Martin Steinebach, Erik Krempel, Christian Jung, Mario Hoffmann

Datenschutz und Datenanalyse

Herausforderungen und Lösungsansätze

Eine der großen Chancen von Big Data sind neue Erkenntnisse, die sich durch die schnelle und flexible Analyse großer Datenmengen ergeben. Um diese Analyse aber rechtskonform durchzuführen, bedarf es der Beachtung des Datenschutzes, sobald auf personenbezogene Daten zugegriffen wird. Mittlerweile existieren verschiedene Ansätze und Konzepte, wie Datenschutz und Datenanalyse vereint werden können.

Dr-Ing. Martin Steinebach

leitet am Fraunhofer SIT die Abteilung Multimedia Sicherheit und IT Forensik. Er vertritt das SIT in der Big Data Allianz der Fraunhofer Gesellschaft.

E-Mail: martin.steinebach@sit.fraunhofer.de

Christian Jung

leitet die Abteilung »Security Engineering« am Fraunhofer-Institut für Experimentelles Software Engineering (IESE) in Kaiserslautern und verantwortet dort das Forschungsfeld Datennutzungskontrolle.
E-Mail: christian.jung@iese.fraunhofer.de

Dipl.-Inform. Erik Krempel

untersucht am Fraunhofer IOSB gemeinsam mit seiner Forschungsgruppe, wie rechtliche und soziale Anforderungen an IKT-Systeme, durch technische Maßnahmen erzwungen werden können.
E-Mail: erik.krempel@iosb.fraunhofer.de

Mario Hoffmann, Dipl.-Inform.

arbeitet seit über 15 Jahren in der IT-Sicherheit auf den Gebieten Identitätsmanagement, mobile Sicherheit, Datenschutz und Cloud Computing.

E-Mail: mario.hoffmann@mykolab.com

1 Einleitung

„Privacy by Design“ gilt seit 2009 als Gestaltungsprinzip für die datenschutzfreundliche Realisierung digitaler Systeme [1]. Vorderstes Ziel ist dabei die Einbettung datenschutzfördernder Prozesse und Technologien bereits während Design und Konzipierung dieser Systeme. Cavoukian stellt sieben Prinzipien vor, darunter Transparenz, Ende-zu-Ende-Sicherheit und die nutzerzentrierte Gestaltung von Lösungen. Sie bilden Richtlinien in der Umsetzung. Doch schon eine Interpretation von Art und Umfang der „eigenen Daten“ führt je nach Perspektive und Anwendungsfeld zu heftigen Kontroversen. Ein besondere Herausforderung bleibt dabei auch Cavoukians Forderung nach voller Funktionalität bei gleichzeitig vollständigem Datenschutz; was in der Praxis in komplexen Systemen bisher als nicht realisierbar angesehen wird.

Vor Kurzem trat nun -mit einer zweijährigen Übergangsfrist- die europäische Datenschutzgrundverordnung (DSGVO) in Kraft, die obiges Gestaltungsprinzip in einen für alle Mitgliedsstaaten einheitlichen europäischen Rahmen gießt und auch für Unternehmen mit Sitz außerhalb Europas bindend ist. Viele Aspekte der DSGVO beinhalteten bereits die bislang gültige Datenschutz-Direktive 95/46/EC aus dem Jahr 1995. Eine im April 2016 veröffentlichte Studie zeigte jedoch, dass über 50% der verbreitetsten Online-Dienste die existierende Gesetzesgrundlage nicht ausreichend umsetzen [2].

Bürgerinnen und Bürger erhalten mit der DSGVO nun neben einem umfassenden Auskunftsrecht über vorgehaltene persönliche Daten, ein Recht auf Korrektur und Vergessen sowie die Möglichkeit der Portabilität eben dieser Daten. Zudem gilt eine umgehende Informationspflicht, sollten personenbezogene Daten unautorisiert in Umlauf gelangen. Auch die Zweckbindung gehört nach wie vor zu den Prinzipien des Datenschutzes.
Gleichzeitig ist unter dem Stichwort „Big Data“ in den vergangenen zwölf Monaten ein wahrer Goldrausch auf die stetig wachsenden personenbeziehbaren Datenmengen aus der zunehmenden Vernetzung von Alltagskontexten, wie Online-Diensten, Smart Home, eHealth und Automotive, entbrannt. Personalisierte Dienstleistungen und Produkte, gezielte Marketingaktionen und Verhaltensprognosen sind dabei nur die Spitze des Eisbergs einer vollständigen Profilbildung einer breiten Bevölkerung.

