das Problem liegt: Viele der genannten Daten sind höchst sensibel und verraten einiges über uns, unseren Gesundheitszustand, unsere Gewohnheiten und Bewegungsmuster. Der Schutz der Privatsphäre, ein an sich altes Thema, wird damit heute so relevant wie nie. Seit 2006 scheint eine Lösung gefunden. „Der neue Goldstandard des Privatsphäre-Schutzes ist Differential Privacy“, sagt Zhikun Zhang. Das Ziel von Differential Privacy ist laut dem Forscher im Grunde einfach: Aus einem bestehenden Datensatz soll soviel wie möglich über eine bestimmte Personengruppe gelernt werden, ohne etwas über die einzelnen Personen in dieser Gruppe zu erfahren.

Was steckt hinter Differential Privacy?

„Zum einen verbirgt sich hinter dem Begriff eine mathematische Definition von Privatsphäre. Es ist eine Art statistische Garantie, dass die Daten einzelner Menschen keinen Einfluss auf das Ergebnis von Abfragen zu größeren Datensätzen haben“, erklärt Zhang. „Zum anderen wird damit oft auch das konkrete Verfahren beschrieben, mit dem Datenbankabfragen so beantwortet werden, dass der Datenschutz gewährleistet bleibt.“ Entwicklerin ist die Kryptografin Cynthia Dwork. Gemeinsam mit Kolleg:innen stellte sie erstmals eine Formel

vor, mit der gemessen werden kann, wie groß die Verletzung der Privatsphäre für eine Person ist, wenn ihre Daten Teil einer größeren Datensammlung sind und damit öffentlich werden.

Rauschen für mehr Privatsphäre

Mit den vielen gesammelten Daten werden heute meist Machine-Learning-Modelle für verschiedene Aufgaben trainiert. So könnte zum Beispiel ein Modell auf einem großen Satz von Daten von Krebspatient:innen wie Blutwerten, Geninformationen und MRT-Befunden dafür trainiert werden, künftig weitaus früher als bisher eine sich entwickelnde Krebserkrankung zu erkennen. Damit die überaus sensiblen medizinischen Daten dabei sicher bleiben, müssen sie in irgendeiner Form anonymisiert werden. Es reicht allerdings nicht, persönlich identifizierende Merkmale wie Name oder Adresse zu entfernen. Denn: Durch mehrere Anfragen und die Kombination von Merkmalen, die auf den ersten Blick wenig aussagekräftig erscheinen, lassen sich häufig eindeutige Rückschlüsse auf Individuen ziehen. Stattdessen wird den Daten sogenanntes Rauschen zugefügt. Dahinter stecken verschiedene Verfahren, um eine Art kontrollierten Zufall bei der Beantwortung von Abfragen einzuführen.

Noch viele Herausforderungen für die Forschung

Wichtig ist, dass die Datenverarbeitung unter diesem Rauschen trotzdem noch ihre statistische Aussagekraft behält. Und das ist nicht die einzige Schwierigkeit. Es müssen oft mehrere spezielle Algorithmen eingesetzt werden und zudem eine Art Buchführung über die Zugriffe geführt werden, denn zu viele Abfragen können auch bei verrauschten Daten wiederum zu viel preisgeben. Die Lösung für diese Probleme können künstlich hergestellte Daten mit starken Privatsphäre-Garantien sein. „Wir veröffentlichen solche synthetischen Daten, die DP erfüllen und die statistischen Eigenschaften der echten Datensätze wiedergeben, aber nicht denselben Limitierungen bei der Verarbeitung unterliegen."

Die Herausforderung bei der Erstellung von synthetischen Daten unter DP besteht laut Zhang darin, möglichst informative statistische Informationen zu identifizieren. Nur so können auch aus komplexen Datensätzen, die etwa Bewegungsmuster von Menschen oder ihre sozialen Verknüpfungen innerhalb von Netzwerken abbilden, so viele nützliche Daten wie möglich extrahiert werden. Er hat zu seiner Forschung verschiedene Paper publiziert, die unter anderem auf dem renommierten USENIX Security Symposium vorgestellt hat.

Facettenreiches Thema

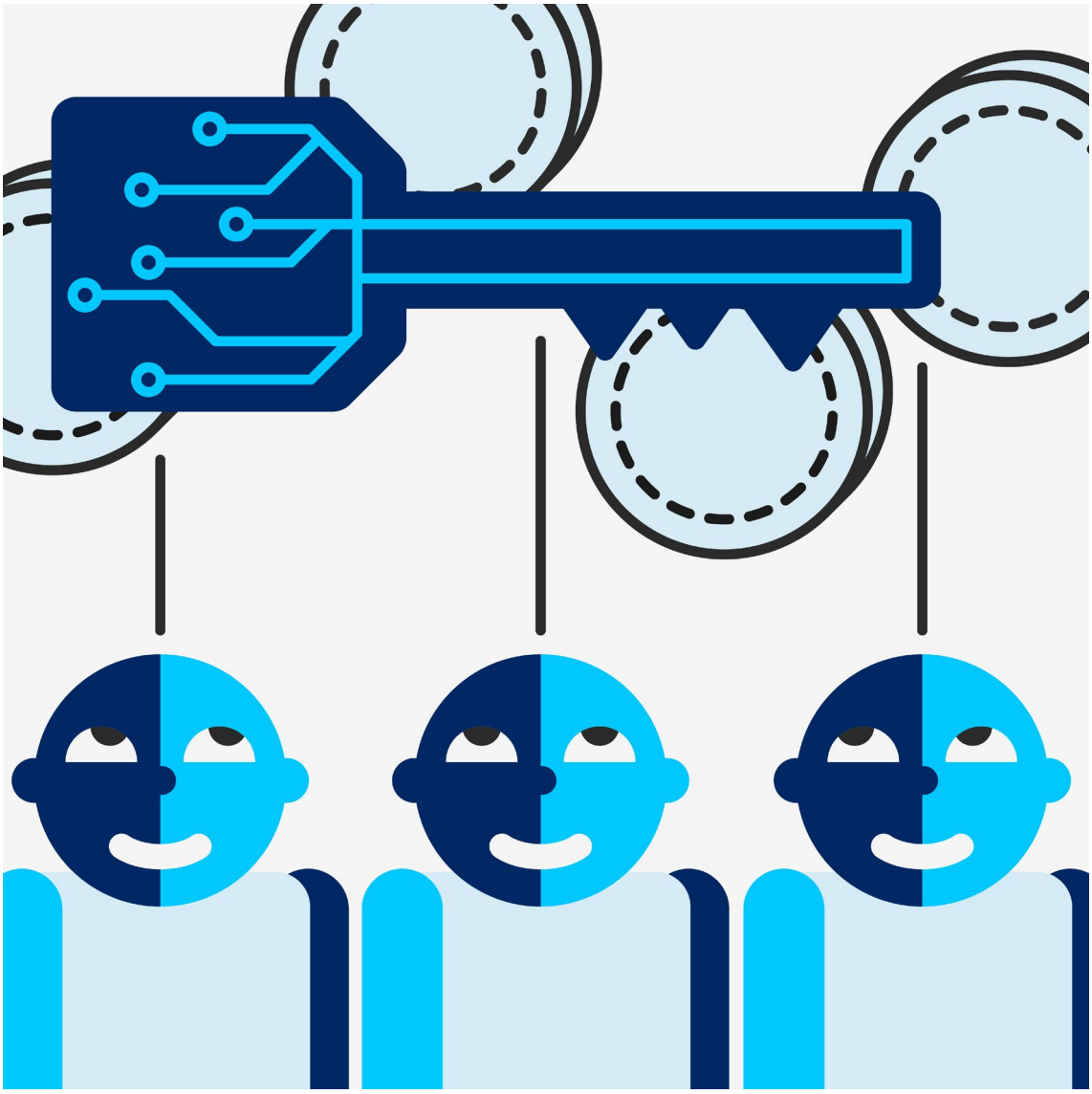
Seit Oktober 2022 forscht Zhang unter der Sonne Kaliforniens. „Ich bin Teilnehmer im CISPA-Stanford-Programm und bin derzeit Gastprofessor an der Stanford University." Differential Privacy bleibt weiterhin ein

Thema, das ihn umtreibt. „Ich forsche derzeit mit einem Kollegen aus Stanford zur Frage, wie es um den Privatsphäre-Schutz innerhalb von Large-Language-Modellen, wie sie etwa in ChatGPT stecken, bestellt ist und welchen Einfluss der Einsatz von Differential Privacy auf solche Modelle haben könnte.“ Goldgräberstimmung.

»Wir veröffentlichen synthetische Daten, die Differential Privacy erfüllen und die statistischen Eigenschaften der echten Datensätze wiedergeben, aber nicht denselben Limitierungen bei der Verarbeitung unterliegen.«

Wang, Haiming; Zhang, Zhikun; Wang, Tianhao; He, Shibo; Backes, Michael; Chen, Jiming; Zhang, Yang (2023) PrivTrace: Differentially Private Trajectory Synthesis by Adaptive Markov Model. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium

Zhang, Zhikun and Wang, Tianhao and Honorio, Jean and Li, Ninghui and Backes, Michael and He, Shibo and Chen, Jiming and Zhang, Yang (2021) PrivSyn: Differentially Private Data Synthesis. In: 30th USENIX Security Symposium, 11-13 Aug 2021, Vancouver, B.C., Canada. Conference: USENIX Security Symposium



© Janine Wichmann-Paulus

Investierten bis vor einigen Jahren nur besonders tech-affine und risikofreudige Spekulant:innen in Bitcoin und Co., mausern sich Kryptowährungen langsam zu einer neuen Anlageklasse an den regulären Finanzmärkten. Um damit sicher handeln zu können, brauchen Besitzer:innen kryptografische Schlüssel und müssen dafür sorgen, dass diese geheim bleiben. Unterstützende Schlüssel-Management-Systeme sind bislang auf Einzelnutzer:innen ausgelegt, in Finanzinstitutionen müssen aber mehrere Menschen Zugriff auf die Schlüssel haben. CISPA-Forscherin und PhD-Studentin Carolyn Guthoff hat eine qualitative Interviewstudie mit 13 Finanzexpert:innen durchgeführt und zeigt auf, wie solche Systeme für neue Anwendungsfelder im Finanzsektor fit gemacht werden müssen. Ihr Paper „Perceptions of Distributed Ledger Technology Key Management“ hat Guthoff auf dem renommierten IEEE Symposium on Security and Privacy (S&P) präsentiert.

Key-Management wird bei Krypto-Fonds zur Herausforderung

Carolyn Guthoff

Eine dezentral verwaltete Währung, auf die keine Bank, kein Staat und keine Behörde Zugriff hat – das ist die Idee hinter Bitcoin. 2008 erstmals in einem Dokument beschrieben, ist Bitcoin auch heute noch der bekannteste, aber längst nicht mehr der einzige Anwendungsfall der ihm zugrundeliegenden sogenannten Distributed-Ledger-Technologie (DLT). Distributed Ledger bedeutet soviel wie „verteiltes Geschäftsbuch" und genau das steckt dahinter: eine Datenbank für Transaktionen, die auf vielen Rechnern liegt und damit nicht zentral von einer Stelle aus, sondern von vielen Nutzenden dezentral verwaltet wird. „Distributed-Ledger-Technologien sind seit 2014 ein absolutes Hype-Thema in den verschiedensten Branchen", erklärt Guthoff.

Weitaus geläufiger als DLT ist vielen Menschen der Begriff der Blockchain. Blockchain ist eine der bekanntesten Distributed-Ledger-Technologien und Grundlage für Kryptowährungen. „Der Name kommt daher, dass in der Blockchain Datenblöcke hintereinander abgespeichert werden. Blockchain-Anwendungen wie Bitcoin oder Ethereum basieren auf derselben Technologie, folgen aber unterschiedlichen Regeln", so Guthoff. Ziel sei aber immer, eine Währung zu haben, die Onlinezahlungen ganz ohne Beteiligung von Finanzinstitutionen möglich machen.

Einzelnutzer-Szenario bei Kryptowährung ist veraltet

Diesen Ursprungsgedanken vor Augen wundert es wenig, dass sich um die Kryptowährungen herum ein eigenes Service- und Verwaltungssystem gebildet hat, das auf einzelne Nutzende ausgerichtet ist. So auch im Bereich des Managements der kryptografischen Schlüssel, das für die Abwicklung der Transaktionen in einer Blockchain elementar ist. Jede Finanztransaktion zwischen zwei Handelspartner:innen auf der Blockchain muss genauestens dokumentiert werden und ist für alle Nutzenden sicht- und nachvollziehbar. Nur so bleibt das System insgesamt vertrauenswürdig und zuverlässig. Dabei besitzen Kryptocoin-Besitzer:innen neben einem öffentlichen auch einen sogenannten Private Key, mit dem sie auf ihre digitale Geldbörse zugreifen können und Transaktionen digital signieren können. Diese Private Keys sind 52 Zeichen lang und werden den Nutzer:innen zufällig zugewiesen. Geht ein solcher Key verloren,

ist auch die mit dem Schlüssel verbundene Kryptowährung endgültig verloren. Daher ist die sichere Speicherung der privaten Schlüssel essenziell.

Studie liefert Designideen für Key-Management in Multi-User-Szenarien

Die Anforderungen an die sichere Verwaltung und Speicherung von kryptografischen Schlüsseln wachsen, wenn mehrere Nutzende Zugriff darauf brauchen. „Das ist zum Beispiel bei Kryptofonds standardmäßig der Fall. Seit einer Gesetzesänderung 2022 sind solche Fonds in Deutschland ein größeres Thema in Finanzinstitutionen geworden und wie bei anderen Fonds auch, werden sie meist von mehreren Mitarbeitenden gemanagt. Zudem müssen sich Kolleg:innen ja auch im Falle von Urlaub oder Krankheit vertreten können“, erklärt Guthoff. Sie hat 13 Mitarbeitende in Finanzinstituten dazu befragt, welche Sicherheits- und Geheimhaltungsanforderungen die Institute für die Schlüsselverwaltung haben und wie sich die Mitarbeitenden ein optimales Schlüsselmanagement vorstellen. „Die Ergebnisse dieser Studie können dabei helfen, Key-Managementlösungen für Finanzinstitutionen entsprechend ihrer Bedürfnisse, sicher und zugleich praktikabel zu designen.“

Eine der größten Herausforderung für das Schlüsselmanagement bei mehreren Nutzenden ist in der Praxis die Fluktuation von Mitarbeitenden. „Hatte ein Mitarbeiter einmal Zugang zu einem Schlüssel, besteht das Risiko, dass er ihn kopiert hat. Auch wenn er zwischenzeitig gekündigt hat oder auf einer anderen Stelle sitzt, kann er dann noch auf die Vermögenswerte zugreifen“, sagt Guthoff. Eine gute Lösung für dieses Problem könnte nach Ansicht einiger Studienteilnehmer:innen die Nutzung eines Programmes sein, das den Einsatz von Schlüsseln für Transaktionen ermöglicht, aber keinen direkten Zugriff auf

»Die Ergebnisse dieser Studie können helfen, Key-Managementlösungen für Finanzinstitutionen entsprechend ihrer Bedürfnisse zu designen.«

den Schlüssel selbst zulässt. „Überhaupt wünschten sich die Teilnehmenden für die Speicherung der Schlüssel überwiegend technische Lösungen, die durch mehrere Faktoren wie TANs und Passwörter abgesichert werden können“, erklärt die Forscherin.

Eine andere wichtige Frage ist, wie mit Haftungs- und Verantwortungsfragen umgegangen wird. Viele der Befragten stellen sich für ein optimales Schlüsselmanagement und die Verteilung entsprechender Zugriffsrechte auf Vermögenswerte Modelle vor, die der Organisationsstruktur ihres Unternehmens entsprechen. „Das heißt, dass zum Beispiel CEOs über höhere Vermögenswerte verfügen können als einfache Angestellte und erweiterte Zugriffrechte bekommen sollen“, erklärt Guthoff. Die meisten Befragten wünschten sich zudem ein Key-Management, das nicht zu viel Hintergrundwissen zu digitalen Signaturen erfordert und möglichst einfach zu bedienen ist. Einen Intermediär zwischen dem Finanzinstitut und der Handelsplattform einzuschalten, der das Key-Management und dessen Absicherung übernehmen könnte, fanden einige Teilnehmende nützlich, sofern ein entsprechendes Vertrauensverhältnis herrscht.

