SecHUMAN
FORSCHUNGSKOLLEG
SICHERHEIT FÜR MENSCHEN IM CYBERSPACE

1. Forschungsthema

Innovative, produktivitätsfreundliche IT-Sicherheitslösungen und Risikomanagement in Organisationen

(Betreuerteam: Prof. M. Angela Sasse, Prof. Annette Kluge)

Abstract

Ziel des Projektes ist es, die Effektivität, Wirtschaftlichkeit und Nachhaltigkeit von IT-Sicherheitsmaßen auf den Prüfstand zu stellen und systematisch zu verbessern. Die konkreten Auswirkungen, die IT-Sicherheitsmaßnahmen auf die Produktivität und das Verhalten in den Organisationen haben, werden identifiziert und quantifiziert. Hierbei gilt es insbesondere IT-Sicherheitsmaßnahmen nicht nur kurz-, sondern langfristig zu prüfen und dabei aufzuzeigen, wie Organisationen lernen können, die Maßnahmen zu verbessern ohne der Produktivität des Unternehmens zu schaden oder Innovationen zu blockieren.

Traditionell hat bei der Entwicklung von IT-Sicherheitsmaßnahmen die Wirtschaftlichkeit dieser Maßnahmen keine große Rolle gespielt. Die Qualität einer Sicherheitsmaßnahme wird danach beurteilt, wie effektiv sie zur Abwehr eines Angriffs oder Schließung einer Schwachstelle dient. Aber dieser traditionelle Ansatz wird nun von einigen Sicherheitsforscher*innen als zu begrenzt beschrieben. Zum Beispiel hat Microsoft Principal Researcher Cormac Herley [1] darauf hingewiesen, dass Effektivität der meisten IT-Sicherheitsmaßnahmen nicht immer nachgewiesen werden kann. Der Mathematiker, Kryptograph und Wissenschaftshistoriker Andrew Odlyzko [2] hat darüber hinaus gefordert, dass auch die Wirtschaftlichkeit und Nachhaltigkeit von Sicherheitsmaßnahmen untersucht werden muss, bevor sie eingesetzt werden. Die US NSA rief bereits 2013 einen Wettbewerb für die besten Science of Cyber Security Papers aus, um die Entwicklung von nachweisbar effektiven und in der Realität einsetzbaren Sicherheitsmaßnahmen zu stimulieren.

Obwohl es seit 2002 den von Ross Anderson und Bruce Schneier gegründeten Workshop on Economics of Security (WEIS) gibt, ist die Erforschung der konkreten Auswirkungen von Sicherheitsmaßnahmen auf Verhalten und Produktivität in Organisationen auf eine Handvoll Fallstudien beschränkt, von denen die bekanntesten von einer Antragstellerin [3, 4, 5] sind. Aber diese dokumentieren die Auswirkungen von Sicherheitsmaßnahmen zu einem Zeitpunkt, und zeigen auf, wie die Maßnahmen verbessert werden könnten. Was z. Zt. völlig fehlt, sind Langzeitstudien, die die Umsetzung solcher Vorschläge mit Messungen zur Produktivität begleiten, und aufzeigen, wie Organisationen lernen, Sicherheitsmaßnahmen zu verbessern, ohne der Produktivität zu schaden. Die Forschung von Prof. Dr. Annette Kluge beschäftigt sich genau mit diesem Thema aus der Perspektive der Wirtschaftspsychologie [6, 7], und hat sowohl individuelles als auch organisatorisches Lernen und Entscheidungsfindung im Rahmen des sehr verwandten Themas Arbeitssicherheit untersucht. Bei Herley [1] und Odlyzko [2] klingt die Abwesenheit von klaren Anforderungen von Seiten der Unternehmen an. Die Sicherheitslieferanten sind teilweise für den jetzigen Zustand verantwortlich – würden die Unternehmen die konkreten Leistungsanforderungen an Sicherheitsmaßnahmen selbst erstellen, so führte dies zu besseren Produkten. Die Tandempartnerinnen würden also die Effektivität von Maßnahmen aus technischer Sicht und im Anwendungskontext in Unternehmen messen.

Referenzen

[1] Cormac Herley: The Unfalsifiability of Security Claims. Proceedings of the 3rd Asian Symposium on Programming Languages and Systems (APLAS ,05) | May 2016.
[2] Andrew Odlyzko: Cybersecurity is not very important. ACM Ubiquity, June 2019, pp. 1-23.
[3] Philip G Inglesant, Martina Angela Sasse: The true cost of unusable password policies. Proceedings of CHI 2010.
[4] Adam Beautement, Martina Angela Sasse, Mike Woham, The Compliance Budget. Proceedings of NSPW 2008.
[5] Iacovos Kirlappos, Simon Parkin, Martina Angela Sasse: Learning from Shadow Security - why understanding non-compliance provides the basis for effective security. ACM SIGCAS Computers and Society, 2015.
[6] Jan Schilling, Annette Kluge: Barriers to organizational learning: An integration of theory and research. International Journal of Management Reviews 2009.
[7] Daniel Putz, Jan Schilling, Annette Kluge: Measuring organizational learning from errors: Development and validation of an integrated model and questionnaire. Management Learning, 2012.