Es zeigt sich, dass hier eine gesellschaftlicher Diskussionsbedarf besteht, bei dem wirtschaftliche Interessen, die Daten als den Rohstoff der Zukunft sehen, welchen es zu gewinnen und zu nutzen gilt, dem Bedürfnis des Bürgers nach Datenschutz und informationeller Selbstbestimmung gegenüberstehen. Eine Abwägung mit dem gesellschaftspolitisch getriebenen Bedürfnis nach innerer Sicherheit sowie dem privatwirtschaftlich betriebenen Durchleuchten des gemeinen Konsumenten kann stets nur in einem engen Rahmen und zweckgebunden erfolgen, so wie es die DSGVO nun detailliert vorschreibt.

Das vorliegende Papier erörtert, wie in Zukunft gesellschaftspolitische und privatwirtschaftliche Interessen mit der Datenschutzgrundverordnung vereinbart werden können. Das Ziel dabei ist es, den Schutz der Privatsphäre technisch zu interpretieren und die Umsetzung der DSGVO in Produkten und Dienstleistungen als europäisches Alleinstellungsmerkmal und nicht als Hemmnis zu verstehen. Die vorgestellten Lösungsansätze sind allesamt solche, die in den Arbeiten der Sicherheitsgruppe der Fraunhofer Big Data Allianz eine Rolle spielen, ein vollständiger Überblick über alle bekannten Methoden wird hier nicht versucht.

2 Herausforderungen der Datenanalyse

Big Data führt zu neuen Herausforderungen im Umgang mit Daten. Große Mengen unterschiedlicher Daten werden zusammengefügt, um neue Methoden der Wertschöpfung zu realisieren, die auf einem vertieften Verständnis des Verhaltens der Personen gründen, über die die Daten gesammelt wurden. Konzepte, die ursprünglich als ausreichend zum Schutz der Privatsphäre betrachtet wurden, reichen nicht mehr aus, wenn immer mehr Daten miteinander verknüpft werden.

Hinsichtlich des Datenschutzes ist zu unterscheiden, ob diese Daten personenbezogen sind oder nicht. Personenbezogene Daten sind alle Daten, die auf eine bestimmbare Person hinweisen oder ihr zugeordnet sind. Einfache Beispiele sind körperliche Merkmale der Person, aber auch ihre Telefonnummer oder ihr Wohnort.

Nicht personenbezogene Daten sind Daten, für die (auch in Zukunft) keine Zuordnung zu handelnden oder betroffenen Personen möglich ist. Das gilt unter anderem für Daten, die sich ausschließlich auf Geräte und Produkte, nicht aber auf ihre Nutzer beziehen, beispielsweise Sensordaten zur Ortung von Transportgütern in der automatisierten Logistik. Wobei zu beachten ist, dass sich zu der Person, die Geräte beaufsichtigt, nutzt oder auch nur mit ihnen interagiert, potentiell auch hier ein Personenbezug herstellen lässt.

In der Praxis stellt diese Trennung immer wieder einer Herausforderung dar. Durch weitere Datenquellen und ein geschicktes Zusammenführen besteht das Risiko, dass Daten personenbezogen werden. So kann eine Unterscheidung immer nur nach jeweils aktuellem Stand des Wissens erfolgen.

Gerade für Big Data ist es eine der anstehenden Herausforderungen den Personenbezug von Daten zu quantifizieren. Bereits seit einigen Jahren versucht man in der Informatik Maße dafür zu finden, wie stark ein bestimmter Datensatz eine betroffene Person identifiziert. Für strukturierte Daten, beispielsweise in Datenbanken, ist dies bereits teilweise gelungen. Hier lassen bestehende Metriken beispielsweise die Aussage zu, dass Rückschlüsse niemals für einzelne Personen möglich sind, sondern nur für Gruppen mit einer Mindestanzahl an Personen. Bessere und aufwendigere Verfahren dieser Art werden in Kapitel 4.1 genauer beschrieben.

Während die Quantifizierung des Personenbezugs bei strukturierten Daten bereits gelingt, steht man bei Multimedia-Daten noch am Anfang. Big Data Anwendungen werden zukünftig neue Rückschlüsse für archivierte Daten erlauben, die heute noch nicht einmal absehbar sind. Im Februar 2016 war es dem neuronalen Netz PlaNet[1] von Google gelungen, die Geolokalisation von beliebigen Bildern deutlich präziser durchzuführen, als dies einem Menschen möglich ist. Wird eine solche Lokalisation mit einer Identifikation erkannter Personen und dem Zeitstempel der Fotos kombiniert, entstehen detaillierte Bewegungsprofile der Betroffenen aus ehemals unkritischen Daten. Hier werden neue Verfahren benötigt, um vor der Verarbeitung von Daten mögliche Datenschutzprobleme zu erkennen und zu vermeiden.