Viele spannende Forschungsfragen

Für Guthoff ist es das erste Paper, das sie auf einer Konferenz einreicht. „Dass meine Arbeit auf der renommierten S&P angenommen wurde, ist toll. Das bestärkt mich.“ Trotz der tiefen Einarbeitung in das Thema Kryptowährung will sich Guthoff in ihrer Forschung künftig nicht auf Finanzthemen fokussieren. „Die Arbeit an diesem Thema war super spannend, aber jetzt werde ich mich erstmal anderen Forschungsfragen zuwenden. Mich interessieren vor allem Themen, bei denen die Vorstellungen und Ansprüche von Sicherheitsforschenden und die Lebensrealitäten der Anwender:innen nicht so richtig zusammenpassen.“ Davon gibt es vermutlich noch einige.

Guthoff, Carolyn; Anell, Simon; Hainzinger, Johann; Dabrowski, Adrian; Krombholz, Katharina (2023) Perceptions of Distributed Ledger Technology Key Management – An Interview Study with Finance Professionals. In: 44th IEEE Symposium on Security and Privacy, 22-25 May 2023 San Francisco, CA, USA. Conference: SP IEEE Symposium on Security and Privacy

Im Jahr 2022 wurden weltweit mehr als 1,14 Milliarden Websites gezählt. Viele davon sind in der Europäischen Union (EU) gehostet oder werden von Menschen aus der EU genutzt. In diesen Fällen findet seit dem Jahr 2018 die europäische Datenschutz-Grundverordnung (DSGVO) Anwendung. Sie verpflichtet Unternehmen und Website-Betreiber:innen sicherzustellen, dass die personenbezogenen Daten ihrer Kund:innen und Nutzer:innen geschützt bleiben. Die CISPA-Forscherin Christine Utz und ihre Kolleg:innen haben nun untersucht, wie die Betreiber:innen von Websites auf die mangelnde Umsetzung der DSGVO sowie der ePrivacy-Richtlinie hingewiesen werden können. Die Ergebnisse haben sie im Paper „Comparing Large-Scale Privacy and Security Notifications“ publiziert, das auf dem Privacy Enhancing Technologies Symposium (PETS) 2023 vorgestellt wurde.

Betreiber:innen von Websites nehmen Sicherheit wichtiger als Datenschutz

Christine Utz

Wird im Internet eine Website geöffnet, ist damit in der Regel auch ein Datenaustausch verbunden. Seitenbetreiber:innen verfolgen zum Beispiel häufig, von welcher IP-Adresse die Website aufgerufen wird. Außerdem geben Kund:innen oft selbst persönliche Daten an, etwa wenn sie Produkte im Internet erwerben und sich diese nach Hause liefern lassen. Mit der seit 2018 gültigen Datenschutz-Grundverordnung (DSGVO) wurde erstmals eine europaweit einheitliche Richtlinie zur Verarbeitung personenbezogener Daten eingeführt. Ziel ist der Schutz der Nutzer:innen vor übermäßiger Datenspeicherung. Bereits die Speicherung einer IP-Adresse gilt als Speicherung persönlicher Daten. Die DSGVO findet für Websites Anwendung, die in der EU gehostet sind oder dort aufgerufen werden können. Verantwortlich für die Umsetzung der Richtlinie sind die Website-Betreiber:innen, während die Kontrolle den nationalen Datenschutzbehörden obliegt.

CISPA-Forscherin Utz untersuchte im Jahr 2019 mit Martin Degeling von der Ruhr-Universität Bochum, wie sich Websites nach der DSGVO-Einführung verändert haben. „Die Haupterkenntnis war, dass sich zwar an der tatsächlichen Praxis des Trackings kaum etwas geändert hatte, aber die Transparenzbemühungen der Websites etwa über das Zurverfügungstellen von Datenschutzerklärungen sowie die Einführung von Cookie-Bannern gestiegen waren“, erzählt Utz. Dies war einer der Ausgangspunkte für ihre aktuelle Studie. Darüber hinaus hatte CISPA-Faculty Dr. Ben Stock, in dessen Gruppe Utz forscht, in einer Studie untersucht, wie Website-Betreiber:innen mit Hilfe von E-Mail-Kampagnen über Sicherheitslücken informiert werden können. „Daraus entstand die Idee, zu untersuchen, ob die Betreiber:innen von Websites mithilfe einer solchen Kampagne auch auf mangelnden Datenschutz hingewiesen werden können“, so Utz weiter.

Studiendesign und Vorgehen bei der Untersuchung

Nach umfangreichen Vorrecherchen erfolgte die konkrete Umsetzung der Studie mit einem Set von ca. 160.000 Websites. Kriterien für die Aufnahme einer Website in die Stichprobe war das Vorhandensein eines

Datenschutzproblems wie etwa dem Fehlen einer Datenschutzerklärung, dem Nicht-Vorhandensein oder zu spätem Anzeigen von Cookie-Bannern sowie Inputfeldern für persönliche Daten ohne Absicherung mit HTTPS. Als Vergleichskriterium wurde darüber hinaus noch ein Sicherheitsproblem in die Studie aufgenommen. Dies war ein ungesicherter Zugang zu einem sogenannten Git Repository, also einer auf einem externen Server gespeicherten Arbeitskopie der Website-Entwickler:innen. Anfang November 2021 wurden die Betreiber:innen automatisiert per Mail angeschrieben und auf die Probleme hingewiesen. Über zwei Monate hat Utz dann bei den Angeschriebenen sowie in einer Kontrollgruppe beobachtet, ob die Probleme auf den Seiten behoben wurden oder nicht. Um tiefergehende Erkenntnisse über die Umsetzung bzw. Nicht-Umsetzung sowie deren Gründe zu erlangen, verschickten die Forschenden mit den E-Mails auch einen Fragebogen und untersuchten ihre E-Mail-Kommunikation mit den Website-Betreiber:innen.

Herausforderungen bei der Umsetzung

Eine Studie mit einer sechsstelligen Stichprobe bringt eine Reihe von Herausforderungen mit sich, die sich unter anderem aus der Automatisierung vieler Arbeitsschritte ergeben. So liegt ein Risiko in falsch-positiven Ergebnissen, weil zum Beispiel automatische Tools zur Durchsuchung der HTML-Quelltexte tatsächlich vorhandene Datenschutzerklärungen, etwa aufgrund uneinheitlicher Benennungen, nicht erkennen. Eine weitere Hürde ist die Auswahl der E-Mail-Adressen, da frühere Studien gezeigt haben, dass die Nutzung von generischen Adressen wie info@- oder webmaster@- Nachteile mit sich bringt. Deswegen wurden, sofern möglich, die E-Mails an konkrete, auf der Website erkannte E-Mail-Adressen geschickt. „Die größte Schwierigkeit war jedoch zu verhindern, dass unsere E-Mails von den Empfänger:innen als Spam eingestuft werden“, erklärt Utz. Um dies zu verhindern, trafen Utz und ihre Kolleg:innen eine Reihe von Vorkehrungen. So wurde ein externer Server für das Hosting genutzt und die E-Mails signiert. Der externe Server sollte auch verhindern, dass alle vom CISPA stammenden E-Mails als Spam eingestuft und aussortiert werden und damit dem Zentrum Schaden entstehen könnte.

Ergebnisse verweisen auf hohe Hürde der Methode

Das wichtigste Ergebnis der Studie war, dass es grundsätzlich möglich ist, mit einer großangelegten Benachrichtigungskampagne Website-Betreiber:innen per E-Mail über Datenschutzprobleme zu informieren. Gleichwohl ist der angesichts der immensen Ressourcen, die für die Durchführung einer solchen Studie nötig sind, recht begrenzt. Dies zeigt sich vor allem daran, dass nur ein sehr kleiner Teil der Informierten überhaupt auf die E-Mails reagierte. Die Anzahl der Websites, auf denen im

Beobachtungszeitraum Veränderungen vorgenommen wurden, bewegte sich im niedrigen einstelligen Prozent-Bereich. Darüber zeigte sich im Vergleich, dass Sicherheitslücken eher behoben werden als Datenschutzprobleme. Einen Grund sieht die CISPA-Forscherin darin, dass Sicherheitslücken mit weniger Aufwand geschlossen werden können.

Weitere Gründe für den beschränkten Erfolg der Kampagne innerhalb des zweimonatigen Untersuchungszeitraums ergab die qualitative Untersuchung der ausgefüllten Fragebögen sowie der E-Mail-Kommunikation mit den Website-Betreiber:innen. Utz konnte hier verschiedene Hürden zur Umsetzung von Änderungen herausarbeiten. Dazu zählten Sprachbarrieren aufgrund mangelnder Englischkenntnisse auf Seiten der Empfänger:innen der E-Mails oder das Einstufen der E-Mails als Spam. Als weitere Hürde erwies sich auch der DSGVO-Bezug selbst. Einige Betreiber:innen bezweifelten, dass die eigene Website überhaupt in den Anwendungsbereich der DSGVO fällt, oder wiesen den Hinweis auf mangelnden Datenschutz pauschal als unzutreffend zurück. Insgesamt zeigten sich die Website-Betreiber:innen weniger offen für Benachrichtigungen über Datenschutzprobleme als über Sicherheitslücken. Gewünscht hätten sich die Teilnehmer:innen auch mehr und detailliertere Informationen zu den datenschutzrelevanten Fragen.

Kooperation mit Datenschutzbehörden als Ziel

Utz' Anliegen ist es, mit ihrer Forschung die Durchsetzungsfähigkeit der DSGVO zu erhöhen. „Datenschutzbehörden haben oft keine Kapazitäten, die mangelnde Umsetzung der DSGVO auf Websites zu erkennen und die Betreiber:innen darauf hinzuweisen", erzählt sie. Forschende und die dahinterstehenden Institutionen verfügen oft nicht über die notwendigen technischen und personellen Ressourcen für solche Projekte. Umgekehrt könnten die Forschenden von der Autorität der Behörden profitieren. „Die Datenschutzbehörden können besser vermitteln, warum die DSGVO wichtig ist", so Utz. Eine Kooperation wäre also eine Win-Win-Situation für alle Beteiligten. Laut Utz ist es jedoch unerlässlich, dass breit angelegte E-Mail-Kampagnen in Zukunft von anderen Maßnahmen flankiert werden, wie etwa Informationskampagnen über den Anwendungsbereich der DSGVO. Des Weiteren schlägt sie die Implementierung eines neuen Standards vor, wie die Betreiber:innen von Websites einfacher erreicht werden können. Die könnte etwa durch eine auf allen Websites hinterlegte Datei privacy.txt geschehen, in der Kontaktinformationen zu den Betreiber:innen bei Datenschutzfragen hinterlegt sind.

Utz, Christine; Michels, Matthias; Degeling, Martin; Marnau, Ninja; Stock, Ben (2023) Comparing Large-Scale Privacy and Security Notifications. In: PETS 2023, July 10–15, 2023, Lausanne, Switzerland. Conference: PETS Privacy Enhancing Technologies Symposium

© Janine Wichmann-Paulus

Die Raumfahrt zum Mond (und darüber hinaus) hat immer schon die Aufmerksamkeit der Öffentlichkeit auf sich gezogen. Die eigentliche Eroberung des Alls findet jedoch in aller Stille in der erdnahen Umlaufbahn, dem „Low Earth Orbit“ (LEO) statt. Mit einer Entfernung von 200 bis 1000 km befindet sich LEO vergleichsweise nah an der Erde und beherbergt eine schnell wachsende Zahl relativ kleiner, relativ preiswerter Satelliten. In ihrer Studie „Space Odyssey: An Experimental Software Security Analysis of Satellites“ untersuchen die beiden CISPA-Faculty Dr. Ali Abbasi und Prof. Dr. Thorsten Holz zusammen mit Forschenden von der Ruhr-Universität Bochum die Sicherheitsprobleme, die mit dem Anbruch dieses neuen Weltraumzeitalters einhergehen. Auf dem IEEE Symposium on Security and Privacy (S&P) im Mai 2023 wurde ihre Publikation mit einem Distinguished Paper Award ausgezeichnet. Das ist eine Ehre, die nur den besten 1 Prozent der eingereichten Arbeiten zuteil wird.

Auffällig im All: Eine Studie zur Satellitensicherheit

Ali Abbasi

Ein Merkmal der so genannten „new space era“ ist die stetig wachsende Anzahl an Satelliten, die die Erde umkreisen. Ein großer Teil davon besteht aus LEO-Satelliten, die aufgrund ihrer geringen Größe und Kosten nicht nur für Staaten und Großunternehmen, sondern auch für kleine Einrichtungen und Firmen zugänglich sind. Amazon beispielsweise bietet Satellitenkommunikation On-Demand an und betreibt Bodenstationen als Dienstleistung. Laut Orbiting Now, einer Website, die Satellitendaten zusammenträgt, wurden im Mai 2023 insgesamt 7.004 aktive LEO-Satelliten gezählt. Abhängig von ihrer Nutzlast können diese Satelliten verschiedenen Zwecken dienen, darunter Erdbeobachtung, Wettervorhersage, Navigation, Kommunikation und Weltraumforschung.

Es war diese plötzliche, breite Verfügbarkeit von LEO-Satelliten, die das Forschungsinteresse der CISPA-Forscher Ali Abbasi und Thorsten Holz weckte. „Es findet gerade ein Paradigmenwechsel statt. Und jeder Paradigmenwechsel bringt auch Sicherheitsprobleme mit sich“, sagt Abbasi. Die lang von Satelliteningenieur:innen gehegte Überzeugung, dass Unsichtbarkeit Sicherheit garantiert, gilt nicht mehr. Abbasi erklärt: „Lange Zeit ging man davon aus, dass Satelliten nicht zugänglich und somit sicher sind. Aber LEO-Satelliten verfügen über zahlreiche Konnektivitätsmöglichkeiten.“ Das Fehlen offizieller Sicherheitsstandards für Satelliten ist ein zusätzlicher Anlass zur Sorge, wie Holz anführt: „Einen Satelliten kann man nur über ein proprietäres Funkprotokoll kontaktieren. Aber die Frequenzen, auf denen Satelliten kommunizieren, sind nicht geregelt.“

Space Oddities, oder: Wie sicher sind Satelliten?

Satelliten werden über ein sogenanntes Bussystem gesteuert und kommunizieren auch darüber. Der Bus besteht aus dem Kommunikationsmodul (COM), das Funknachrichten von der Bodenstation empfängt, und dem Command and Data Handling System (CDHS), das alle eingehenden Befehle verarbeitet und ausführt. Das COM ist sozusagen das Ohr des Satelliten, während das CDHS als sein Gehirn fungiert: Es trägt eine Computerplattform, die auf der Grundlage einer vorinstallierten, bordseitigen Software, der Firmware, arbeitet. Satelliten ähneln somit anderen, gängigeren Computersystemen und sind auch ähnlich anfällig für Software-Angriffe.

Ausgehend von der Annahme, dass Satellitensysteme weniger sicher sind als moderne Windows-, Linux- oder MacOS-Systeme, konzentrierten sich die Forscher auf die Angriffsflächen der Satelliten-Firmware.

Als Ausgangspunkt für ihre Untersuchung erstellten sie eine Taxonomie möglicher Bedrohungen für die Satelliten-Firmware. Sie identifizierten drei übergreifende Angreiferziele und skizzierten alle möglichen Angriffswege, die zu deren Verwirklichung genutzt werden könnten. Das ultimative Ziel eines Angriffs kann darin bestehen, entweder die Verfügbarkeit des Satelliten zu beeinträchtigen, Zugang zu Satellitendaten zu erhalten, oder die Kontrolle über den gesamten Satelliten zu erlangen. Dieses letztgenannte Angreiferziel birgt gleichzeitig das größte Schadenspotenzial: Wird ein Satellit gekapert und für den Angriff auf einen anderen genutzt, können die entstehenden Trümmer einen Dominoeffekt auslösen, bei dem der Weltraum mit losen Satellitenteilen überschwemmt wird. Dieser als Kessler-Syndrom bezeichnete Effekt ist nach Abbasis Worten jedoch größtenteils „Hollywood-Kram".

„Wenn Angreifende erst einmal Zugang erlangen, hat man Pech": Schwachstellen in der Firmware

Ob Hollywood-Kram oder nicht, das Forscherteam um den Doktoranden Johannes Willbold von der Ruhr-Universität Bochum konnte erfolgreich Fehlerzustände auf dem CDHS auslösen und im angewandten Teil der Studie die volle Kontrolle über zwei von drei realen LEO-Satelliten übernehmen. Zum Zweck einer Sicherheitsanalyse hatten die Forschenden den Zugang zu den Firmware-Images dieser drei Satelliten erhalten. Zuvor hatten sie über einen längeren Zeitraum vertrauensvolle Beziehungen zu den institutionellen Eigentümern der Satelliten aufgebaut.