2. Forschungsthema

Kognitive Aspekte des Hackens

(Betreuerteam: Prof. Christof Paar, Prof. Nikol Rummel)

Abstract

Das Tandemprojekt behandelt einen wichtigen Baustein der modernen IT-Sicherheit, die Obfuskation. Dieses Thema wird bisher in der Literatur praktisch nur als technisches behandelt, obwohl der Mensch hier eine zentrale Rolle spielt. Ein Hauptziel besteht dabei darin, Hardware-Obfuskation zu entwickeln, die auf dem Verständnis der menschlichen Kognition basieren, zu entwickeln. In einem weiteren Schritt sollen Methoden entwickelt werden, um die kognitiven Vorgänge beim Software-Reversen zu analysieren, wobei u.a. Methoden des Eye-Trackings eingesetzt werden sollen.

In der Tandempromotion „Lernprozesse in der IT-Sicherheit“ in der ersten SecHuman-Förderphase wurde weltweit das erste Mal der Frage nachgegangen, welche kognitiven Prozesse bei Angriffen auf Computer-Hardware ablaufen. Ausgangspunkt hierbei ist es, das Verstehen einer Hardware-Schaltung in Form eines integrierten Schaltkreises (IC) als einen Problemlöseprozess zu betrachten. Hierbei ergibt sich die Möglichkeit, auf die Methoden der Problemlöseforschung zuzugreifen. In der ersten Tandempromotion konnten viele wichtige Ergebnisse erzielt werden, die in den Konferenzbeiträgen [3-6] beschrieben sind sowie in den eingeladenen Vorträgen auf der SHB (Security and Human Behavior Workshop) 2018 und 2019 an der CMU bzw. Harvard [1, 2].

Ein wesentlicher Teil der bisherigen Forschung bestand in der Konzeption und Durchführung komplexer Studien, mit denen sich die Lernvorgänge beim Hardware-Reverse-Engineering untersuchen lassen. Dies hat zu umfangreichen Datenmaterial geführt, aus dem aktuell eine psychologisch orientierte Einreichung in einer renommierten Kognitions-Fachzeitschriftin Arbeit ist und eine tech-

nisch-orientierte Publikation auf der SOUPS 2020 stattfand. Kennzeichnend für die bisherigen Forschungen war die extrem enge Zusammenarbeit der beiden Tandempromovierenden sowie der betreuenden Hochschullehrer*innen Paar und Rummel.

Ziel der Tandempromotion „Kognitive Aspekte des Hackens“ ist es, aufbauend auf den bisherigen Ergebnissen, die beiden folgenden wissenschaftlich vielversprechenden Forschungsfragestellungen zu untersuchen:
Fragestellung 1: Kognitive Obfuskation. In modernen ICs wird im großen Stil Hardware-Obfuskation eingesetzt. Dies dient sowohl dem Schutz geistigen Eigentums, aber auch dem Erschweren von Hardware-Manipulationen. Diesem weit verbreiteten Ansatz steht eine ausgesprochen beschränkte Behandlung in der wissenschaftlichen Literatur gegenüber. Die wenigsten existierenden Techniken beschäftigen sich zumeist mit sehr speziellen Methoden wie FSM-Obfuskation, die zudem zumeist gebrochen wurden. Ziel der Tandemforschung ist es, Obfuskationsmethoden zu entwickeln, die auf dem Verständnis der Kognitionsvorgänge beim Hardware-Reverse-Engineering basieren. Hierbei werden Effekte ausgenutzt, bei denen menschlichen Angreifer*innen die größten Schwierigkeiten bereiten. Ein wichtiger Teil der Forschung besteht darin, Hypothesen zu kognitiv-schweren Reverse-Engineering-Problemen zu erstellen und diese experimentell zu verifizieren. Ziel ist es, Richtlinien für starke Obfuskationsmethoden zu formulieren. Dies wäre ein wissenschaftliches Novum und hätte zudem eine starke transdisziplinäre Komponente, da es einen großen Bedarf in der Industrie für solche Lösungen gibt.