Die datenschutzrechtliche Bewertung von Daten stellt jedoch keineswegs nur eine technische Herausforderung dar. Bisher unterscheiden das Recht lediglich zwischen personenbezogenen und nicht personenbezogenen Daten. Diese absolute Unterscheidung wird zukünftig nicht immer erreicht werden können. Wenn für eine Datenbank lediglich eine quantifizierte Aussage, beispielsweise dass für weniger als 5% der Betroffenen eine Offenlegung ihrer Namen möglich ist, getroffen werden kann, ist es bisher völlig unklar, was dies für die Rechtssprechung bedeutet.

3 Lösungsansatz Privacy by Design

Um in Zukunft Systeme so zu gestalten, dass sie dem Datenschutz einfacher genügen können, gibt es die Prinzipien und Konzepte des „Privacy by Design" (PbD). Hier sind Herangehensweisen zusammengefasst, die, wenn sie beachtet werden, zu Systemen führen, die Nutzbarkeit und Datenschutz vereinen. Einen Einstieg in die Denkweise bieten beispielsweise die sieben Prinzipien von Ann Cavoukian [1]. Hier wird unter anderem gefordert, dass datenschutzfreundliche Voreinstellungen bei Wahlmöglichkeiten gegeben sind, ebenso aber auch, dass Datenschutz nicht zu Systemen führen darf, die für ihren eigentlichen Zweck unbrauchbar werden.

Erforderlich für Privacy by Design ist, wie der Name schon sagt, dass Privacy -also das Thema Datenschutz- bereits in der Planungsphase eines Systems berücksichtigt wird. Dies steht im deutlichen Unterschied zu den vorherrschenden Ansätzen, die Privacy, ähnlich wie Sicherheit allgemein, im Nachhinein über bestehende Systeme legen. Die Folge hiervon sind mangelhaft integrierte Lösungen, die Lücken aufweisen, welche regelmäßig von Angreifern ausgenutzt werden. Weiterhin sind die Kosten hinsichtlich von Aufwand und Performanzverlust deutlich höher.

Kritisiert wird an Privacy by Design, dass derzeit in erster Linie Anforderungen und Prinzipien diskutiert werden, ohne konkrete Handlungsempfehlungen oder beispielhafte Implementierungen in komplexen Systemen zu liefern. Dementsprechend sieht sich jeder, der sich mit der Umsetzung beschäftigen möchte, vor

1 https://www.technologyreview.com/s/600889/google-unveils-neural-network-with-superhuman-ability-to-determine-the-location-of-almost/

der Herausforderung, eine konkrete Lösung umzusetzen und dabei möglichst viele Aspekte zu berücksichtigen.

Die European Union Agency for Network and Information Security (ENISA) hat ein Dokument veröffentlich [3], welches sich dem Thema PbD aus juristischer Sicht annähert. Hier werden acht Prinzipien aufgestellt, denen teilweise auch Mechanismen zur Umsetzung zugeordnet sind. Diese werden in vier datenbezogene und vier organisatorische Prinzipen aufgeteilt.

- **Minimize** ist analog zur Datensparsamkeit definiert. Dahinter liegt die Forderung, immer nur so viele Daten zu erfassen, wie für die Lösung der vorliegenden Aufgabe notwendig sind.
- **Hide** adressiert die Verschleierung von Daten auf eine Weise, die es Menschen unmöglich macht, diese zu interpretieren, wenn sie in ihren Besitz gelangen sollten. Konkret bedeutet dies, dass Daten zum einen verschlüsselt transportiert und aufbewahrt werden sollen, zum anderen sollen Anonymisierungsmechanismen verwendet werden.
- **Separate** fordert eine verteilte Datenhaltung und –analyse. Es sollen nicht alle Informationen über eine Person an einem Ort gespeichert und verarbeitet werden, so dass sich bei einem Datenverlust der Schaden für diese in Grenzen hält und keine umfassende Profile über sie erstellt werden können. Dies verbietet insbesondere auch das Zusammenführen von unterschiedlichen Datenbanken mit Personenbezug.
- **Aggregate** führt personenbezogene Daten von Individuen möglichst frühzeitig zu Gruppen zusammen. So können im Sinne der Anonymisierung keine Rückschlüsse mehr auf individuelle Personen getroffen werden. Technisch kann dies mit k-anonymity und verwendeten Konzepten umgesetzt werden.
- **Inform** teilt Personen unmittelbar mit, welche Daten über sie gesammelt wurden, warum das geschah und wie die Daten geschützt werden. Weiterhin wird über eine Weitergabe der Daten sowie gegebenenfalls über einen Datenverlust durch Angriffe informiert.
- **Control** bezeichnet den Umstand, dass Personen die Kontrolle über Daten behalten sollen, die über sie gesammelt wurden.
- **Enforce** erfordert die Umsetzung der für den gegebenen Fall gültigen Datenschutzgesetze. Dies beinhaltet technische und organisatorische Maßnahmen.
- **Demonstrate** weist nach, dass Datenschutzmaßnahmen tatsächlich umgesetzt wurden. Hierzu sind Protokolldateien und Auditmöglichkeiten erforderlich.