Die Ergebnisse dieser Fallstudien verdeutlichen, dass eine wissenschaftliche Auseinandersetzung mit der Sicherheit von Satelliten überfällig gewesen ist. Die wichtigsten Erkenntnisse der Studie betreffen die Sicherheit des COM. Als Eingangspunkt für Funknachrichten von der Bodenstation sollte das COM idealerweise eine Türsteherfunktion ausüben und verdächtige Befehle abwehren. Erfüllt es diese Rolle nicht, kann das CDHS durch unvorhergesehene Eingaben angegriffen werden. Wenn es diesen Eingaben dann gelingt, einen Fehlerzustand in der Onboard-Software auszulösen, greifen sie direkt in das Gehirn des Satelliten ein.

Obwohl es sich bei Satelliten um hochkomplexe Systeme handelt, sind die von den Forschenden aufgedeckten Software-Schwachstellen erstaunlich gewöhnlich. Holz erklärt: „In der Linux- oder Windows-Welt untersuchen wir Softwarefehler dieser Art schon seit vielen Jahren. Aber bei den vorliegenden, eingebetteten Systemen sind die Schutzmaßnahmen 20 Jahre hinter

dem zurück, was wir von herkömmlichen Systemen kennen.“ Willbold betont, dass der wirksamste Schutzmechanismus derzeit außerhalb des Systems liegt: „Die Barriere ist der Zugang. Aber wenn Angreifende erst einmal Zugang erlangen, hat man Pech gehabt.“

Verantwortungsvolle Offenlegung: Für mehr Satellitensicherheit

Bevor sie ihre Studie veröffentlichten, meldeten die Forschenden alle entdeckten Softwareprobleme an die Eigentümer der drei Satelliten. Dieses Verfahren, die so genannte verantwortungsvolle Offenlegung, ist Teil ihres beruflichen Verhaltenskodex. Aber auch für die Förderung der Satellitensicherheit ist sie unerlässlich. Künftig können diese Systeme nur geschützt werden, wenn Forschende, Betreiber:innen und Entwickler:innen zusammenarbeiten, wie Abbasi betont: „Diejenigen, die ihre Satelliten-Firmware mit uns geteilt haben, haben Mut bewiesen. Ihnen liegt die Cybersicherheit wirklich am Herzen, denn kurzfristig haben sie damit nichts gewonnen: Da gibt es vielleicht ein Problem mit ihrer Software, aber mit jeder Software gibt es Probleme. Aber langfristig haben sie geholfen, Weltraumsysteme zu schützen.“

Willbold, Johannes; Schloegel, Moritz; Vögele, Manuel; Gerhardt, Maximilian; Holz, Thorsten; Abbasi, Ali (2023) Space Odyssey: An Experimental Software Security Analysis of Satellites. In: 44th IEEE Symposium on Security and Privacy, 22-25 May 2023 San Francisco, CA, USA. Conference: SP IEEE Symposium on Security and Privacy

CISPA-Faculty Dr. Michael Schwarz beschäftigt sich seit Jahren mit Seitenkanalangriffen. Unter anderem war er an der Entdeckung von Platypus und Meltdown beteiligt. Das sind Cyberangriffe, bei denen Daten über einen Umweg, einen sogenannten Seitenkanal, gestohlen werden. Dazu werden Informationen ausgenutzt, die der Prozessor unfreiwillig bei der Verarbeitung abgibt, wie etwa das Laufzeitverhalten oder der Energieverbrauch. Mit Collide+Power haben Schwarz und sein PhD-Student Lukas Gerlach gemeinsam mit einer Forschergruppe der TU Graz jetzt einen neuen strombasierten Seitenkanalangriff entdeckt, der unmittelbar auf die Central Processing Unit (CPU) abzielt und theoretisch alle Prozessoren treffen kann. Ihr Paper „Collide+Power: Leaking Inaccessible Data with Software-based Power“ haben sie 2023 auf dem USENIX Security Symposium vorgestellt.

Collide+Power: Neuer Seitenkanalangriff betrifft alle Prozessoren

Michael Schwarz

Mit Collide+Power können Daten direkt vom Prozessor des Computers abgegriffen werden. Alle Daten, die ein Computersystem verarbeiten soll, müssen durch die Central Processing Unit (CPU) geschleust werden. Die CPU verfügt über interne Zwischenspeicher, oder Caches, in denen bereits verarbeitete Daten vorgehalten werden, damit sie für zukünftige Arbeitsprozesse schnell wieder bereitstehen. Werden die Daten im Cache nun durch neue Daten überschrieben, etwa, weil man in seinem Passwortmanager auf ein weiteres Passwort zugreift, wird Strom verbraucht. Dabei gilt eine physikalische Besonderheit: Je mehr Daten im Cache verändert werden, umso mehr Strom wird benötigt.

Kollision der Daten im Cache

Diesen Effekt macht Collide+Power sich zu nutze. Der für den Angriff programmierte Schadcode bewirkt, dass der Cache mit den Angreifer:innen bekannten Daten gefüllt wird. Greifen User:innen nun auf ein Programm – etwa ihren Passwortmanager – zu, werden die Daten der Angreifer:innen im Cache mit dem Passwort überschrieben: Angreifer- und User-Daten „kollidieren“ sozusagen. Anhand des Stromverbrauchs der CPU während des Überschreibungsvorgangs können die Angreifer:innen Rückschlüsse auf das Passwort ziehen. Denn: „Je ähnlicher sich die geladenen Daten und die Daten aus dem Zielprogramm sind, desto geringer ist der Stromverbrauch – und den kann man sehr genau messen“, erklärt Schwarz.

Natürlich finden in den Zwischenspeichern eines Computers viele Rechenprozesse parallel statt, zum Beispiel weil verschiedene Programme gleichzeitig geöffnet sind. Wie können die Angreifer:innen also den Teil der Berechnungen im Cache identifizieren, den sie ausbeuten wollen? Gerlach erläutert: „Der eingeschleuste Schadcode bewirkt, dass die Daten aus dem Programm, das angegriffen werden soll, zigfach im Cache geladen werden.“ Diese stetig wiederholten Ladevorgänge erlauben den Angreifer:innen Rückschlüsse darauf, welche Datensätze für sie relevant sind.

Stromverbrauch erlaubt Rückschlüsse auf Daten

Möglich ist diese Art des Datenklaus, da im Speicher eines Computers alle Werte mit Nullen und Einsen kodiert sind. Dabei wird jeder einzelne Wert mit mehreren Stellen

abgebildet, die je entweder mit einer Eins oder einer Null besetzt werden: Für ein Byte, das mit acht dieser Stellen abgebildet wird, wäre eine Eins kodiert mit 0000 0001, eine Zwei mit 0000 0010. Um eine Eins im Cache mit einer Zwei zu überschreiben, müssen also zwei Stellen, die letzten beiden, geändert werden. Überschreibt man die Eins mit einer Null, die mit 0000 0000 kodiert ist, ändert sich allein die letzte Stelle. Dazu braucht es weniger Strom. Durch einen Abgleich des Stromverbrauchs „errät“ Collide+Power nacheinander jede der Stellen eines Wertes.

Sehr viele Wiederholungen dieses „Ratevorgangs“ sind nötig, um alle Stellen eines Wertes und dadurch das Geheimnis zu erbeuten. Das macht das Verfahren sehr aufwendig und zeitintensiv. Mit dem aktuell programmierten Schadcode würde das Extrahieren einer Kreditkartennummer beispielsweise vier bis fünf Stunden in Anspruch nehmen, so schätzen die Forschenden. „Allerdings ist das auch nur unser Testcode. Wenn man es ernst meint, könnte man den Code ganz sicher optimieren“, sagt Schwarz.

»Wir Forscher können nur aufzeigen, dass es möglich ist. Wie gefährlich es ist, müssen die Hersteller selbst beurteilen.«

Collide+Power schließt eine Forschungslücke

Mit Collide+Power schließen die Forschenden eine Lücke in der Erkennung von strombasierten Seitenkanalangriffen. Es ist der erste Seitenkanalangriff, der Strommesswerte dazu verwendet, Daten unmittelbar vom Prozessor des Computers abzuleiten. Da die Hardware selbst zum Ziel des Angriffs wird, ist diese Form des Angriffs nicht zu verhindern. Die Hersteller können nur ihrer Informationspflicht nachkommen und Hinweise formulieren. Bislang, sagt Michael Schwarz, ist Collide+Power aus der Praxis noch nicht bekannt: „Wir Forscher können nur aufzeigen, dass es möglich ist. Wie gefährlich es ist, müssen die Hersteller selbst beurteilen." Allerdings, fügt Lukas Gerlach an, „man verliert die Garantie, dass Daten unangreifbar bleiben."

Kogler, Andreas; Juffinger, Jonas; Giner, Lukas; Gerlach, Lukas; Schwarzl, Martin; Schwarz, Michael; Gruss, Daniel; Mangard, Stefan (2023) Collide+Power: Leaking Inaccessible Data with Software-based Power Side Channels. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium

© Janine Wichmann-Paulus

„Als Mobilfunkforscher ist man in gewisser Weise Gefangener seiner geografischen Position“, sagt CISPA-Forscher Dr. Adrian Dabrowski. Damit meint er, dass Mobilfunkforschende wegen der großen Zahl von Anbieter:innen und Netzen bislang nur mit immensem Aufwand Tests und Messungen in ausländischen Mobilfunknetzen vornehmen können. Zusammen mit Gabriel Gegenhuber von der Universität Wien und weiteren Forschungskollegen hat Dabrowski deshalb Mobile-Atlas entwickelt, eine Art Infrastruktur, die die Testung quer durch Europa erlaubt – egal von welchem Standort aus. Sein Paper und den neuen Ansatz des „Geographically Decoupled Measurements in Cellular Networks for Security and Privacy Research” stellt er auch auf dem renommierten USENIX Security Symposium 2023 vor.

MobileAtlas: Eine Kartografie der Mobilfunk-Sicherheit

Adrian Dabrowski

2G, 3G, 4G, 5G – was sich anhört wie die Ziehung beim Bingo bezeichnet die aktuell verwendeten Mobilfunkstandards. Der jüngste Standard der 5. Generation – dafür stehen all die Gs – ist noch im Aufbau. Der älteste, 2G, wurde schon in den 1990er-Jahren Jahren eingeführt und ist noch immer in Verwendung. „2G wird vor allem für Sprachübertragung oder für einfache smarte Geräte verwendet; etwa ein Getränkeautomat, der anzeigt, dass er nachgefüllt werden muss“, erklärt Dabrowski. Das darauffolgende 3G wurde 2021 in Deutschland abgeschaltet und durch 4G, auch LTE genannt, ersetzt. Mit 4G lassen sich unterwegs zum Beispiel Streamingdienste nutzen oder Videotelefonie durchführen. Mittlerweile gelten diese nebeneinander existierenden Mobilfunkstandards weltweit. Durch das sogenannte Roaming sollen Mobilfunk-Kund:innen auch im Ausland die mit ihrem Mobilfunk-Anbieter:innen vereinbarten Services nutzen können und den versprochenen Sicherheits- und Privatsphäreschutz genießen.

Ist das „Roam-Like-At-Home-Prinzip“ ein leeres Versprechen?

Die Rede ist hier vom sogenannten Roam-Like-At-Home-Prinzip, das EU-Bürger:innen in der 2022 neugefassten EU-Roamingverordnung versprochen wird. Die Bundesnetzagentur schreibt dazu: „Durch die Neufassung der Roaming-Verordnung gilt auf Reisen in der EU nicht nur der gleiche Preis wie zuhause, sondern auch grundsätzlich die gleiche Qualität.“ Dabrowski bezweifelt, dass dieses Versprechen eingehalten werden kann.

»Wenn man genau hinschaut, gibt es keine Konsistenz zwischen Roaming- und Nicht-Roaming-Verbindungen«

„Beim Roaming arbeiten das Heimatnetzwerk und das Netzwerk des Landes, in dem ich zu Gast bin, zusammen. Sie wollen einen Service anbieten, der auch hinsichtlich Privatsphäre und Sicherheit so konsistent ist wie der im Heimatnetz. Die technische Umsetzung ist dabei aber komplett verschieden.“ So werde zum Beispiel bei einem Urlaub in der Schweiz die Telefonverbindung direkt übers Schweizer Netz hergestellt, während die Internetverbindung den Umweg über Deutschland nehme. Im Heimnetzwerk ginge beides den direkten Weg. „Wenn man genau hinschaut, gibt es keine Konsistenz zwischen Roaming- und Nicht-Roaming-Verbindungen“, erklärt der Forscher. Die Mobilfunk-Anbieter:innen hätten extrem viel Gestaltungsspielraum und seien bislang kaum zu kontrollieren. Das gilt laut Dabrowski auch hinsichtlich der Sicherheit der Netze.

Grenzüberschreitende Tests bislang kaum möglich

Das Problem: Bislang sind Tests und Messungen über die Grenzen hinweg extrem aufwendig. „Europa ist extrem zersplittert. In jedem Land gibt es viele Mobilfunk-Anbieter:innen. Deutschland ist mit seinen nur drei Anbieter:innen die Ausnahme. Wenn ich feststelle, dass in einem unserer inländischen Mobilfunknetz eine Sicherheitslücke ist, und prüfen will, ob das in anderen Netzen auch der Fall ist, habe ich derzeit zwei Möglichkeiten: Entweder ich reise viel herum und teste jedes Netz in jedem Land in jeder Konstellation, oder ich statte in jedem Land möglichst viele Geräte mit möglichst vielen verschiedenen SIM-Karten von unterschiedlichen Anbieter:innen aus. In kürzester Zeit habe ich so 1000 SIM-Karten, 1000 Verträge und eine Privatinsolvenz.“

„Entkoppelte Messungen“ sind die Lösung

Die Lösung könnte ein von den Forschenden entwickeltes Framework sein, das die geografische Trennung der SIM-Karte vom Mobilfunkmodem erlaubt. Das Modem ist eine Komponente in mobilen Endgeräten wie etwa Smartphones, das die Verbindung zwischen den Geräten und einem Mobilfunknetz herstellt. Seine Aufgabe ist es, die Funkdaten in die richtige Form zu bringen, an die Sendemasten senden und von dort zu empfangen. Die SIM-Karte dient zur Identifikation der Nutzer:innen und ordnet das Smartphone einem bestimmten Netz zu. Dabrowski erklärt, was das alles mit seinem Framework zu tun hat: „Normalerweise sind SIM-Karte und Telefon eine Einheit. Wir trennen diese Einheit auf und entfernen die SIM-Karte aus dem Telefon. Wir simulieren das Kommunikationsprotokoll übers Internet und können so quasi virtuell reisen. Nochmal einfacher an einem Beispiel erklärt: Wir verbinden einmal die SIM-Karte mit unserer Messstation in Deutschland und können so tun, als wären wir in Deutschland. Dann trennen wir sie und verbinden sie zu unserer Messstation in Frankreich und können so

tun als seien wir da. Wir brauchen dafür nur noch ein Endgerät in Deutschland oder eben in Frankreich.“