Fragestellung 2: Software-Obfuskation. Es wird erwartet, dass im Rahmen der Tandempromotion auch wissenschaftlich fundierte Grundlagen für starke Software-Obfuskation geschaffen werden können. Obwohl es hier mehr Literatur als im Hardware-Fall gibt, wird die gesamte Software-Obfuskationsforschung praktisch ohne Betrachtung des menschlichen Verhalten durchgeführt. Als zweite Fragestellung soll von dem Tandem eine Methodik entwickelt werden und erste Untersuchungen durchgeführt werden, mit denen das menschliche Vorgehen beim Software-Reversen analysiert werden kann. Hier sollen u.a. Methoden des Eye-Tracking genutzt werden.

Referenzen

[1] Christof Paar, Nikol Rummel: Cognitive Obfuscation: The Human Factor in Hacking (II). Workshop on Security and Human Behavior, 2019, Harvard Univ., June 5-6, 2019.
[2] Christof Paar, Nikol Rummel. Cognitive Obfuscation: The Human Factor in Hacking. Workshop on Security and Human Behavior, 2018, Carnegie Mellon Univ., June 24-25, 2018.
[3] Steffen Becker, Carina Wiesen, Nils Albartus, Christof Paar, Nikol Rummel: Promoting the Acquisition of Hardware Reverse Engineering Skills. 2019 IEEE Frontiers in Education Conference (FIE), Cincinnati, OH, USA. 2019.
[4] Steffen Becker, Carina Wiesen, Christof Paar, Nikol Rummel: Wie arbeiten Reverse Engineers? Datenschutz und Datensicherheit 43(11): 686-690 (2019)
[5] Carina Wiesen, Nils Albartus, Max Hoffmann, Steffen Becker, Sebastian Wallat, Marc Fyrbiak, Nikol Rummel, Christof Paar: Towards cognitive obfuscation: impeding hardware reverse engineering based on psychological insights. ASP-DAC 2019: 104-111

[6] Marc Fyrbiak, Sebastian Strauss, Christian Kison, Sebastian Wallat, Malte Elson, Nikol Rummel, Christof Paar: Hardware reverse engineering: Overview and open challenges. IVSW 2017: 88-94

3. Forschungsthema

Ethische Schwachstellenanalyse: Automatisierte Methoden und ethische Orientierung

(Betreuerteam: Prof. Thorsten Holz, Jun.-Prof. Sebastian Weydner-Volkmann)

Abstract

Im Rahmen des Tandems soll an Methoden geforscht werden, die es ermöglichen, ein gegebenes IT-System (klassischer Computer, IoT-Gerät) automatisiert auf Schwachstellen untersuchen zu können. Dazu werden sowohl statische als auch dynamische Analysetechniken entwickelt. Oft tritt hierbei bei Wissenschaftler*innen ein ethisches Dilemma auf, wenn Schwachstellen gefunden werden. Daher soll von der/dem Tandempartner*in im Sinne einer deskriptiven Ethik etablierte Wertkonzepte und bewährte Handlungsregeln herausgearbeitet und innerhalb allgemeinerer ethischer Prinzipien systematisiert werden. Trotz der unterschiedlichen Disziplinen ist ein enger Austausch innerhalb des Promotionstandems erforderlich, um möglichst praxistaugliche Hinweise für den Umgang mit Wertkonflikten zu entwickeln, und so der Schritt hin zu einer normativen Bereichsethik, einer Ethik für IT-Sicherheit, zu vollziehen.

In der IT-Sicherheitsforschung ist die systematische Suche nach Schwachstellen etwa in verbreiteten Computersystemen, kritischen Anwendungen und öffentlich angebotenen Web-Diensten ein zentrales Forschungsthema. Dabei wird das Ziel verfolgt, die Sicherheit und Verlässlichkeit in der Informationstechnik zu stärken und so auch die Angreifbarkeit digitalisierter Gesellschaften zu senken. In ihrer konkreten Arbeit stoßen Forschende jedoch immer wieder auf Situationen, in denen dieses Ziel mit anderen Wertvorstellungen kollidiert, etwa wenn Schwachstellen öffentlich bekannt werden, bevor diese in kritischen Systemen geschlossen worden sind. Entsprechend ist der Umgang mit Schwachstellen ein in der Praxis viel diskutiertes Problem, für das immer noch viele offene Fragen bestehen.