Betrachtet man diese Punkte, wird offensichtlich, welche Herausforderung eine umfassende Implementierung von PbD für eine Big Data Lösung darstellt. Alleine das Separieren von personenbezogenen Daten über mehrere Standorte hinweg führt zu Laufzeitverlusten und Komplexitätsanforderungen, die kein Unternehmen freiwillig umsetzen wird.

Inform und Control erfordern umfassende Schnittstellen, die jedem Benutzer eine direkte Rückmeldung über die Nutzung seiner Daten mit einem späteren Eingriffsrecht geben. Soll dieser Eingriff dann auch umgehend in der Ergebnisse von Analysen eingehen, wie von Cavoukian und Jonas in [4] angeregt, so erfordert dies gegebenenfalls eine kontinuierliche Anpassung von Analyseergebnissen. Realisierbar sind solche umfassenden Anforderungen offensichtlich nur dann, wenn diese wirklich bereits in der Entwurfsphase eines Systems berücksichtigt werden, da hierzu spezifische Kontrollflüsse für Daten erforderlich sind.

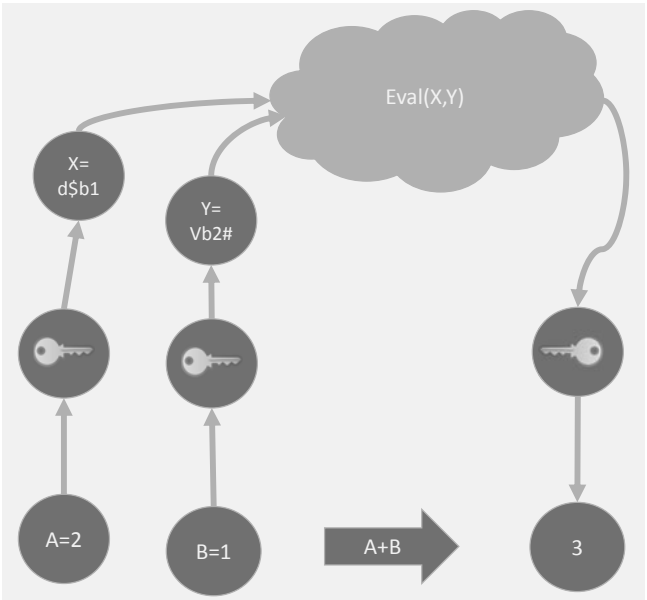
4 Privatheit und Datenweitergabe

Ein verbreiteter Ansatz bei der Handhabe personenbezogener Daten ist es, den Personenbezug aus den Daten zu entfernen, bevor diese gespeichert, verarbeitete oder weitergegeben werden. Somit entfallen die Anforderungen des Datenschutzes. Es bleibt allerdings eine nicht abschließend geklärte Herausforderung, wie erfolgreich das Entfernen des Bezugs wirklich ist: Die Verfügbarkeit von Big Data-Verfahren, die komplexe Zusammenhänge und verschiedenen Datenquellen verknüpfen können, kann in der Praxis die notwendigen Hürden deutlich erhöhen. Ein Beispiel zeigt die Arbeit von de Montjoye et al.[5], die das Aufdecken von anonymisierten Kreditkarteninformationen adressiert.

4.1 Anonymisieren et al.

Beim Anonymisieren werden identifizierende Merkmale aus den Datensätzen gelöscht. Dieser Vorgang soll nach dem DSGVO nicht oder nur mit unverhältnismäßig hohem Aufwand umkehrbar sein. Oft bestehen an diese Anonymität bestimmte Vorgaben, die beschreiben, wie groß eine Gruppe von Personen mindestens sein muss, auf die mittels der vorhandenen Daten eingegrenzt werden kann. Hier spricht man von k-Anonymität [6], wobei k die Größe der nicht unterscheidbaren Personengruppe bestimmt. Da sich k-Anonymität als angreifbar erwiesen hat und einer Re-Identifizierung von Personen durch Angriffe möglich ist [7], wurden mit l-diversity und t-closeness Nachfolger geschaffen, die die Zuverlässigkeit der Anonymität erhöhen. L-diversity fordert, dass es keine Klassen von sensiblen Daten gibt, deren Inhalt quasi identisch ist. Dies wird erreicht, indem in jeder Klasse mindestens l deutlich unterschiedliche Attribute vorhanden sind. So kann aus der Klasse nicht auf ein spezifisches Attribut gefolgert werden. T-closeness erweitert diese Forderung noch um den Umstand, dass die Verteilung eines Attributes in einer Klasse zumindest ähnlich zu der Verteilung in der Datensammlung ist, aus der die Klasse entnommen wurde.