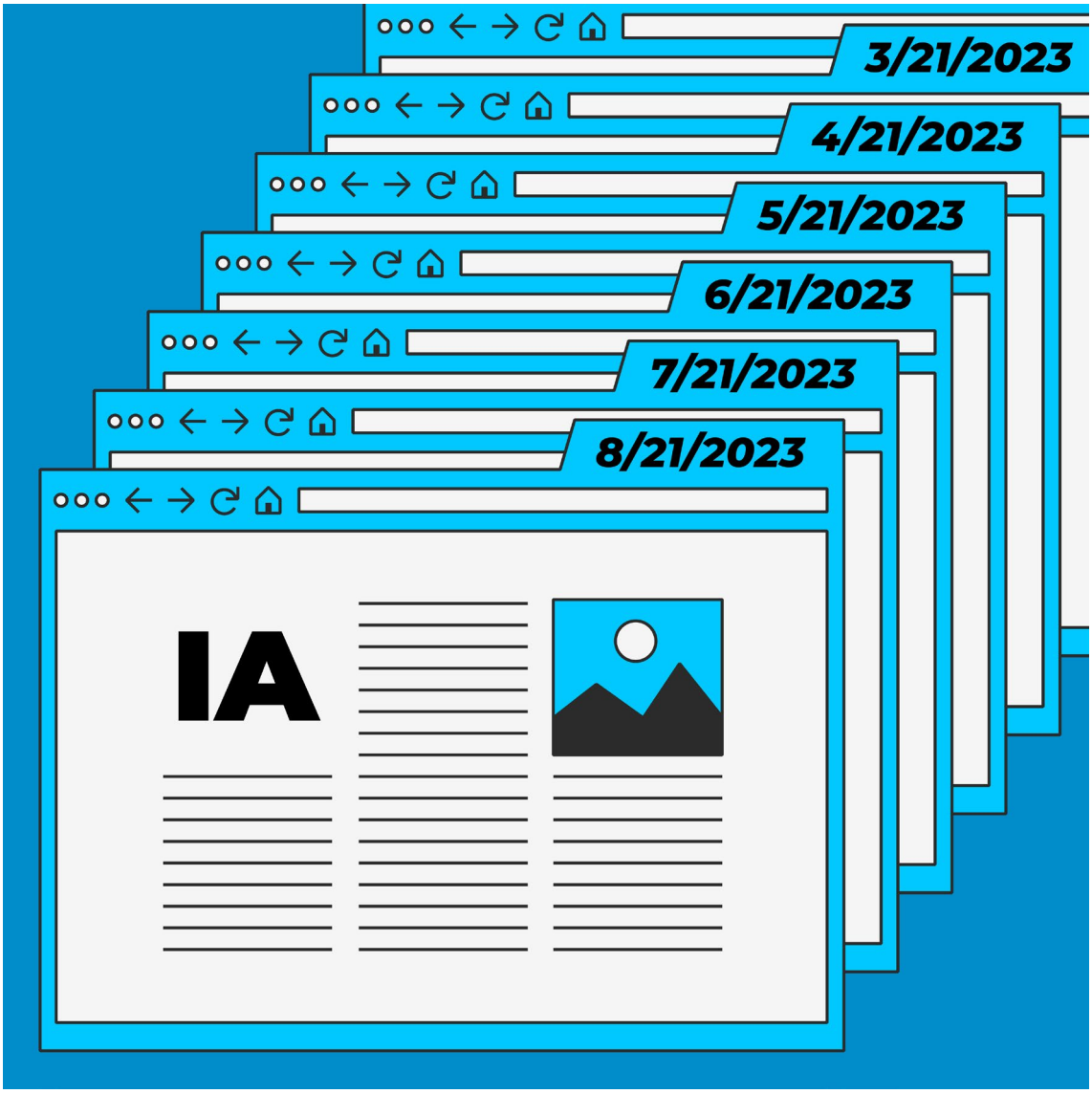
Kostengünstig und Open Source

Die daraus resultierende Mess- und Testplattform, die für die Standards 2G bis 4G funktioniert, bietet laut Dabrowski eine kontrollierte Experimentierumgebung die erweiterbar und kostengünstig ist. „Zudem ist unser Ansatz Open Source, sodass andere Forschende Standorte, SIM-Karten und Messskripte dazu beitragen können.“ Die Forschenden machen die Plattform unter dem Namen MobileAtlas zugänglich und nutzbar. Das Tool dürfte dabei nicht nur für Wissenschaftler:innen interessant sein. „Mobilfunk-Anbieter:innen könnten damit auch erstmals prüfen, ob ihre Roamingpartner ihre Versprechen halten.“ Der Name Mobile Atlas kommt dabei nicht von Ungefähr. Er wurde laut Dabrowski abgeleitet vom Namen der seit 2010 existierenden Internet-Testplattform RIPE Atlas „RIPE NCC ist die europäische Internetverwaltung. Der RIPE Atlas ist ein globales Netz von Messgeräten, die die Konnektivität und Erreichbarkeit des Internets messen.“

Die Grenzen der Grenzenlosigkeit

Mit dem MobileAtlas gibt es bislang in zehn Ländern Messstationen und die für die Messungen geeignete Infrastruktur. Dabrowski hofft, dass sich das Messnetzwerk durch die Hilfe anderer Forschender schnell vergrößert. „Allerdings werden wir auch schauen müssen, dass kein Unfug mit den SIM-Karten getrieben wird, damit uns keine Kosten entstehen. Ob wir MobileAtlas so umfänglich anbieten können wie RIPE NCC ihre Plattform, muss sich noch zeigen.“ Dass sich mit ihrem Ansatz interessante Informationen zu Tage fördern lassen, haben Dabrowski und seine Kollegen bereits bewiesen: „Wir haben zum Beispiel entdeckt, dass sich in einigen Mobilfunknetzen bestimmte Dienste so tarnen lassen, dass der dafür anfallende Datenverkehr nicht vom im Tarif enthaltenen Datenvolumen abgezogen wird. Für Endnutzer:innen schlimmer sind allerdings die Sicherheitsproblematiken, die wir ebenfalls nachweisen konnten. So haben wir zum Teil problematische Firewall-Konfigurationen gefunden oder versteckte SIM-Kartenkommunikation mit dem Heimnetzwerk aufgedeckt.“ Allzu beunruhigend sind die Befunde dabei nicht. Eine Ausnutzung dieser Probleme würde sehr gezielte Angriffe und versierte Angreifer:innen voraussetzen. „Aber solche Lücken sind nie gut. Und jetzt haben wir die Möglichkeit, die Anbieter:innen darauf hinzuweisen.“

Gegenhuber, Gabriel Karl; Mayer, Wilfried; Weippl, Edgar; Dabrowski, Adrian (2023) MobileAtlas: Geographically Decoupled Measurements in Cellular Networks for Security and Privacy Research. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium



© Lea Mosbach

Qualität in der Wissenschaft wird unter anderem daran gemessen, ob Studien von anderen Wissenschaftler:innen reproduziert werden können und dabei das gleiche Ergebnis herauskommt. Gerade bei Studien zu Internet-Sicherheitsmechanismen ist das jedoch eine große Herausforderung, da Websites sich ständig verändernde Gebilde sind. CISPA-Forscher Florian Hantke und seine Kolleg:innen aus dem Team von CISPA-Faculty Dr. Ben Stock haben nun in Kooperation mit Forschenden der Universität Ca' Foscari in Venedig mit einem neuen Ansatz, Web-Archive statt Live-Analysen für diese Studien zu nutzen, vielversprechende Ergebnisse erzielt.

Ein neuer Standard? Die Nutzung von Web-Archiven für Live-Analysen zur Sicherheit von Websites

Florian Hantke

Studien zur Sicherheit von Websites nehmen im Forschungsgebiet der Informationssicherheit einen breiten Raum ein. Dabei ist der Standard in der Forschung bis heute oft die Live-Analyse. Das bedeutet, dass bestimmte Parameter zur Sicherheit von Websites in dem Moment gemessen werden, in dem die Forschenden auf eine Website zugreifen. Problematisch ist, dass dies immer nur eine Momentaufnahme darstellt: Was in einem Moment „live" ist, kann einen Tag später schon veraltet sein. „Das Web ist so random, dass es extrem komplex ist, Experimente zu reproduzieren", so CISPA-Forscher Florian Hantke. Deswegen ist es bei Live-Analysen fast unmöglich, Experimente unter gleichen Bedingungen zu wiederholen. Für Hantke stellt dies ein grundsätzliches Problem dar: „Experimente sollten immer reproduzierbar sein, weil ein Experiment sonst an Relevanz verliert. Sonst könnte jeder einfach behaupten, das Internet wäre sicher." Eine Alternative, mit der das Kriterium der Reproduzierbarkeit gewährleistet werden kann, könnte laut Hantke theoretisch die Nutzung von Web-Archiven darstellen. Web-Archive speichern in regelmäßigen Abständen Kopien existierender Websites, sogenannte „snapshots", auf externen Servern. Dort können sie versehen mit Datum und Timecode abgerufen werden. Anders als an Live-Websites gibt es an den gespeicherten Kopien keine Veränderungen mehr. Das bekannteste Web-Archiv ist das Internet Archive. In der Forschung werden Web-Archive bisher vor allem für historische Analysen, aber nicht für Live-Analysen verwendet. Hantke erklärt dies damit, dass „viele Leute denken, in den Archiven wären nicht alle wichtigen Daten vorhanden."

Internet Archive anderen Web-Archiven überlegen

CISPA-Forscher Hantke und seine Kolleg:innen wollten nun wissen, wie gut sich Web-Archive für Live-Analysen zur Überprüfung von Sicherheitsmechanismen von Websites eignen. Dafür mussten sie herausfinden, welches der existierenden Web-Archive die genauesten

Kopien speichert. Konkret untersuchten sie dafür eine Reihe öffentlicher Web-Archive hinsichtlich des Umfangs und der Qualität der hinterlegten Daten der 5.000 wichtigsten Websites für den Zeitraum von Januar 2016 bis Juli 2022. Im Vergleich der verschiedenen Web-Archive zeigte das Internet Archive (IA) die besten Resultate. Die Qualität des Archivs ist so gut, dass die Autor:innen um Hantke unter bestimmten Voraussetzungen sogar eine Arbeit mit dem IA als alleiniger Quelle empfehlen. Die Datenqualität des IA überprüften sie anhand einer Fallstudie zu zwei Mechanismen, die zum Standard vieler Websites gehören: den sogenannten Security Headern sowie Java-Script-Inclusions. Darüber hinaus zeigten sie auf, dass das IA so regelmäßig Kopien von Websites speichert, dass auch detailliertere Analysen möglich sind, deren Qualität Live-Analysen in nichts nachsteht. Zusätzlich ermöglicht das IA die Analyse von mehreren Snapshots einer Website im gleichen Zeitraum, was Hantke als „Neighborhood" bezeichnet. Dies ermöglicht etwaige kurzzeitige Ausreißer in den Daten, wie zum Beispiel Serverprobleme einer Website, zu glätten. Durch das von den Forschenden genutzte Verfahren, öffentlich zugängliche Web-Archive zu nutzen, werden Studien einfacher reproduzierbar. Langfristig kann dadurch die Qualität der Forschung gesteigert und können Sicherheitsmechnismen von Websites besser überprüft werden.

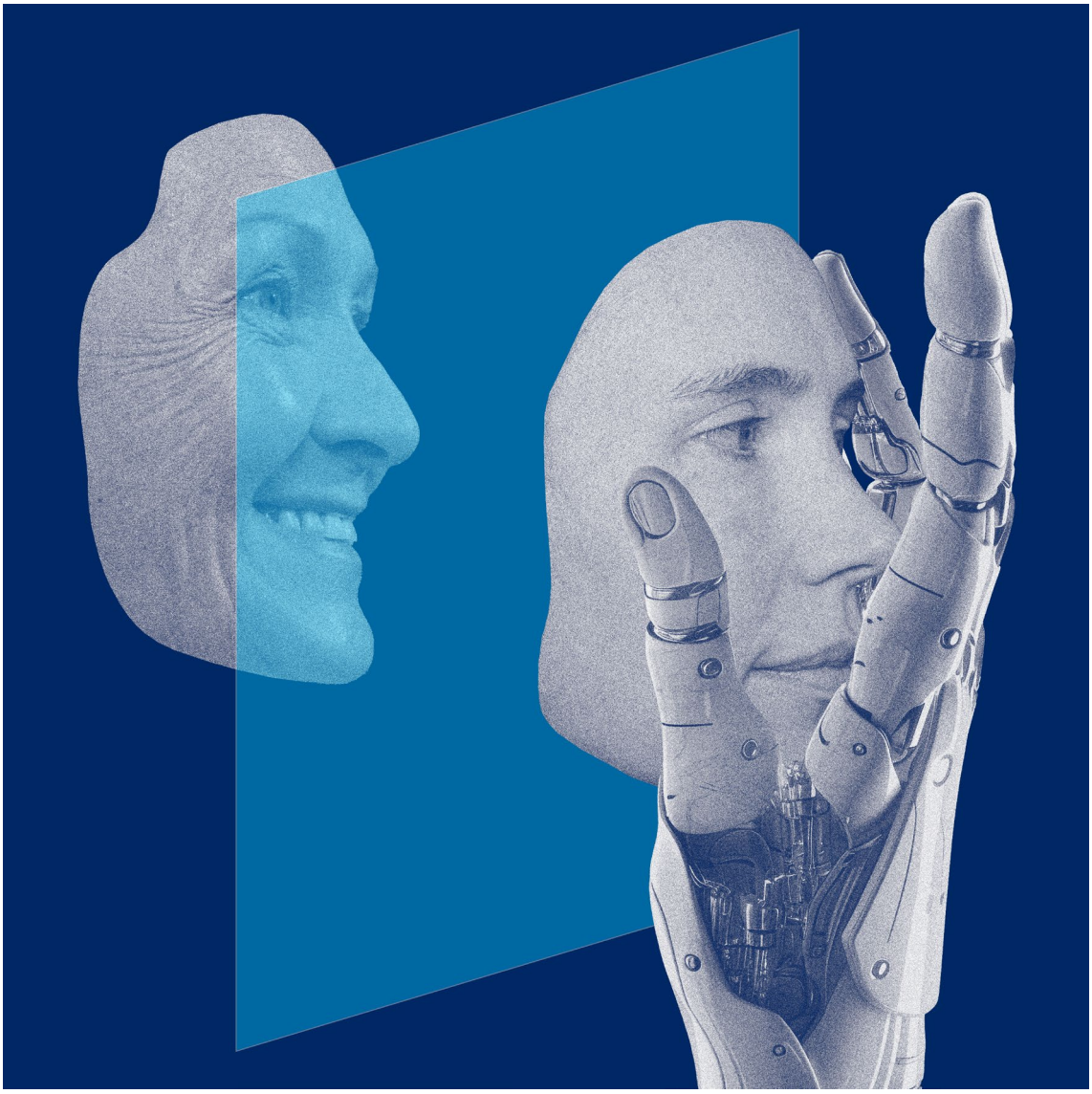
Herausforderungen bei der Nutzung von Web-Archiven

Gleichwohl gibt es auch bei der Nutzung von Web-Archiven für die Live-Analyse einiges zu bedenken. „Ein großer Nachteil ist die langsame Geschwindigkeit", erklärt Hantke. So ist die Bearbeitung großer Datenmengen bei einer klassischen Live-Analyse wesentlich schneller, da der Zugriff auf in den Web-Archiven gespeicherte Daten sehr langsam ist. Umgehen ließe sich dies jedoch durch Kooperationen mit den Archiven, wie etwa dem von Hantke und Kolleg:innen favorisierten IA, um einen besseren Zugriff auf die Daten zu bekommen. „Zu beachten sind auch die unterschiedlichen Vantage Points", so Hantke weiter. Das sind die Zugriffsstellen, von wo auf der Welt auf die Websites zugegriffen wird. Diese entscheiden auch darüber, wie genau eine Website aussieht, die im Archiv gespeichert wird. „Bei Sicherheitsthemen sind die Unterschiede eher vernachlässigbar, aber bei Analysen etwa zur Implementierung der DSGVO wird der Zugriffsort schon wichtig", erklärt der CISPA-Forscher. Denn spezifische Features, die für die Datenschutzgrundverordnung (DSGVO) relevant sind, werden oft nur auf europäischen Websites angezeigt. Eine in den USA gespeicherte Kopie würde hier also nicht helfen. Deswegen muss für jede neue Forschungsfrage überprüft werden, ob die Arbeit mit Web-Archiven in Frage kommt.

Produktive Forschungstätigkeit als PhD

Florian Hantke arbeitet jetzt seit gut einem Jahr als PhD beim CISPA. Da er mit seiner Frau in Erlangen lebt, arbeitet er viel im Homeoffice. Danach gefragt, ob ihm dabei zu Hause nicht spezielles Equipment für die Forschung fehlt, erzählt er, dass eine sichere VPN-Verbindung zum CISPA-Server in Saarbrücken völlig ausreiche. „Ich schicke dann einfach eine Anweisung an den Server und lasse die Analysen dort laufen“, so Hantke. Die Resultate kann er dann später bequem abrufen. Das Paper zu den Web-Archiven ist bereits seine zweite Veröffentlichung. „Ich bin ganz zufrieden mit meinem Output“, gesteht er lachend. Und für den Sommer ist bereits ein weiteres Paper in Planung. Aber vorher hofft er, dass es noch mehr Interesse an seinen Erkenntnissen zur Nutzung von Web-Archiven für Sicherheitsanalysen gibt. Das Management von Internet Archive hat auf jeden Fall schon Interesse signalisiert. Und in Zusammenarbeit mit den an der Studie beteiligten Co-Autor:innen von der Universität Ca’ Foscari in Venedig ist auch ein öffentlich zugängliches Projekt für Web-Sicherheitsanalysen in Planung, das auch andere Forscher:innen nutzen können.