An das recht alte Konzept einer „Hackerethik“ anknüpfend wurden mit Blick auf solche Wertkonflikte u.a. durch Berufsverbände und die scientific community mittlerweile eine Reihe ethischer Verhaltenskodizes und best practices formuliert, die für einen verantwortungsbewussten Umgang mit Sicherheitslücken Orientierung bieten sollen (z.B. zu responsible disclosure oder die gängige Frist von 90 Tagen, bis Informationen zu einer Schwachstelle publik gemacht werden). Gleichzeitig werden diese Kodizes durch neue gesellschaftliche Entwicklungen, aber auch durch neue Methoden zur Entdeckung von Schwachstellen, immer wieder herausgefordert, wie aktuell etwa durch die Entdeckung von Fehlern in Computerprozessoren oder mobilen Kommunikationssystemen wie LTE. Im Rahmen dieses Tandems soll aus der IT-Sicherheits-Perspektive an Methoden geforscht werden, um automatisiert ein gegebenes System effizient und effektiv auf mögliche Schwachstellen hin untersuchen zu können. Dazu werden sowohl statische als auch dynamische Analysetechniken entwickelt; ein Fokus liegt auf automatisierten Methoden aus dem Bereich des Fuzz Testings: Bei

dieser Technik wird das zu untersuchende System mit speziell gewählten Zufallsdaten ausgeführt, um mögliche Programmierfehler zu entdecken. Hier lässt sich nochmals eine ganz eigene Dynamik ethischer Konflikte erwarten, insbesondere aufgrund der zu erwartenden hohen Anzahl an entdeckten Schwachstellen.

In der moralphilosophischen Forschung hat das Feld der IT-Sicherheit bis auf wenige Ausnahmen bislang allerdings kaum eigens Aufmerksamkeit gefunden. Im Sinne einer deskriptiven Ethik soll im Rahmen des Tandems der Versuch unternommen werden, etablierte Wertkonzepte und bewährte Handlungsregeln herauszuarbeiten und innerhalb allgemeinerer ethischer Prinzipien zu systematisieren. Gerade mit Blick auf neuere Entwicklungen sollen über einen engen Austausch innerhalb des Promotionstandems möglichst auch praxistaugliche Hinweise für den Umgang mit aktuell entstehenden Wertkonflikten innerhalb der IT-Sicherheitsforschung erarbeitet und so der Schritt hin zu einer normativen Bereichsethik vollzogen werden. Eine solche angewandte Ethik für IT-Sicherheit zielt letztlich darauf, eine umfassendere ethische Orientierung für wirtschaftliche Akteure und die scientific community zu bieten – auch mit Blick auf die Entdeckung neuartiger sicherheitsrelevanter Schwachstellen.

4. Forschungsthema

Schutz vor Re-Identifizierung bei der Verlinkung von Daten

(Betreuerteam: Prof. Maike Buchin, Prof. Estrid Sørensen)

Abstract

De-Anonymisierung ist durch den „Erfolg“ von Social Networks zusammen mit der Fülle anderer digitaler Daten zu einem großen Problem geworden. In diesem Projekt sollen Methoden zur De- und Re-Identifizierung von Datensätzen untersucht sowie das Risiko durch die dabei entstandene Verlinkung bewertet werden. In diesem Tandem sollen Verfahren zur De- und Re-Identifizierung sowohl technisch als auch sozialanthropologisch betrachtet und beurteilt werden. Dazu soll das Risiko der Re-Identifizierung mit statistischen Methoden ermittelt werden.

Heutzutage werden immer größere Mengen von digitalen Daten erzeugt. Um aktuelle Datenschutzstandards – z. B. die DSGVO – einzuhalten, wird ein Teil dieser Daten typischerweise de-identifiziert zugänglich gemacht. Die wachsenden Datenmengen, die umfassenderen Dateninhalte sowie die verbesserten Methoden zur Verlinkung von Daten haben auch zu einem erhöhten Wert von Daten (Big Data) geführt. Zunehmend werden durch die Verlinkung verschiedener Datensätze neue, noch aussagekräftigere Datensätze erstellt. Häufig handelt es sich dabei dann um vertrauliche Daten. Zum Beispiel steigt bei der Verlinkung von Umfragedaten und geographisch detaillierten Raumdaten die Wahrscheinlichkeit, dass einzelne Befragte re-identifiziert werden können, auch wenn in beiden Ursprungs-Datensätzen bereits Anonymisierungsmaßnahmen zur Anwendung kamen. Bei der Verlinkung von Datensätzen entsteht somit ein Risiko der Re-Identifizierung von Individuen. In einem Paper, das für viel Debatte gesorgt hat, haben Luc Rocher und Kolleg*innen (2019) kürzlich postuliert, dass durch die Verwendung von 15 demographische Merkmalen 99.98% der US-Amerikaner in jeg-

lichem Datensatz korrekt re-identifiziert werden können. Diese Zahl von Merkmalen entsteht leicht bei der Verlinkung von Datensätzen. Rocher et al. schlussfolgern, dass es dabei unwahrscheinlich ist, dass Datensätze die Forderungen der DSGVO und ähnlicher Datenschutz-Standards nach Anonymisierung erfüllen. Dies erweist sowohl eine rechtliche wie auch eine technische Herausforderung existierender De-Identifikations und release-and-forget-Modelle.