Hantke, Florian; Calzavara, Stefano; Wilhelm, Moritz; Rabitti, Alvise; Stock, Ben (2023) You Call This Archaeology? Evaluating Web Archives for Reproducible Web Security Measurements. In: ACM CCS 2023, 26-30 Nov 2023, Copenhagen, Denmark. Conference: CCS ACM Conference on Computer and Communications Security



© Lea Mosbach

Die Allgegenwart von Bildern im Internet auf der einen sowie die exponentielle Lernkurve von KI-Bildgeneratoren auf der anderen Seite erhöhen auch das Risiko von Bildmanipulationen mit böswilliger Absicht. CISPA-Forscher Zheng Li und seine Kolleg:innen haben nun ein Verfahren getestet, mit dem dies teilweise verhindert werden kann. Die Ergebnisse ihrer Studie haben sie im Aufsatz „UnGANable: Defending Against GAN-based Face Manipulation“ beim renommierten USENIX Security Symposium publiziert.

Test eines neuen Verfahrens zum Schutz vor Deepfakes

Zheng Li

Ein wesentliches Merkmal zeitgenössischer Online-Kommunikation in sozialen Netzwerken ist der Austausch von digitalen Bildern zwischen verschiedenen Nutzer:innen. Einmal hochgeladen, bleiben die Bilder meist für sehr lange Zeit verfügbar. Damit ist auch einer manipulativen Verwendung Tür und Tor geöffnet. Neben dem Identitätsklau, bei dem mit realen Bildern falsche Profile angelegt werden, ist ein weiteres Risiko, dass diese Bilder in KI-Bildgeneratoren eingespeist und für Deepfakes genutzt werden. Deepfakes sind Manipulationen an Bildern, die mit dem bloßen Auge nicht erkennbar sind. Insbesondere Politiker:innen und Personen des öffentlichen Lebens sind diesem Risiko ausgesetzt. „Meist wissen die Personen auf den Fotos nichts von der Manipulation und können sich nicht mal dagegen wehren“, so CISPA-Forscher Zheng Li. Gezielte Desinformationen werden damit noch einfacher. „Deshalb stellen Deepfakes auch eine echte Gefahr für die Demokratie dar.“ Erstellt werden können Deepfakes mit Hilfe verschiedener Verfahren auf Basis künstlicher Intelligenz, wie etwa durch die Nutzung sogenannter GANs.

»Meist wissen die Personen auf den Fotos nichts von der Manipulation und können sich nicht mal dagegen wehren.«

GAN ist die Abkürzung für Generative Adversarial Network und bezeichnet ein Modell des maschinellen Lernens. GANs bestehen aus zwei künstlichen neuronalen Netzen, die miteinander kommunizieren. Vereinfacht gesagt generiert eines der beiden Netze neue Daten, also zum Beispiel Bilder, während das andere diese Daten ausgehend von einem bestehenden Datensatz bewertet. Dazu vergleicht es die Unterschiede zwischen den bestehenden und den neu generierten Daten. Diese Bewertung wird an das generierende Netz zurückgespielt und von diesem für Verbesserungen genutzt, etwa damit die generierten Bilder realen Bildern immer ähnlicher werden, was auch den Algorithmus an sich verbessert. Im gleichen Maße, wie die Bildauflösung immer besser sowie der Look immer fotorealistischer wird, erweitern sich auch die Möglichkeiten, Bilder realer Personen zu manipulieren. Hier geht es vor allem um den Bereich der „Face Manipulation" – auf Deutsch übersetzt Bildmanipulation von Gesichtern –, bei der einzelne Merkmale, wie etwa der Gesichtsausdruck oder die Haarfarbe, verändert werden können. Eine wichtige Methode zur Generierung von Deepfakes ist die GAN-Inversion, ein spezielles Verfahren zur Verarbeitung von Bildern in KI-Bildgeneratoren.

Die Entwicklung von UnGANable

„Unser Ausgangspunkt war die Erkenntnis, dass es bis dato keine Möglichkeit gab, Deepfakes zu verhindern, die auf der Methode der GAN-Inversion basieren", erzählt Li. „Unser Verfahren haben wir deswegen UnGANable genannt", so der CISPA-Forscher weiter. „Vereinfacht gesagt versucht UnGANable, Bilder von Gesichtern vor Deepfakes zu schützen." Damit GANs Bilder überhaupt verarbeiten können, müssen sie sie zuerst in mathematische Vektoren, den sogenannten „latent code", umwandeln. Dies wird als GAN-Inversion bezeichnet und stellt eine Art Bildkomprimierung dar. Mit Hilfe des „latent code" eines realen Bildes kann ein Generator neue Bilder, die ihrem realen Vor-Bild täuschend ähnlich sind.

An dieser Stelle greift nun das von Li und seinen Kolleg:innen entwickelte Verfahren, das die GAN-Inversion und damit die Fälschungen erschwert. Dafür produziert UnGANable auf Ebene der mathematischen Vektoren maximale Abweichungen, in der Fachsprache „noise" genannt, die auf Bildebene jedoch nicht sichtbar sind und die Umwandlung in „latent code" erschweren. Damit läuft das GAN – einfach gesprochen – quasi trocken, weil es keine Daten findet, mit deren Hilfe neue Bilder erstellt werden können. Und wenn keine dem Originalbild ähnlichen Kopien auf Basis des „latent code" erstellt werden können, ist auch keine Bildmanipulation möglich. Tests mit UnGANable bei verschiedenen GAN-Inversion-Verfahren ergaben zufriedenstellende Resultate. Darüber hinaus konnten Li und Kolleg:innen nachweisen, dass

ihr Verfahren auch besser schützt als alternative Mechanismen wie etwa das Programm Fawkes. Das von einer Forschungsgruppe des Sand Lab in Chicago entwickelte Programm arbeitet mit einem Verzerrungsalgorithmus, der an Fotos mit dem menschlichen Auge nicht wahrnehmbare Veränderungen auf Pixelebene vornimmt.

Anwendungsbereiche des Verfahrens

Die Arbeit des CISPA-Forschers ist ein wichtiger Schritt hin zur Entwicklung neuer Verfahren zum Schutz vor „Face Manipulation". „Mir ist es wichtig, die Menschen vor der böswilligen Manipulation ihrer Bilder zu schützen", erklärt Li. Der Code für das von ihm mitentwickelte Verfahren ist Open-Source, also öffentlich zugänglich, und kann so auch von anderen Forschenden genutzt werden. Und wer im Umgang mit Codes versiert ist, kann diesen auch bereits jetzt dafür nutzen, die eigenen Bilder vor missbräuchlicher Nutzung zu schützen. Aber damit die breite Masse der User:innen dies anwenden kann, müsste noch eine entsprechende Software programmiert werden. Während CISPA-Forscher Li sich schon wieder anderen Projekten zugewendet hat, forschen einige seiner Kolleg:innen weiter an ähnlichen Fragestellungen, etwa ob der gefundene Mechanismus auch für KI-Verfahren genutzt werden kann, die aus Texteingaben Bilder generieren. „Und vielleicht lässt sich das Verfahren in Zukunft auch für Videos einsetzen", überlegt Li. Sicher ist in jedem Fall, dass angesichts der exponentiellen Lernkurve von GAN-basierten Verfahren zur Bildgenerierung und damit auch zur Bildmanipulation die Notwendigkeit zur Entwicklung von Abwehrmechanismen weiter steigen wird.

Li, Zheng; Yu, Ning; Salem, Ahmed; Backes, Michael; Fritz, Mario; Zhang, Yang (2023) UnGANable: Defending Against GAN-based Face Manipulation. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium

Forscher: ***Zheng Li***
Autor: ***Felix Koltermann***

© Janine Wichmann-Paulus

Messengerdienste bieten durch standardmäßige Ende-zu-Ende-Verschlüsselung eine relativ große Sicherheit. Aber immer nur so lange am anderen Ende tatsächlich die richtige Person chattet. Nur wenige Menschen wissen, dass eine Authentifizierung der Chat-Partner:innen entscheidend ist, um Angriffe auf den Chatverlauf zu verhindern.
Der Frage, warum dieser Akt nur selten stattfindet, ist Matthias Fassl aus der Forschungsgruppe von CISPA-Faculty Dr. Katharina Krombholz in einem Selbstversuch nachgegangen. Die Ergebnisse wurden 2023 als Paper bei der CHI Conference on Human Factors in Computing Systems veröffentlicht.

Ein Selbstversuch zeigt Schwierigkeiten beim Durchführen von Authentifizierungszeremonien

Matthias Fassl

Schon seit vielen Jahren sind Messengerdienste wie Signal, Threema oder WhatsApp eine der beliebtesten und am weitesten verbreiteten Form des digitalen Austauschs zwischen Personen. Getauscht werden nicht nur Textnachrichten, sondern auch Bilder, Dokumente und Sprachnachrichten, sowohl privater als auch dienstlicher Natur. Umso wichtiger ist die Frage, wie sicher die Nutzung dieser Dienste ist. Heute ist bei vielen Messengerdiensten eine Ende-zu-Ende-Verschlüsselung der Normalfall. Dies bedeutet, dass die Nachrichten, „sobald sie ein Gerät verlassen, so verschlüsselt werden, dass nur das Empfängergerät sie entschlüsseln kann“, erklärt Fassl. „Die große Unsicherheit ist die Frage, ob am Ende auch wirklich der richtige Mensch sitzt,“ so der Forscher weiter. „Eine der möglichen Sicherheitslücken ist ein Man-In-The-Middle Angriff, wo jemand vorgibt, zum Beispiel dein Freund Paul zu sein. Um einen solchen Angriff abzuwehren, müssen Nutzende kontrollieren, dass der Schlüssel, mit dem sich der Text entschlüsseln lässt, auch zum richtigen Empfänger gehört. Das passiert mit Hilfe von Authentifizierungszeremonien.“ Konkret bedeutet dies, dass sich zwei User:innen treffen und über auf ihren Smartphones angezeigte QR-Codes gegenseitig authentifizieren.

Methodisches Neuland zur Deckung einer Forschungslück

Die Herausforderung besteht jedoch darin, dass Nutzer:innen nur selten Authentifizierungszeremonien durchführen. Dies hat nach Ansicht von Fassl zum einen damit zu tun, dass hinter der Ende-zu-Ende-Verschlüsselung das Konzept „trust-on-first-use“ steckt. Dabei wird davon ausgegangen wird, dass User:innen denjenigen Kontakten, die sie zu einem Messenger hinzufügen, vertrauen und dies durch die Kontaktaufnahme über den Messenger bestätigen. Die tatsächliche Verschlüsselung der Chats findet dann im Hintergrund statt. Aus diesem Grund wissen viele gar nicht, dass erst die tatsächliche Authentifizierung der Chat-Partner:innen größtmögliche Sicherheit bietet. Zahlen und Studien dazu, wie oft User:innen Authentifizierungszeremonien durchführen, gibt es nach Aussage des CISPA-Forschers kaum.

Genau hier setzt Fassls Forschungsinteresse an: Er will wissen, was die Umsetzung der Zeremonie so schwer macht. „Im Laufe der Zeit bin ich darauf gekommen, dass vielleicht auch Faktoren eine Rolle spielen können, die nicht das User-Interface betreffen, sondern wie wir miteinander interagieren“, erzählt er.

Für seine Studie hat sich Fassl für die Methode der Autoethnographie entschieden. „Ethnographische Ansätze sind relativ praktisch, um soziale und kulturelle Faktoren zwischen mehreren Menschen zu untersuchen, die sozial interagieren“, erzählt er. „Eine Autoethnographie ist dasselbe, nur mit der eigenen Person. Das ist ein Sonderfall und nicht so gerne gesehen, weil die Untersuchungsperson und die untersuchende Person dieselbe sind.“ Gleichwohl gebe es auch Vorteile eines autoethnographischen Vorgehens, da man „nicht immer alles punktgenau erfassen muss, weil sich Dinge auch nachträglich aus der Erinnerung ergänzen lassen.“ Herausfordernd aufgrund der Methodenwahl war hingegen die Publikation der Ergebnisse. „Dadurch, dass die Methode nicht so gern gesehen ist, war es etwas schwer, das zu publizieren. Es war im Cybersecurity-Bereich auch erst die zweite Autoethnographie, die ich gefunden habe.“

Unterschätzter Aufwand zur Durchführung der Zeremonien

Umso interessanter – auch in den Augen der Reviewer:innen – waren die Ergebnisse, die Fassl über die mehrmonatige Selbstbeobachtung zusammentragen konnte. So konnte er nachweisen, dass die größte Herausforderung der Authentifizierungszeremonien der Planungs- und Organisationsaufwand ist. „Ich muss mich nicht nur mit Leuten treffen, sondern auch überlegen, wie ich die Zeremonie ins Gespräch einbaue“, erklärt der CISPA-Forscher. „Ich persönlich bin für die Studie alle meine Kontakte in den Messengern durchgegangen und habe geschaut, wo es schon einen grünen Haken gibt und wen ich noch authentifizieren muss. Ich habe dann versucht, das relativ systematisch abzuarbeiten.“ In diesem Prozess gibt es laut seiner Beobachtungen auch noch vor der eigentlichen Zeremonie verschiedene Stellen, wo es zu Brüchen kommen kann. „Das sind Momente, wo die Leute aussteigen, weil sie die Zeremonie vergessen oder während der sozialen Interaktion spannendere Unterhaltungsthemen aufkommen.“

Oft musste er seinen Gesprächspartner:innen auch erst einmal erklären, was es mit Authentifizierungszeremonien auf sich hat. Hier zeigte sich, dass persönliche Faktoren eine entscheidende Rolle spielen. Die können jedoch individuell sehr unterschiedlich sein. „Bei mir hat am stärksten Einfluss gehabt, dass meine Kontakte wissen, dass ich in der Wissenschaft im Bereich Cybersicherheit arbeite“, erzählt er. „Das heißt wenn ich vorschlage, einen Sicherheitsmechanismus auszuprobieren, kommt

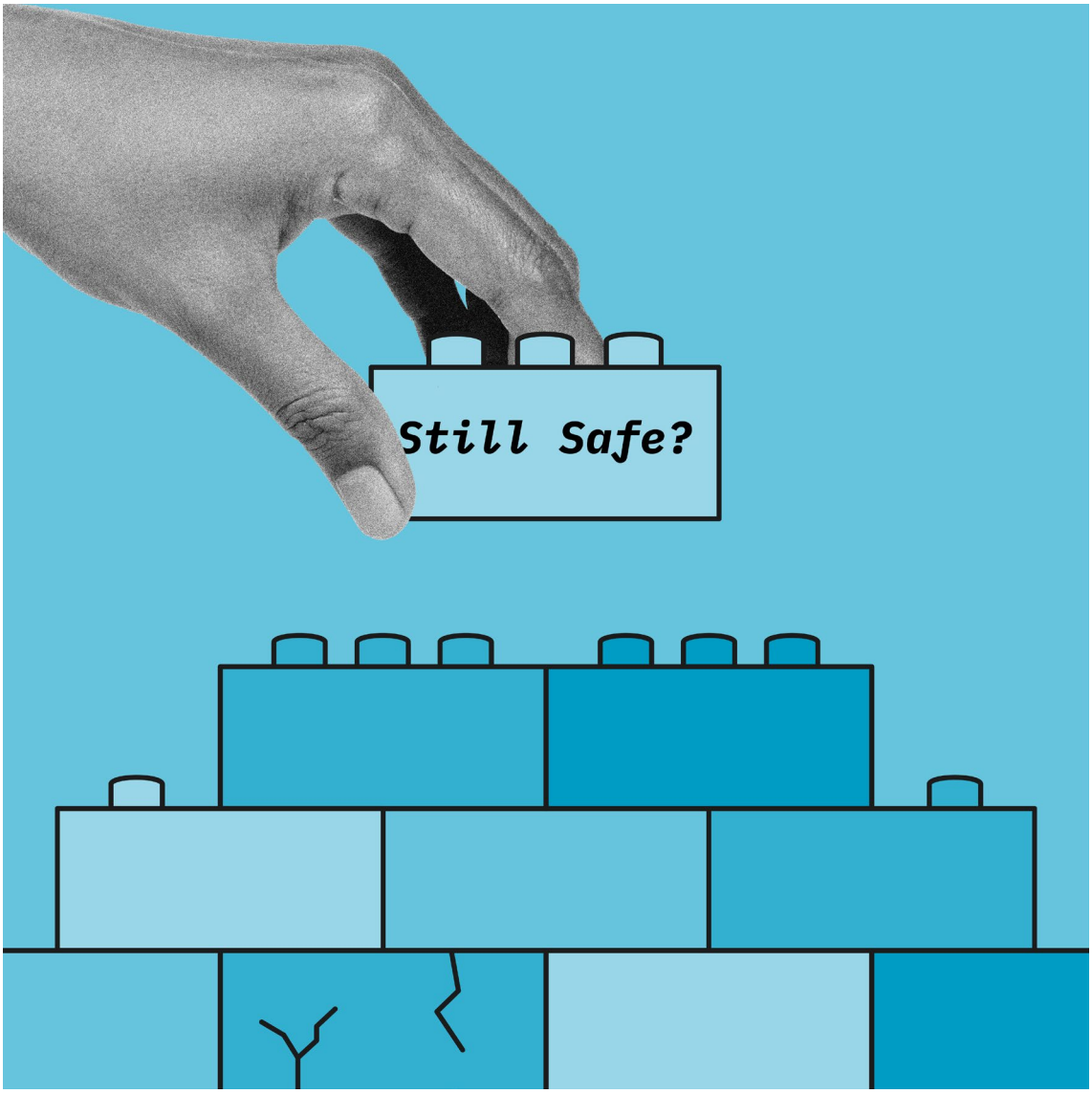
da sehr viel Autorität mit. Widerspruch mir gegenüber würde da mangelnde Wertschätzung ausdrücken.“ Hier zeigt sich, wie wichtig bei solch einem autoethnographischen Vorgehen die Einordnung der eigenen Erfahrung ist, wie er weiter fortführt: „Was ich in der Studie beschrieben habe, war aufgrund meiner persönlichen Faktoren vermutlich noch ein relativ positiver Ausblick. Andere Leute hätten es vermutlich noch viel schwerer gehabt, diese Zeremonien durchzuführen.“

Konsequenzen und mögliche Änderungen am Design

Ganz grundsätzlich ist Fassl wichtig, die enge Verzahnung von Sicherheitsthemen mit menschlichen Faktoren herauszustellen. „Ich glaube, dass hinter jeder technischen Sicherheit eigentlich immer ein menschlicher Faktor steht, der etwas beschützen oder etwas vermeiden will.“ Gerade bei den Authentifizierungszeremonien sei dies besonders wichtig, so der Forscher weiter: „Der Unterschied von normalen Sicherheitsmechanismen zu Authentifizierungszeremonien ist, dass wir erste oft nur für uns umsetzen. Letztere sind insofern ein Sonderfall, da wir zusammenarbeiten müssen, um unsere gemeinsame Sicherheit zu gewährleisten.“ Die Technik ist dabei immer nur ein Teil der Lösung, was der Begriff des Social-technical Gap zu umschreiben versucht. Der Social-technical Gap beschreibt den Unterschied zwischen dem, was die Technik erlaubt und dem, was User:innen auch umsetzen können und wollen. „In diesem Fall heißt das, dass die Authentifizierungszeremonie ja auch irgendwie in den Alltag und in Gespräche eingebaut werden muss“, erklärt Fassl.

„Meine Überlegungen gehen in die Richtung, den User:innen Organisationsaufwand abzunehmen“, so der CISPA-Forscher weiter. „Technische Unterstützung könnte dabei helfen den Social-technical Gap zu überbrücken. Das wäre unter anderem mit automatisierten Benachrichtigungen auf dem Smartphone zu passenden Zeitpunkten möglich. Die Personen könnten natürlich sagen, kein Bedarf. Aber es wäre eine praktische Erinnerung.“ Ideen, neue Lösungen selbst auszuprobieren, gibt es viele auf Seiten von Fassl und seinen Kolleg:innen. Und auch wenn aktuell kein konkretes Forschungsprojekt damit verknüpft ist, ist sich Fassl sicher: „Ich glaube das Thema ist noch nicht tot.“

Fassl, Matthias; Krombholz, Katharina (2023) Why I Can’t Authenticate – Understanding the Low Adoption of Authentication Ceremonies with Autoethnography. In: CHI23, 23-28 April 2023, Hamburg, Germany. Conference: CHI International Conference on Human Factors in Computing Systems



© Lea Mosbach

Gleich zwei Distinguished Paper Awards gab es im Jahr 2023 auf dem renommierten USENIX Security Symposium für Forschungspaper, an denen Alexander Dax mitgearbeitet hat. Der CISPA-Forscher und PhD-Student freut sich über so viel Zuspruch aus der Community. Eine der beiden in der Forschungsgemeinschaft begehrten Auszeichnungen hat er für sein Paper „Hash gone bad: Automated discovery of protocol attacks that exploit hash function weaknesses“ erhalten. Darin zeigt der Saarländer auf, dass automatisierte Sicherheitsanalysen von Internetprotokollen oft ungenau sind, weil sie von falschen Voraussetzungen – in diesem Fall perfekten Hashfunktionen – ausgehen. Er erklärt uns, was Hashfunktionen sind, wozu sie eingesetzt werden und wie er mit seiner Forschung die automatisierte Analyse der Protokolle verbessern will.

Automatisierte Protokollanalysen im Realitätscheck

Alexander Dax

Damit im Internet Daten sicher hin und her geschickt werden können, kommen verschiedene Protokolle zum Einsatz. Sie regeln, wer wann wem was und in welcher Form schicken darf. Eines der bekanntesten Internet-Protokolle im Dauereinsatz ist das sogenannte TLS, kurz für Transport Layer Security. Mit TLS wird vor allem geregelt, wie die Kommunikation zwischen Webanwendungen verschlüsselt wird. So kommunizieren zum Beispiel Browser wie Google Chrome und Mozilla bei jedem Aufruf einer Website mit einem Webserver. Damit diese Kommunikation nicht von Angreifer:innen unterwandert werden kann, muss im ersten Schritt vor der eigentlichen Kommunikation eine sichere Verbindung aufgebaut werden. So wird erstmal sichergestellt, dass die Kommunikationspartner:innen sind, wer sie vorgeben und nicht irgendein:e Dritte:r sich dazwischenschalten kann. Ist das geklärt, können kryptografische Schlüssel ausgetauscht werden und so eine vertrauliche Kommunikation ermöglichen. Soweit so gut, aber wie kann das jetzt sicher passieren?

Hashfunktionen als Sicherheitsgarant

„Nahezu jedes Sicherheitsprotokoll nutzt Hashfunktionen“, erklärt Dax. Damit lässt sich ein Prüfwert und damit eine Art digitaler Fingerabdruck erstellen. Mit diesem lässt sich prüfen, ob Daten auf dem Weg von A nach B manipuliert wurden. „Diese Funktionen nehmen irgendeinen Wert, egal welcher Größe, und machen daraus einen kleineren Wert mit fixer Größe“, erklärt Dax. Damit alleine lässt sich noch nicht viel anfangen, die Funktionen müssen zusätzlich bestimmte Eigenschaften aufweisen. „Dazu gehört, dass ein bestimmter Dateninhalt, etwa ein Passwort, mit derselben Hashfunktion berechnet, immer denselben Wert ergeben muss. Umgekehrt darf es aber nicht möglich sein, aus dem Hashwert auf den Dateninhalt zurückzuschließen.“ Eine weitere wichtige Eigenschaft von Hashfunktionen ist, dass verschiedene Ursprungsdaten nicht zum selben Hashwert umgerechnet werden dürfen. „Man spricht von Kollisionen, wenn das passiert“, sagt Dax. Und genau hier kommen sich Theorie und Praxis in die Quere. „In der Realität gibt es keine perfekten Hashfunktionen. Es ist immer nur eine Frage der Zeit, bis es Kollisionen gibt. Zudem hat sich der Stand der Technik geändert. Bei alten Hashfunktionen ist es mittlerweile

möglich, mit genug Rechenpower solange verschiedene Werte durchzuprobieren, bis der Ursprungswert für einen Hashwert herausgefunden ist. Das nennt man Brute-Force-Angriff“, sagt Dax.

Netze müssen zukunftssicher sein

Solche Angriffe sind für moderne Hashfunktionen laut Dax sehr aufwendig und daher bislang kein alltägliches Problem. „Allerdings entwickelt sich die Technik sehr schnell weiter und wir müssen dafür sorgen, dass unsere Netze auch zukunftssicher sind.“ Und damit kommt Dax’ Forschung rund um Tools für die automatisierte Sicherheitsanalyse von Protokollen ins Spiel. „Es reicht nicht, zu behaupten, dass ein Protokoll sicher ist. Wir müssen es auch formal beweisen können. Das heißt, wir brauchen präzise mathematische Definitionen davon, wie sich das Protokoll verhält und dann lässt sich berechnen, wie sicher es ist.“ Diese Prüfverfahren sind enorm aufwendig, weshalb sie mittlerweile automatisiert wurden. „Es gibt dazu Tools wie zum Beispiel den Tamarin Prover oder Proverif, die die Arbeit für uns übernehmen können. Das Problem ist: Diese Tools arbeiten bislang häufig nur mit modellhaften Abbildungen von Hashfunktionen, die in dieser Form perfekt sind. Wir wissen aber, dass sie es in der Praxis oft eben nicht sind.“

»Die Technik entwickelt sich sehr schnell weiter und wir müssen dafür sorgen, dass unsere Netze auch zukunftssicher sind.«

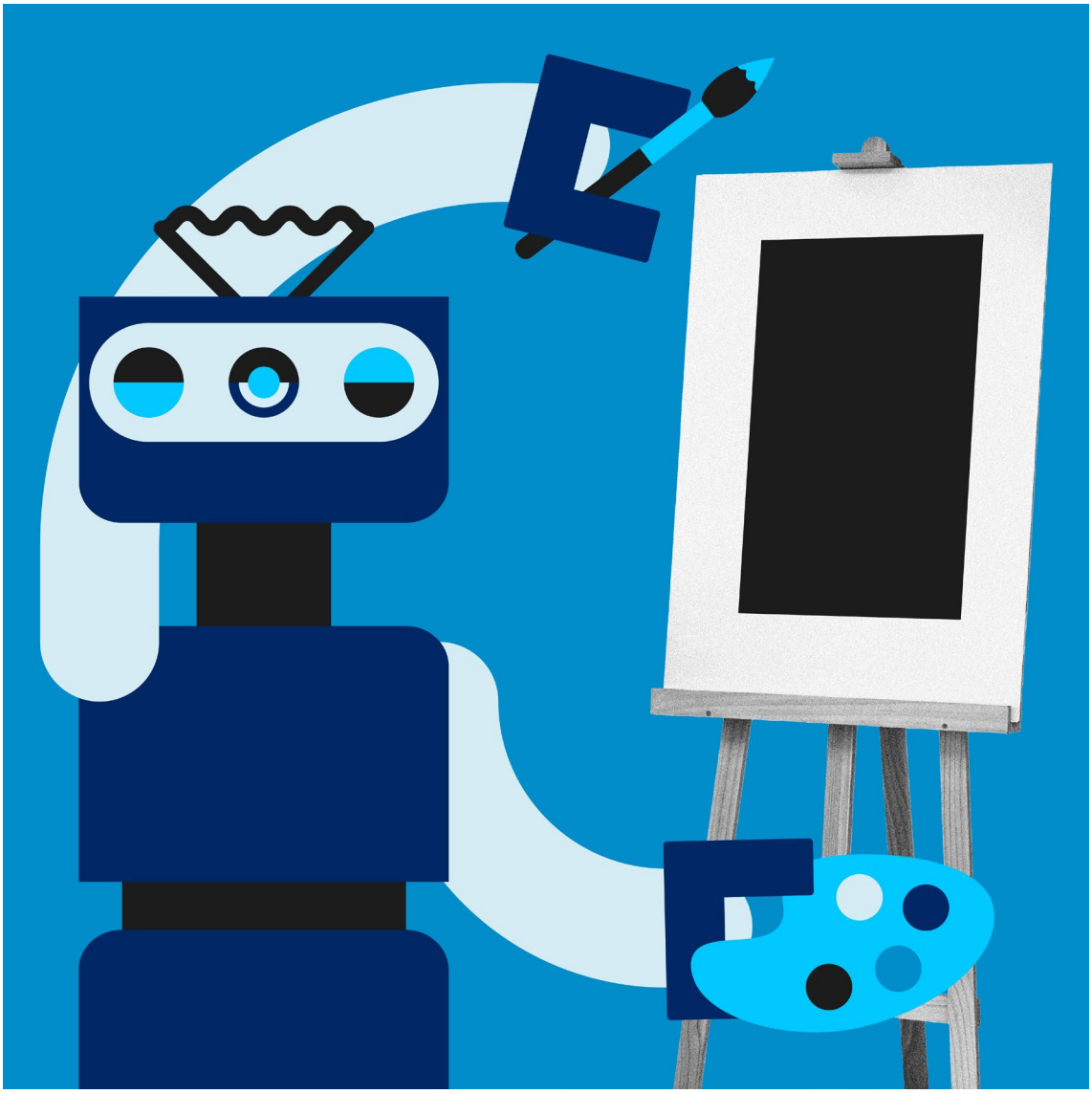
Zu perfekt ist auch nicht gut

Das anzuerkennen, ist der erste Schritt zur Verbesserung der Tools. Und es hat noch einen weiteren Vorteil: „Wir haben verschiedene Varianten von schwachen Hashfunktionen modelliert und in den Tamarin Prover und das Tool Proverif eingebaut. Wir wollen so auch herausfinden, wie groß der Einfluss von verschiedenen Schwächen in den Hashfunktionen auf die Gesamtsicherheit des Protokolls ist.“ Formale Sicherheitsbeweise von Protokollen sind dabei kein nerdiger Forschenden-Kram, sondern längst auch in den großen Tech-Unternehmen der Welt angekommen. „Viele große Unternehmen wie zum Beispiel Google beschäftigen Kryptografen, um zu prüfen, wie sicher die eingesetzten Protokolle sind. Das ist manuell sehr aufwendig und selbst die Kontrolle von automatisierten Sicherheitsanalysen erfordert derzeit noch viel Aufwand. Wir wollen die Tools so gut machen, dass dafür künftig deutlich weniger Personal und Aufwand erforderlich ist und die automatisierte Prüfung echte Protokoll-Sicherheit garantieren kann.“

An der Quelle

Dax arbeitet in der Gruppe von CISPA-Faculty Prof. Dr. Cas Cremers und sitzt somit an der Quelle für Forschungsfragen rund um die automatisierte Prüfung von Protokollen. Cremers und Kolleg:innen haben vor einigen Jahren den bereits erwähnten Tamarin Prover entwickelt, der von Unternehmen wie Mozilla und Amazon genutzt wird. „Meine Forschung ist Teil eines größeren Projektes zur Verbesserung der automatisierten Sicherheitsanalyse. Ich arbeite schon seit Jahren daran mit. Dass meine Forschung zu Hashfunktionen jetzt in ein ausgezeichnetes Paper gemündet ist, ist toll“, sagt Dax. Er ist mittlerweile eine Art CISPA-Urgestein. „Ich bin schon seit 2016 mit dabei, war zunächst Hiwi bei Michael Backes, dann in der Gruppe von Robert Künnemann und jetzt bin ich als Doktorand bei Cas. Irgendwie bin ich mit dem CISPA mitgewachsen.“

Cheval, Vincent; Cremers, Cas; Dax, Alexander; Hirschi, Lucca; Jacomme, Charlie; Kremer, Steve (2023) Hash Gone Bad: Automated discovery of protocol attacks that exploit hash function weaknesses. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium



© Lea Mosbach

Im Jahr 2023 haben KI-Bildgeneratoren eine nie dagewesene Popularität erfahren. Mit wenigen Klicks lassen sich alle Arten von Bildern erstellen: Auch menschenverachtendes Bildmaterial und Hass-Memes können dazu gehören. CISPA-Forscherin Yiting Qu aus dem Team von CISPA-Faculty Dr. Yang Zhang hat nun untersucht, wie hoch der Anteil dieser Bilder unter den bekanntesten KI-Bildgeneratoren ist und wie sich deren Erstellung mit effektiven Filtern verhindern lässt. Das dazugehörige Paper „Unsafe Diffusion: On the Generation of Unsafe Images and Hateful Memes From Text-To-Image Models“ wurde bei der renommierten ACM Conference on Computer and Communications Security (CCS) publiziert.

Neu entwickelter Filter soll verhindern, dass KI-Bildgeneratoren „unsichere Bilder“ verbreiten

Yiting Qu

Wenn heute von KI-Bildgeneratoren die Rede ist, dann geht es häufig um sogenannte Text-zu-Bild-Modelle. Dies bedeutet, dass Nutzer:innen durch die Eingabe bestimmter Textinformationen in ein KI-Modell ein digitales Bild generieren lassen. Die Art der Texteingabe bestimmt dabei nicht nur den Inhalt des Bildes, sondern auch den Stil. Je umfangreicher das Trainingsmaterial des KI-Bildgenerators war, umso mehr Möglichkeiten der Bildgenerierung haben die Nutzer:innen. Zu den bekanntesten Text-zu-Bild-Generatoren gehören Stable Diffusion, Latent Diffusion oder DALL·E. „Die Menschen verwenden diese KI-Tools, um alle Arten von Bildern zu zeichnen“, erzählt die CISPA-Forscherin Yiting Qu. „Ich habe allerdings festgestellt, dass einige diese Tools auch nutzen, um etwa pornografische oder verstörende Bilder zu erzeugen. Die Text-zu-Bild-Modelle bergen also ein Risiko in sich.“ Problematisch werde es vor allem dann, wenn diese Bilder an Mainstream-Plattformen weitergegeben werden und dort eine breite Zirkulation erfahren.