Das Re-Identifizierungsrisiko gilt insbesondere bei der Verlinkung von Daten, das heißt, wenn verschiedene Datensätze, wie Raumdaten, Zeitreihen sowie Internetdaten (z.B. Social Media) miteinander verknüpft werden. In diesem Projekt sollen die Methoden zur De- und Re-Identifizierung von Datensätzen untersucht werden sowie auch das Risiko der De- und Re-Identifizierung von Datensätzen bewertet werden, die aus einer Verlinkung entstanden sind. Dabei soll einerseits das Risiko der Re-Identifizierung mit statistischen Methoden ermittelt werden und andererseits sollen verschiedene Methoden und Verfahren zur De- und Re-Identifizierung sowohl mathematisch-informatisch wie auch sozialanthropologisch betrachtet und bewertet werden. Das Secure Data Center des GESIS Leibniz-Instituts für Sozialwissenschaften fungiert in diesem Tandem als Praxispartner, vor allem soll mit dem Team „Data Linking & Data Security“ zusammengearbeitet werden. Das Secure Data Center bietet Zugang zu Forschungsdaten, die aus Datenschutzgründen besonderen Zugangsbeschränkungen unterliegen, und berät zur Forschung mit sensitiven Forschungsdaten. Die Verlinkung von Forschungsdaten besonders mit Social-Media-Daten ist ein Schwerpunkt der Arbeiten.

Im mathematischen-informatischen Teil des Tandems soll untersucht werden, wie sich die Verknüpfung von Daten auf die Re- und De-identifikation auswirkt, und welche Methoden hier zum Einsatz kommen können. Ebenfalls soll untersucht werden, wie die Ergebnisse der Verknüpfung Nutzenden zur Verfügung gestellt werden können. Insbesondere interessieren wir uns für die Verknüpfung von Daten, durch die eine Bewegung einer Person in Raum und Zeit herleitbar und damit ihre Identifikation möglich ist. Wir betrachten Möglichkeiten, um für solche Daten bestehende Methoden der De-Identifikation zu erweitern und damit sicherer zu machen. Dazu betrachten wir insbesondere, welche räumlichen sowie attributbezogenen Vergröberungen und Manipulationen zu verknüpfender bzw. verknüpfter Daten datenschutzrechtlich unbedenklich sind. Ebenfalls betrachten wir Konzepte zur Bereitstellung aggregierter oder synthetischer Daten. Das Ziel der De-Identifikation ist die Bereitstellung der Daten und wir untersuchen daher verschiedene Möglichkeiten und deren Wert für Nutzende der Daten.

Im sozialanthropologischen Teil des Projekts werden einerseits die Bedeutung von wissenschaftlichen Datenpraktiken sowie der Konfiguration der soziomateriellen Infrastruktur am Daten-Zugriffspunkt für den Schutz vor Re-Identifizierung untersucht. Andererseits soll beobachtet und analysiert werden, wie sozialwissenschaftliche Datenpraktiken und dadurch auch epistemologische Praktiken der Sozialwissenschaft sich transformieren durch die Risikoeinschätzungen der Re-Identifizierung sowie ihren Schutz. Durch teilnehmende Beobachtungen, Interviews und Dokumentenanalyse soll die Bedeutung von Datenpraktiken für den Schutz vor Re- Identifizierung untersucht werden. Dabei stehen die Begrenzung der Datennutzung auf kontrollierte wissenschaftliche Kreise sogenannte „trusted communities“ im Fokus der Untersuchung, wie auch die formellen und informellen Normen,

Belohnungs- und Kontrollsysteme der Wissenschaft. Darüber hinaus soll erforscht werden, wie epistemische Praktiken durch veränderte Datenpraktiken beeinflusst werden, z.B. ob Änderungen in Fragestellungen und Themen sowie auch in der Kategorisierung und Theoretisierung von untersuchten Phänomenen beobachtet werden können.

Eine enge Zusammenarbeit der zwei Tandemprojekte wird erwartet, indem beide an Datensätzen, Verfahren zum Schutz vor Re-Identifizierung sowie Datenpraktiken des GESIS Secure Data Center forschen werden. Einerseits werden Verfahren zum Schutz vor Re-Identifizierung, die durch das mathematisch-informatische Projekt entwickelt werden in Datensätzen implementiert, an denen Sozialwissenschaftler*innen arbeiten und ihren Datenpraktiken anpassen. Diese sind Gegenstand der sozialanthropologischen Forschung. Andererseits weisen die Erkenntnisse des sozialanthropologischen Projekts über wissenschaftliche Datenpraktiken auf die Arten von Verlinkungen hin, die besonders schutzbedürftig sind. Die Zusammenarbeit der zwei Tandemprojekte erhöht sowohl die Breite wie auch die Relevanz der Ergebnisse zum Schutz vor Re-Identifizierung bei der Verlinkung von Datensätzen.