Der Begriff „unsichere Bilder“

Für den von Qu und ihren Kolleg:innen beobachtenden Umstand, dass die KI-Bildgeneratoren mit einfachen Anweisungen dazu gebracht werden können, Bilder menschenverachtenden oder pornografischen Inhalts zu generieren, arbeiten sie mit dem Begriff „unsichere Bilder“. „Derzeit gibt es in der Forschungsgemeinschaft keine allgemeingültige Definition, was ein unsicheres Bild ist und was nicht. Daher haben wir einen datenbasierten Ansatz verfolgt, um zu definieren, was unsichere Bilder sind“, erklärt Qu. „Für unsere Analyse haben wir mit Hilfe von Stable Diffusion Tausende von Bildern generiert“, so die Forscherin weiter. „Die haben wir dann gruppiert und auf der Grundlage ihrer Bedeutungen in verschiedene Cluster eingeteilt. Die wichtigsten fünf Cluster beinhalten Bilder mit sexuell expliziten, gewalttätigen, verstörenden, hasserfüllten und politischen Inhalten.“ Um das Risiko der Generierung menschen-

verachtenden Bildmaterials durch KI-Bildgeneratoren konkret quantifizieren zu können, fütterten Qu und ihre Kolleg:innen im Anschluss vier der bekanntesten KI-Bildgeneratoren, Stable Diffusion, Latent Diffusion, DALL·E 2 und DALL·E mini, mit spezifischen Sets hunderter von Texteingaben, den sogenannten Prompts. Die Sets von Texteingaben stammten aus zwei Quellen: der in rechtsextremen Kreisen beliebten Online-Plattform 4chan sowie der Lexica-Website. „Wir haben uns für diese beiden entschieden, da sie bereits in früheren Arbeiten zur Untersuchung von unsicheren Online-Inhalten verwendet wurden“, erklärt Qu. Ziel war herauszufinden, ob die Bild-Generatoren aus diesen Prompts sogenannte „unsichere Bilder“ erzeugen oder nicht. Das Ergebnis war, dass über alle vier Generatoren hinweg 14,56 Prozent aller generierten Bilder in die Kategorie „unsichere Bilder“ fielen. Mit 18,92 Prozent lag der Anteil bei Stable Diffusion am höchsten.

Filterfunktionen blockieren Bildgenerierung

Eine Möglichkeit, die Verbreitung von menschenverachtendem Bildmaterial zu verhindern, besteht darin, die KI-Bildgeneratoren so zu programmieren, dass sie dieses Bildmaterial gar nicht erst herstellen beziehungsweise diese Bilder nicht ausgeben. „Ich kann am Beispiel von Stable Diffusion erklären wie das funktioniert“, erzählt Qu. „Sie definieren mehrere unsichere Wörter, wie etwa Nacktheit. Wenn dann ein Bild erzeugt wird, wird der Abstand zwischen dem Bild und dem als unsicher definierten Wort, wie etwa Nacktheit, berechnet. Wenn dieser Abstand kleiner als ein Schwellenwert ist, wird das Bild durch ein schwarzes Farbfeld ersetzt.” Dass in Qus Studie von Stable Diffussion so viele „unsichere Bilder“ erzeugt wurden zeigt, dass die existierenden Filter ihre Aufgabe nicht zufriedenstellend lösen. Aus diesem Grund entwickelte die Forscherin einen eigenen Filter, der im Vergleich mit einer wesentlich höheren Trefferquote punkten kann.

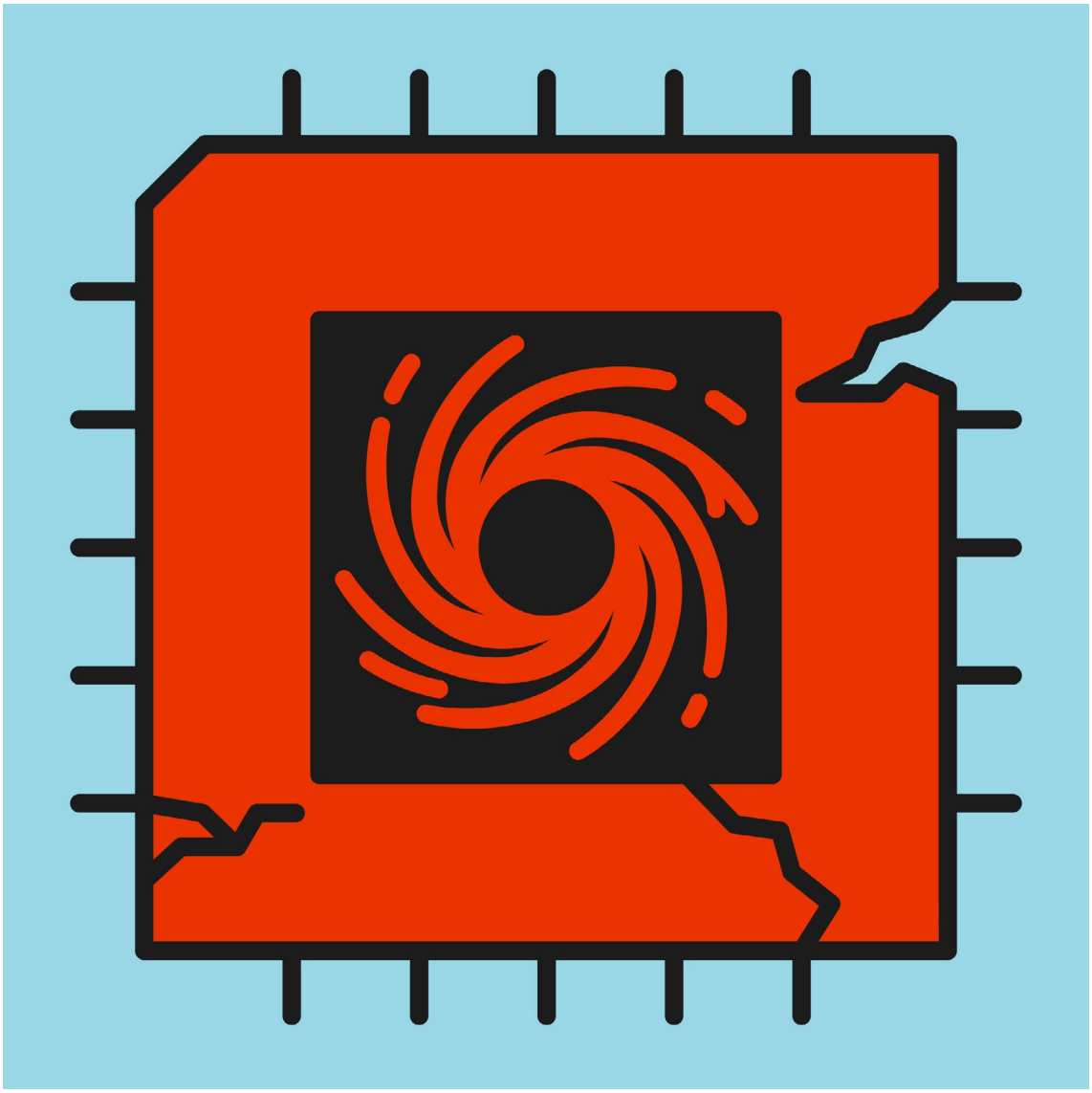
Die Verhinderung der Bildgenerierung ist jedoch nicht die einzige Möglichkeit, wie Qu erklärt: „Wir schlagen drei Abhilfemaßnahmen vor, die der Lieferkette von Text-zu-Bild-Modellen folgen. Zunächst sollten Entwickler:innen in der Trainings- oder Abstimmungsphase die Trainingsdaten kuratieren, also die Anzahl unsicherer Bilder reduzieren.“ Denn „unsichere Bilder“ in den Trainingsdaten seien der Hauptgrund, warum das Modell später Risiken birgt. „Darüber hinaus können Entwickler:innen die Eingabeaufforderungen der Nutzer:innen regulieren, zum Beispiel durch die Entfernung unsicherer Schlüsselwörter.“ Die dritte Möglichkeit betrifft die Verbreitung nach der Bildgenerierung. „Sind unsichere Bilder bereits generiert, muss es eine Möglichkeit geben, diese Bilder zu klassifizieren und online löschen zu können.“

Für letzteres wiederum bräuchte es dann Filterfunktionen für die Plattformen, auf denen diese Bilder zirkulieren. Bei all diesen Maßnahmen besteht die Herausforderung darin, das richtige Maß zu finden. „Es braucht einen Kompromiss zwischen Freiheit und Sicherheit des Inhalts. Aber wenn es darum geht, zu verhindern, dass diese Bilder auf Mainstream-Plattformen breite Zirkulation erfahren, halte ich eine strenge Regulierung für sinnvoll“, so die CISPA-Forscherin. Qu hofft, mit ihrer Forschung dazu beitragen zu können, dass in Zukunft weniger Hass-Bilder im Internet zirkulieren.

»Sind unsichere Bilder bereits generiert, muss es eine Möglichkeit geben, diese Bilder zu klassifizieren und online löschen zu können.«

Qu, Yiting; Shen, Xinyue; He, Xinlei; Backes, Michael; Zannettou, Savvas; Zhang, Yang (2023) Unsafe Diffusion: On the Generation of Unsafe Images and Hateful Memes From Text-To-Image Models. In: CCS 2023, 26-30 Nov 2023, Copenhagen, Denmark. Conference: CCS ACM Conference on Computer and Communications Security

Forscherin: Yiting Qu
Autor: Felix Koltermann



© Lea Mosbach

Mit sogenannter Secure Encrypted Virtualization (SEV) will der Chiphersteller AMD vor allem Cloud-Dienste sicherer machen. Doch auch auf die aktuellsten Versionen des Sicherheitsfeatures, SEV-ES (Encrypted State) und SEV-SNP (Secure Nested Paging), war bis vor Kurzem noch ein softwarebasierter Fehler-Angriff möglich. Entdeckt hat das CISPA-Forscher Ruiyi Zhang, der im Team von CISPA-Faculty Dr. Michael Schwarz forscht. Die von ihm konstruierte Angriffsart nennt sich CacheWarp und ermöglicht Angreifer:innen im schlimmsten Fall umfassende Zugriffsrechte auf Daten und sogar die Möglichkeit, sie zu verändern. AMD hat die Lücke nach eigenen Angaben durch ein Update geschlossen. Sein Paper „CacheWarp: Software-based Fault Injection using Selective State Reset“ stellt Zhang 2024 auf dem USENIX Security Symposium vor.

Schwachstelle in AMD-Sicherheitsfeature entdeckt

Ruiyi Zhang

Die Nutzung großer Cloud-Plattformen boomt. „Cloud-Dienste erlauben es Unternehmen, flexibel Rechenpower und Speicherplatz einzukaufen, wann immer sie es brauchen“, erklärt Riuyi Zhang. Die Sicherheit der Dienste ist essentiell, wurde in der Vergangenheit aber bereits durch entdeckte Schwachstellen und potentielle Angriffsmöglichkeiten erschüttert. „Die Grundlage von Clouddiensten ist die sogenannte Virtualisierung, mit der sich Hardwarekomponenten und damit verbunden auch Personal einsparen lassen“, sagt Zhang. Bei der Virtualisierung werden laut dem Forscher innerhalb eines physischen Rechners mehrere virtuelle Maschinen erstellt. Virtuelle Maschinen sind quasi softwarebasierte Rechner, die alles haben, was ein normaler Computer auch hat: einen eigenen Arbeitsspeicher, eine CPU, ein eigenes Betriebssystem.
Mit Virtualisierung lassen sich also aus einem Rechner mit entsprechender Rechenpower viele machen.

Sicherheitsfeature mit Schwächen

Für die Verteilung der Ressourcen und die entsprechende Trennung von Prozessen ist der sogenannte Hypervisor zuständig. Es handelt sich dabei um eine Software, die die Ressourcen wie Arbeitsspeicher und Rechenleistung verteilt und die Betriebssysteme isoliert. Der Hypervisor fungiert also als eine Art Host für die virtuellen Maschinen. Da dieser nicht zum Angriffspunkt werden darf, hat der Prozessorhersteller AMD die erste Generation von Secure Encrypted Virtualization (SEV) vorgestellt. Die Idee hinter SEV: Für jede laufende virtuelle Maschine wird der Arbeitsspeicher mit einem separaten Schlüssel verschlüsselt, was einen übergreifenden Datenzugriff und den Zugriff durch einen nicht-vertrauenswürdigen oder von Angreifer:innen übernommenen Hypervisor unmöglich machen soll. „Schnell wurden mehrere Sicherheitslücken bekannt. Zudem wurde Verschlüsselung bei SEV-ES und SEV anfänglich ohne Identitätsprüfung verwendet. Dadurch konnten Daten manipuliert werden. Und nicht alle Teile des Speichers waren verschlüsselt“, erklärt Michael Schwarz. Der CISPA-Faculty ist Experte für Sicherheitslücken in CPUs und war an der Entdeckung von mehreren solcher Lücken beteiligt, darunter Spectre, Meltdown und ZombieLoad. AMD reagierte auf die Probleme und entwickelte SEV weiter zu den Feautures SEV-ES (Encrypted State) und zuletzt

SEV-SNP (Secure Nested Paging). Laut AMD verspricht SEV-SNP eine starke Speicherintegrität. Hypervisor-Attacken sollten damit unterbunden werden.

Ein paar Zeilen Code

Etwa eine halbe Minute, Zugang zu einem Serverraum und ein paar Zeilen Code bräuchte Zhang, um sich Zugang auf alle virtuellen Maschinen zu verschaffen und dort mit Administratorrechten alles einzusehen und zu verändern, was er möchte. Herauszufinden, wie genau das möglich ist, erforderte monatelange Arbeit. „Cache Warp ist unseres Wissens nach bislang der einzige softwarebasierte Angriff, mit dem SEV-SNP derartig ausgehebelt werden kann“, erklärt Zhang.