5. Forschungsthema

Privatautonomes Entscheiden als Sicherheitsrisiko

(Betreuerteam: Prof. Markus Dürmuth, Prof. Tobias Gostomzyk)

Abstract

Datenschutzrechtliche Einwilligungen setzen auf einen freiwilligen, selbstbestimmten Zustimmungsakt von Privatpersonen („informed consent“ bzw. „notice and consent“). Tatsächlich weisen aber viele Studien darauf hin, dass viele Einwilligungen – etwa beim Akzeptieren von Cookies – uninformiert getroffen werden. Ziel der Tandemforschung ist es, zum einen die Prüfung von Autonomie (u.a. Informiertheit) einer sozialwissenschaftlichen und rechtlichen Bestandsaufnahme zu unterziehen, und zum anderen dabei alternative Mechanismen zu „notice and consent“ zu entwickeln, die für menschliche Nutzer*innen zu bevorzugen sind.

Moderne IT-Systeme, das Internet und insbesondere Big Data führen zu einer Erosion datenschutzrechtlicher Grundprinzipien, die sich größtenteils in den 1980er Jahren herausgebildet haben. Die Einwilligung ist ein Beispiel hierfür: Personenbezogene Daten dürfen nur erfasst werden, wenn eine wirksame Einwilligung vorliegt oder ein gesetzlicher Erlaubnistatbestand dies gestattet, beispielsweise die Wahrnehmung berechtigter Interessen. Voraussetzungen für eine wirksame Einwilligung sind aber unter anderem, dass sie in informierter Weise bzw. in Kenntnis der Sachlage erfolgt. Hierzu muss die oder der Einwilligende zumindest wissen, wer welche personenbezogenen Daten für welche Zwecke verwenden können soll und an wen er diese überdies weitergeben darf (dazu etwa Bitter/Buchmüller/Uecker, Datenschutzrecht, in: Th. Hoeren (Hrsg.), Big Data und Recht, 2014, S. 72 ff.). An diese Einwilligung ist die spätere Datennutzung gebunden.

Die Einwilligung - genuiner Ausdruck informationeller Selbstbestimmung - wird unter den Bedingungen moderner IT und Big Data einer Belastungsprobe ausgesetzt. Sie betrifft insbesondere den

Zustimmungsakt und die Zweckbindung der Datennutzung: Datenschutzrechtliche Einwilligungen setzen auf einen freiwilligen, selbstbestimmten Zustimmungsakt von Privatpersonen („informed consent“ bzw. „notice and consent“). Tatsächlich weisen aber viele Studien darauf hin, dass viele Einwilligungen, etwa beim Akzeptieren von Cookies, uninformiert getroffen werden (etwa J.A. Obar/A. Oeldorf-Hirsch, The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services, Information, Communication & Society, pp. 1-20, 2018; I. Pollach, What's wrong with online privacy policies?, Communications of the ACM, vol. 50, no. 9, pp. 103-108, 2007; D. Susser, Notice After Notice-and-Consent: Why Privacy disclosures are valuable even if consent frameworks aren't, Journal of Information Policy 9: 37–62, 2019). Weder werden Datenschutzerklärungen vor einer Einwilligung gelesen („clicking-without-reading“) oder – sofern ausnahmsweise doch – meist nicht oder nur teilweise zutreffend verstanden. Auch die Zweckbindung der Einwilligung wird durch technische Entwicklungen in Frage gestellt. Das betrifft gerade den Bereich big data/machine learning, wo von Nutzer*innen erhobene Daten z.B. in das Training und die Selektion von Algorithmen und Modellen einfließen. Hier ist der Zweck der Daten oft nicht mehr klar erkennbar und durch den Einsatz von trainierten Modellen in immer neuen Einsatzzwecken extrem wandelbar. Das widerspricht einer Informiertheit ex ante, von der das Prinzip der datenschutzrechtlichen Einwilligung ausgeht.

Es stellt sich mithin die Frage, welche Alternativen es zu einer Einwilligung im herkömmlichen Sinne gibt: Diese werden in der juristischen Literatur bislang nur vereinzelt und eher apodiktisch diskutiert (etwa L. Specht/L. Bienemann, Zur Zukunft der datenschutzrechtlichen Einwilligung, in: Beilage 1 zu K&R 9/2018, S. 22-23, S. Augsberg/U. Ulmenstein, Modifizierte Einwilligungserfordernisse - Kann das Datenschutzrecht vom Gesundheitsrecht lernen? | GesR 2018, 341-346). Eine umfassende Aufarbeitung hierzu fehlt, soweit ersichtlich. Denkbar wären z.B. ein stärkeres Datenschutzbewusstsein bei Einwilligung durch eine wirtschaftliche Betrachtung von Daten oder die Limitierung der Einwilligung der Reichweite von Einwilligungen durch den Gesetzgeber und eine intensivere Aufsichtspraxis („rote Linien“); das gilt gerade für den Bereich asymmetrischer Machtverhältnisse (dazu etwa M. Kamp/M. Rost, Kritik an der Einwilligung, DuD 2013, 80-84).

Produktiv für eine Weiterentwicklung erscheinen aber vor allem die Verknüpfung einer rechtlichen und einer technischen Perspektive („privacy by design“): Ein möglicher Ansatz wäre eine standardisierte Schnittstelle für Privatsphäre-Einstellungen, ein Ansatz der in der Vergangenheit aber wiederholt gescheitert ist, vgl. den Do-Not-Track Standard (J. Mayer and J. Mitchell. Third-Party Web Tracking: Policy and Technology. IEEE SP 2012.). Erfolgsversprechender erscheinen daher Ansätze, die ohne weitere Mitwirkung der Seitenbetreiber auskommen, z.B. erscheint es möglich, „Cookie-Banner“ automatisiert auszufüllen, basierend auf vom Benutzer einmal vorgenommenen Privatsphäre-Einstellungen. Auch wenn heute nur eine Minderheit der eingesetzten Cookie-Banner den rechtlichen Anforderungen genügen, könnte es unter Umständen möglich sein, eine Struktur der Cookie-Banner einzufordern, die ein automatisiertes Ausfüllen erleichtert.

Ein anderer Ansatz könnte bei der Einwilligung selbst auf technische Assistenten setzen, welche die eigene Einwilligungspraxis dokumentieren (Wo welche Einwilligungen zu welchem Zeitpunkt un-

ter Verwendung welcher Datenschutzerklärungen?) und Risikobewertungen vornehmen. Denkbar wäre aber auch ein fortlaufendes Monitoring von erteilten Einwilligungen bzw. dem Gebrauch von personenbezogenen Daten, die – gerade wenn sie in einem neuen, datenschutzrechtlich relevanten Kontext verwendet werden sollen – einen Hinweis sendet, ob erneut eingewilligt bzw. widersprochen werden soll.

Letztlich also soll die Prüfung von Autonomie (u.a. Informiertheit) als Grundannahme einer Einwilligung einer sozialwissenschaftlichen und rechtlichen Bestandsaufnahme unterzogen werden. Sich hieraus ergebende Herausforderungen sollen insbesondere im Verbund mit technischen Lösungen weiterentwickelt werden.

Mögliche übergreifende Forschungsfragen:

- Welche Chancen und Risiken bringt der „notice and consent"-Ansatz aus rechtlicher und sozialwissenschaftlicher Perspektive?
- Wie lassen sich der bisherige „notice and consent"-Ansatz im Verbund einer rechtlichen und technischen Perspektive modifizieren bzw. funktionale Äquivalente zum bisherigen Ansatz finden?
- Welche alternativen Mechanismen zu „notice and consent" sind aus Nutzersicht zu bevorzugen, weil sie die rechtliche Zielvorstellung „informationelle Selbstbestimmung" im Zeitpunkt der Erteilung einer datenschutzrechtlichen Einwilligung und hierüber hinaus unterstützen?

der Beantwortung dieser Fragen käme ein Mix aus Methoden zum Einsatz. Zentral wären auf der einen Seite ein Mix aus Nutzerbefragungen und Experimenten (sowohl in-lab also auch online), um Intentionen und Handlungen in Bezug auf Einwilligungen zu untersuchen. Sie wären eng mit rechtlichen Fragestellungen zu verknüpfen, um auch hier Aussagekraft zu erhalten.
Wir haben somit einen Themenkomplex identifiziert, der in sehr enger Abstimmung zwischen den Tandempartnern bearbeitet werden kann. Verbunden werden hier eine rechtliche, technische und sozialwissenschaftliche Perspektive, die auch die PI's fachlich abdecken.