»Cache Warp ist unseres Wissens nach bislang der einzige softwarebasierte Angriff, mit dem SEV-SNP derartig ausgehebelt werden kann.«

Ein Computer geht auf Zeitreise

„Zunächst müssen wir uns in ein System einloggen können. Dabei hilft uns eine Technik, die wir TimeWarp genannt haben“, sagt Schwarz. Sie nutzt laut dem Forscher aus, dass Computer sich in gewissen Situation im Speicher merken, welchen Code sie als nächstes ausführen müssen. „Wir können rückgängig machen, was der Computer sich als nächsten Schritt gespeichert hat. Dadurch führt der Computer Code aus, den er davor schon einmal ausgeführt hat, weil er eine veraltete sogenannte Rücksprungadresse aus dem Speicher liest. Der Computer reist somit in der Zeit zurück. Allerdings wird der alte Code mit neuen Daten ausgeführt, was zu unvorhergesehenen Effekten führt. Wenn man diese Technik clever einsetzt, kann man so die Programmlogik verändern“, erklärt Schwarz. Zhang fügt hinzu: „Mit Hilfe von TimeWarp ändern wir die Programmlogik einer virtuellen Machine also so, dass sie uns das Einloggen erlaubt, ohne das Passwort zu kennen.“

Die 0 steht für Erfolg

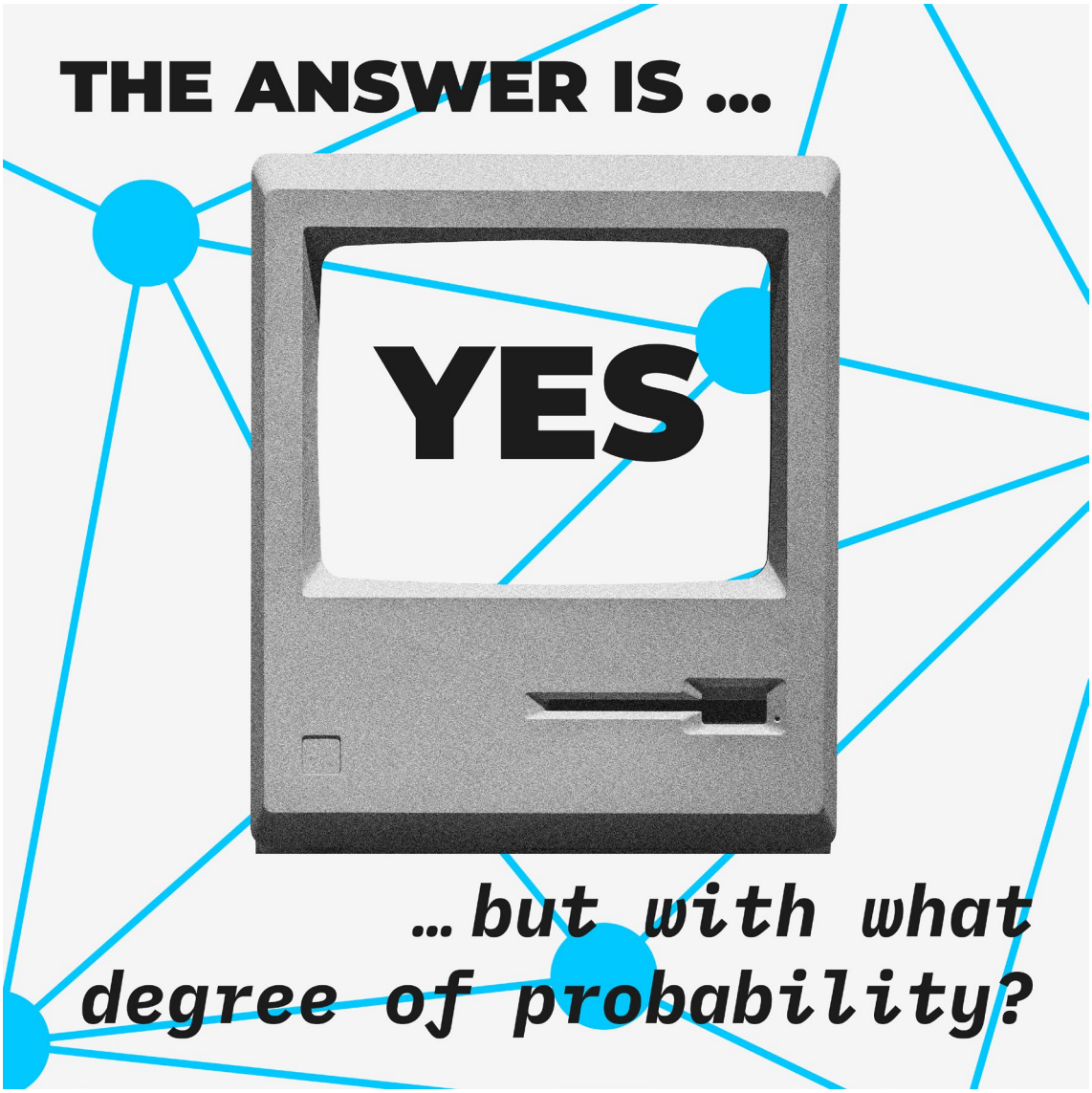
Kombiniert mit der zweiten Technik, dem sogenannten „Dropforge“, ist es möglich, auch den Cache zu manipulieren und dort Veränderungen an Daten rückgängig machen. „Auch wenn es nicht intuitiv erscheint, kann man damit auch Administratorrechte bekommen. Dabei nutzt man Details in der Programmlogik aus“, sagt Schwarz. In der Informatik stehe eine „0“ oft dafür, dass etwas erfolgreich war. Andere Zahlen stehen hingegen für etwaige Fehlercodes. „0“ ist laut Schwarz jedoch auch der Standardwert von Daten, wenn nichts anderes gespeichert wird. „Wenn ein System testet, ob der jeweilige Nutzer Administrator ist oder nicht, dann liefert die Abfrage auch ‚0‘ zurück, wenn man Administrator ist. Ist man kein Administrator, wird eine andere Zahl zurückgegeben. Mit ‚Dropforge‘ kann man diese Antwort rückgängig machen. Egal, ob man Administrator ist oder nicht, es steht der initiale Wert ‚0‘ im Speicher. Das System nimmt dann an, dass man Administrator ist“, erklärt Schwarz. „Mit dieser Kombination haben wir uneingeschränkten Zugriff auf die virtuelle Maschine“, ergänzt Zhang.

Vertrauen ist gut

In ihrem Paper „CacheWarp: Software-based Fault Injection Selective State Reset“ haben die Forscher nicht nur die Angriffstechniken beschrieben, sondern auch eine Lösung zur Entschärfung der Angriffsmöglichkeiten vorgeschlagen. Zudem wollen sie ein Testing-Tool für die Schwachstellen Open Source zur Verfügung stellen. „Wir wollen uns nicht auf die Aussage verlassen, dass etwas sicher ist. Wir wollen es prüfen können“, erklärt Schwarz. Seit Entdeckung von CacheWarp sind die Forschenden auch im Austausch mit AMD: Der Hersteller hat ihnen gegenüber angegeben, die Lücke inzwischen geschlossen zu haben.

Zhang, Ruiyi; Gerlach, Lukas; Weber, Daniel; Hetterich, Lorenz; Lü, Youheng; Kogler, Andreas; Schwarz, Michael (2024) CacheWarp: Software-based Fault Injection using Selective State Reset. In: 32nd USENIX Security Symposium, 9-11 Aug 2024, Anaheim, CA, USA. Conference: USENIX Security Symposium

Forscher: Ruiyi Zhang
Autorin: Annabelle Theobald



© Lea Mosbach

Viele Methoden des maschinellen Lernens (ML) könnten ganze Bereiche unseres Lebens revolutionieren. Damit ein ML-Algorithmus vertrauenswürdig sein kann, müssen die Nutzer:innen jedoch wissen, wie sicher das Modell in Bezug auf die Vorhersage ist. Bislang zählt bei der Bewertung vor allem die Genauigkeit. Die gibt jedoch keine Auskunft darüber, mit welcher Wahrscheinlichkeit ein Modell einzelne Eingaben verarbeitet. Die Herausforderung steigt, wenn es um komplexe Datennetze mit vielen Verbindungen geht, wie bei der Arzneimittelforschung oder der medizinischen Diagnostik. Dabei können genau diese Verbindungen in den Daten ein besseres Verständnis der Wahrscheinlichkeit fördern. CISPA-Forscher Soroush H. Zargarbashi hat für sein auf der International Conference on Machine Learning (ICML) 2023 vorgestelltes Paper „Conformal Prediction Sets for Graph Neural Networks“ nun eine neue Methode zur Unsicherheitsquantifizierung getestet.

Neues Verfahren zur Unsicherheitsquantifizierung von Anwendungen maschinellen Lernens

Soroush Zargarbashi

Die technische Grundlage vieler Anwendungen des maschinellen Lernens bilden sogenannte Graph Neural Networks (GNN). „In vielen realen Szenarien haben wir es mit Graphen zu tun. Zwischen den Datenpunkten gibt es sinnvolle Verbindungen, und mit GNN berücksichtigen wir diese Verbindungen“, erklärt CISPA-Forscher Soroush H. Zargarbashi. Graphen sind eine Art abstrakte Datenstruktur, die aus zwei Elementen besteht, den Knoten und den Verbindungen zwischen den Knoten, den sogenannten Kanten. Graphen können zum Beispiel soziale Netzwerke, Sensorennetze, wissenschaftliche Arbeiten mit ihren Referenzen oder ähnliches modellieren. Dabei gibt es eine Besonderheit, die bei bestimmten Anwendungsbereichen zu Problemen führt, so Zargarbashi weiter: „Wenn wir ein Modell als Blackbox verwenden, führt eine Eingabe immer zu einer Ausgabe, zum Beispiel wenn ein Auto eine Situation erkennt und beschließt, nach links zu steuern. Wenn man aber nicht weiß, wie sicher das Modell in Bezug auf diese bestimmte Ausgabe ist, wird es höchst unzuverlässig, insbesondere in sicherheitskritischen Bereichen, in denen der Benutzer eine Unsicherheitsabschätzung des Modells benötigt.“ Problematisch ist, dass die Vorhersagequalität der Modelle von den Modellen oft als besser angegeben wird, als sie ist und der Unsicherheitsfaktor der Prognosen zu niedrig angegeben wird.

Warum es wichtig ist, dass Modelle eine zuverlässige Unsicherheitsschätzung für ihre Aussagen liefern, illustriert Zargarbashi an einem Beispiel: „Nehmen wir an, ein Arzt verwendet ein medizinisches Diagnosesystem, um zu entscheiden, ob ein Patient eine bestimmte Krankheit hat. Dann ist es sehr wichtig, dass das Modell dies mit hoher Sicherheit vorhersagt. Denn wenn das Modell dies nicht leisten kann, muss man weitere Diagnosen durchführen. Diese Vorhersagen zu präzisieren, ist die Idee hinter der Quantifizierung der Unsicherheit.“ Ein Beispiel sind Verfahren, bei denen KI eingesetzt wird, um durch die automatische Analyse von MRT-Bildern festzustellen, ob ein Organ krebsartig ist oder nicht. In diesem Fall ist es wichtig, die Qualität der Vorhersage für jede

einzelne Eingabe zu kennen. Mit anderen Worten: Eine 90-prozentige Trefferquote kann für die anderen 10 Prozent sehr riskant sein.

Conformal Prediction als Lösung für die Unsicherheitsquantifizierung

„Es gibt Methoden, um diese Unsicherheit zu quantifizieren, aber sie sind rechenintensiv, schwer anzuwenden und, was am schlimmsten ist, viele von ihnen funktionieren nicht bei Graphen“, sagt Zargarbashi. Viele dieser Methoden erfordern in der Regel Änderungen an der Modellarchitektur oder ein neues Training des Modells. „Es gibt jedoch ein wachsendes Interesse an einem alternativen Ansatz, der als Conformal Prediction bekannt ist“, so der CISPA-Forscher weiter. Conformal Prediction (CP) ist ein seit Ende der 1990er-Jahre bekanntes statistisches Verfahren zur Erstellung von Vorhersagesätzen, ohne Annahmen über den Vorhersagealgorithmus treffen zu müssen. Zargarbashi erläutert, dass CP wie eine Art Hülle um das Modell herum arbeitet und eine Prognose mit einer vom Benutzer festgelegten Wahrscheinlichkeitsgarantie für die richtige Antwort liefert. Aber wie genau funktioniert das? „Für einen neuen Patienten beispielsweise kann man den Algorithmus so einstellen, dass er Sätze erzeugt, die mit einer Wahrscheinlichkeit von 95 Prozent die richtige Antwort liefern. Das funktioniert für jedes Modell, auch für solche, die nur zu 60 Prozent richtig sind“, erklärt der CISPA-Forscher. „Man braucht nur eine Zufallsstichprobe früherer Patienten mit ihrer richtigen Diagnose. Auf diese Weise haben wir für jeden Patienten eine Reihe möglicher Diagnosen, von denen wir wissen, dass sie mit sehr hoher Wahrscheinlichkeit die richtige Antwort enthalten.“

Maschinelles Lernen vertrauenswürdig machen

Im Grunde ist die von Zargarbashi und seinen Kollegen entwickelte Methode eine Variante der Unsicherheitsquantifizierung, die nicht nur mit Graphen arbeitet, sondern auch die Informationen aus den Beziehungen zwischen Datenpunkten nutzt. Ihre Methode ist nicht rechenintensiv und einfach zu implementieren, sofern zusätzliche Daten verfügbar sind. Entscheidender Vorteil dieser Herangehensweise ist laut Zargarbashi, dass CP „modellunabhängig ist, was bedeutet, dass es egal ist, welches Modell verwendet wird. Man muss also Modelle nicht von Grund auf neu trainieren.“ Um die Studie praktisch umsetzen zu können, war die Entwicklung einer Methode namens „Diffusion Adaptive Prediction Sets“ notwendig. Es nutzt die Verbindungen zwischen den Datenpunkten, um die Qualität der Unsicherheitsabschätzung zu verbessern. Eingebettet ist die ausführliche empirische Analyse dieser Methode im veröffentlichten Paper in eine umfassende theoretische Studie darüber, wann CP für GNNs anwendbar ist. Mit ihrer Studie leisten Zargarbashi und seine Kollegen einen wichtigen Beitrag

dazu, wie auf Graphen basierende Modelle maschinellen Lernens vertrauenswürdiger gemacht werden können.

»Es gibt Methoden, um diese Unsicherheit zu quantifizieren, aber sie sind rechenintensiv, schwer anzuwenden und, was am schlimmsten ist, viele von ihnen funktionieren nicht bei Graphen.«

Zargarbashi, Soroush H.; Antonelli, Simone; Bojchevski, Aleksandar (2023) Conformal Prediction Sets for Graph Neural Networks. In: Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202. Conference: International Conference on Machine Learning

ÜBER DAS CISPA

Das CISPA Helmholtz-Zentrum für Informationssicherheit ist eine Großforschungseinrichtung des Bundes innerhalb der Helmholtz-Gemeinschaft. CISPA-Wissenschaftler:innen erforschen die Informationssicherheit in all ihren Facetten. Sie betreiben modernste Grundlagenforschung sowie innovative anwendungsorientierte Forschung und arbeiten an den drängenden Herausforderungen der Cybersicherheit, der künstlichen Intelligenz und des Datenschutzes. CISPA-Forschungsergebnisse finden Einzug in industrielle Anwendungen und Produkte, die weltweit verfügbar sind. Damit stärkt das CISPA die Konkurrenzfähigkeit Deutschlands und Europas.

Das CISPA bietet ein Forschungsumfeld von Weltrang und stellt einer großen Zahl an Forscher:innen umfangreiche Ressourcen zur Verfügung. Darüber hinaus fördert das CISPA in besonderem Maße auch die grundständige und postgraduale Bildung von Cybersicherheitsstudierenden. Das Zentrum hat sich zum Ziel gesetzt, eine Kaderschmiede für die nächste Generation an Cybersicherheitsexpert:innen und wissenschaftlichen Führungskräften in diesem Bereich zu werden. Das CISPA ist in Saarbrücken und St. Ingbert situiert. Die Lage des Zentrums in direkter Nachbarschaft zu Frankreich und Luxemburg ist ideal für grenzüberschreitende Kollaborationen mit anderen Forschungsinstitutionen.

Aktuell konzentriert sich unsere Forschung auf die folgenden sechs Forschungsbereiche:

Algorithmische Grundlagen und Kryptographie

Vertrauenswürdige Informationsverarbeitung

Verlässliche Sicherheitsgarantien

Erkennung und Vermeidung von Cyberangriffen

Sichere vernetzte und mobile Systeme

Empirische und verhaltensorientierte Sicherheit

IMPRESSUM

CISPA – Helmholtz-Zentrum
für Informationssicherheit gGmbH
Stuhlsatzenhaus 5
66123 Saarbrücken, Deutschland

Herausgeber

Sebastian Klöckner

Verantwortliche Redaktion

Felix Koltermann,
Eva Michely,
Annabelle Theobald

Redaktion

Lea Mosbach,
Janine Wichmann-Paulus

Illustration

Janine Wichmann-Paulus

Gestaltung

Stephanie Bremerich,
Tobias Ebelshäuser

Fotografie

Mai 2024

Stand des Impressums

T: +49 681 87083 2867
M: pr@cispa.de
W: https://cispa.de/

Kontakt Corporate Communications

Neuer Ansatz verbessert automatisierte Schwachstellensuche in Prozessoren

Warum visuelle digitale Zertifikate bislang nur theoretisch sicher sind

Entwicklung eines Open-Source-Prototyps für die 2-Faktor-Authentifizierung

Neue Spezifikationssprache revolutioniert automatisierte Softwaretests

Ein neues digitales System zur Verteilung humanitärer Hilfe vereint Datenschutz und Rechenschaftspflicht

Der neue Goldstandard: Differential Privacy weitergedacht

Key-Management wird bei Krypto-Fonds zur Herausforderung

Betreiber:innen von Websites nehmen Sicherheit wichtiger als Datenschutz

Auffällig im All: Eine Studie zur Satellitensicherheit

Collide+Power: Neuer Seitenkanalangriff betrifft alle Prozessoren

MobileAtlas: Eine Kartografie der Mobilfunk-Sicherheit

Ein neuer Standard? Die Nutzung von Web-Archiven für Live-Analysen zur Sicherheit von Websites

Test eines neuen Verfahrens zum Schutz vor Deepfakes

Ein Selbstversuch zeigt Schwierigkeiten beim Durchführen von Authentifizierungszeremonien

Automatisierte Protokollanalysen im Realitätscheck

Neu entwickelter Filter soll verhindern, dass KI-Bildgeneratoren „unsichere Bilder“ verbreiten

Schwachstelle in AMD-Sicherheitsfeature entdeckt

Neues Verfahren zur Unsicherheitsquantifizierung von Anwendungen maschinellen Lernens

CISPA
HELMHOLTZ-ZENTRUM FÜR INFORMATIONSSICHERHEIT