6. Forschungsthema

Autorprofilerstellung der Verfasser*innen von Hate Speech

(Betreuerteam: Prof. Dorothea Kolossa, Prof. Karin Pittner, Dr. Kerstin Kucharczik-Kohrt)

Abstract

In diesem Promotionstandem sollen linguistische Analyseverfahren mit maschinellen Lernverfahren kombiniert werden, um Hassmails automatisiert identifizieren und klassifizieren zu können. Aus der linguistischen Perspektive lautet dabei die Fragestellung, ob anhand von sprachlichen Merkmalen verschiedene Typen von Hassmails zu erkennen sind, die zu einer Profilierung der Verfasser*innen führen können. Eine anspruchsvolle technische Fragestellung ist, inwieweit maschinell erkennbare spezifische sprachliche Merkmale oder Merkmalskombinationen Aufschluss über Einstellungen wie Extremismus, Gewaltbereitschaft etc. und die Zugehörigkeit der Verfasser*innen zu bestimmten Gruppierungen geben können.

Hassmails sind nicht nur eine Bedrohung für den Einzelnen, sondern zunehmend auch für das politische System. Zudem stellen sie aufgrund ihrer Masse und der häufigen Verschleierung der Identität der Verfasser*innen ein wachsendes Problem für Ermittler dar.

In dem Projekt sollen linguistische Analyseverfahren mit maschinellen Lernverfahren kombiniert werden, um Hassmails automatisiert identifizieren und klassifizieren zu können.

Aus der linguistischen Perspektive lautet die Fragestellung, ob anhand von sprachlichen Merkmalen verschiedene Typen von Hassmails zu erkennen sind, die zu einer Profilierung der Verfasser/innen (z.B. hinsichtlich ihrer Muttersprache, regionalen Einordnung, Alter und Bildungsgrad, eventuell auch Geschlecht, vorgetäuschte Identitäten) führen können [Ehr2018]. Eine weitergehende Fragestellung ist, inwieweit spezifische sprachliche Merkmale oder Merkmalskombinationen Aufschluss über Einstellungen wie Extremismus, Gewaltbereitschaft etc. und die Zugehörigkeit der Verfasser/innen zu bestimmten Gruppierungen geben können. Dabei sind Fragen der singulären respektive multiplen Autorschaft von Hassmails ebenso relevant wie Aspekte der individuellen respektive gruppenbezogenen Adressierung (insbesondere hinsichtlich extremistischer Inhalte).

Zu diesem Zweck sollen mit Hilfe maschinellen Lernens Analyseverfahren entwickelt werden, die geeignet sind, entsprechende Merkmale zu identifizieren und die so eine automatisierte Klassifikation von Hassmails ermöglichen. Der erste Schwerpunkt liegt hierbei auf der Interpretierbarkeit der Analyseergebnisse, die durch eine enge Verzahnung linguistisch relevanter Sprachmerkmale mit maschinell gelernten Entscheidungsfunktionen ermöglicht wird. Dies kann beispielsweise durch die Verwendung von Attention-Modellen, wie dem Transformer-Modell [Dev2019] geschehen, die wie in [Boe2019-2] die Grundlage ihrer Entscheidungsfunktion ausgeben und der Analyse zugänglich machen können. Der zweite Schwerpunkt liegt in der Entwicklung von Konfidenzmaßen für die einzelnen Entscheidungen. Hierfür hat es sich in dem vorausgehenden Projekt bewährt, mit Hilfe von siamesischen Netzen auch die Entscheidungsmetrik selbst zu optimieren [Boe2019-1], um so hochzuverlässige Entscheidungen klar identifizieren zu können, und um Problemfälle automatisch auszusortieren, beispielsweise für eine weitergehende Analyse und/oder zum Zweck einer weiteren Verbesserung des Klassifikationsmodells.

Die Ergebnisse des Projekts können neben einer Anwendung für die Verbesserung der automatischen Identifikation und Klassifikation von Hassmails für ein Beratungsportal nutzbar gemacht werden, in dem Betroffenen Ratschläge für eine passende Reaktion bzw. rechtliche Schritte gegeben werden.

Referenzen

[Ehr2018] S. Ehrhardt (2018): Authorship attribution analysis. In: Visconti, Jacqueline (ed.): Handbook of Communication in the Legal Sphere. Berlin/Boston: de Gruyter, 169-200.
[Dev2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT.
[Boe2019-1] B. Boenninghoff, R. M. Nickel, S. Zeiler, D. Kolossa: „ Similarity Learning for Authorship Verification in Social Media," Proc. ICASSP, Brighton, UK, May 2019.

[Boe2019-2] B. Boenninghoff, S. Hessler, D. Kolossa, R. Nickel: „Explainable Authorship Verification in Social Media via Attention-based Similarity Learning," accepted for publication at IEEE Int. Conference on Big Data, Los Angeles, CA, USA, December 9-12, 2019.