Abbildung 1 | Homomorphe Verschlüsslung erlaubt Operationen auf verschlüsselten Daten



Beim Pseudonymisieren werden die Namen oder andere identifizierende Merkmale nicht einfach gelöscht, sondern durch ein Pseudonym ersetzt. Wer dieses Pseudonym kennt, kann den zur Person gehörenden Datensatz weiterhin identifizieren. Ein anderer Weg ist die Datenaggregation: Hier werden mehrere Datensätze zusammengefasst. Aus diesen zusammengefassten Werten kann dann nicht mehr auf individuelle Datensätze geschlossen werden. Schlussendlich ist auch noch die Datensynthese zu nennen. Sie basiert darauf, keine personenbezogenen Daten mehr direkt weiterzugeben, wie verschleiert diese auch sein mögen, sondern aus diesen Daten Modelle abzuleiten, anhand derer dann neue Daten synthetisiert werden. Diese Daten sollen dann die gleichen statistischen Eigenschaften haben und so äquivalente Untersuchungen und Analysen ermöglichen.

4.2 Homomorphe Verschlüsselung

Ein anderer Fall liegt vor, wenn Daten von einer dritten Partei verarbeitet werden sollen, ohne dass diese auf die Daten zugreifen kann. Ein Beispiel wäre ein Unternehmen, welches vertrauliche Daten erfasst, um diese von einem anderen Unternehmen analysieren zu lassen. Die Herausforderung hierbei liegt darin, dass die Analyse auf verschlüsselten Daten geschehen muss: Die Daten werden verschlüsselt übertragen und verarbeitet und erst das zurückerhaltene Ergebnis wird wieder entschlüsselt. Nur so ist sichergestellt, dass die Dritte Partei keinen Einblick in die vertraulichen Daten erhalten kann. Diese Idee wurde bereits 1978 von Rivest et al. eingeführt [8]. Das damalige motivierende Beispiel war eine externe Lohnbuchhaltung, die Idee lässt sich aber auf heutige Cloud-Dienste übertragen.

Die Methodik dahinter ist die homomorphe Verschlüsselung, eine Technologie, die verspricht die Sicherheit von Cloud- und Big Data Systemen deutlich erhöhen zu können, heute allerdings noch eher ein Forschungsgegenstand ist. Derzeit bekannte Lösungen führen zu einem selbst für Big Data-Verhältnisse sehr hohen Mehraufwand an benötigter Rechenleistung und Datenvolumen. Grundsätzlich gilt, wie in Abbildung 1 gezeigt, dass Daten verschlüsselt werden und auf diese dann eine Funktion angewandt wird, welche als Ergebnis wiederum verschlüsselte Daten erzeugt. Diese Daten können dann wieder entschlüsselt werden und das Ergebnis der Funktion auf den verschlüsselten Daten entspricht der Operation auf unverschlüsselten Daten.

Eine vollständige homomorphe Verschlüsselung gilt heute noch nicht als praktikabel. Sie wäre erreicht, wenn beliebige Operationen auf den verschlüsselten Daten durchgeführt werden können. Es existieren aber bereits partiell-homomorphe Verfahren, die nur eine Teilmenge aller Verfahren unterstützen und quasi-homomorphe Verfahren, die zwar beliebige Operationen erlauben, aber einen Rechenfehler einbringen, der auch als Rauschen bezeichnet wird.

5 Datennutzungskontrolle

Das Forschungsfeld der Datennutzungskontrolle (engl. Data Usage Control) erweitert klassische Zugriffskontrollmechanismen (engl. Access Control) und entwickelt Lösungsansätze zur umfassenden Steuerung der Datennutzung. Die grundlegende Idee: Der Dateneigentümer soll umfassende Kontrollmöglichkeiten erhalten, mit denen die Nutzung seiner Daten gesteuert wird.

Hierzu werden die klassischen Mechanismen der Zugriffskontrolle erweitert, damit die Nutzung von Daten kontrolliert werden kann und der Dateneigentümer jederzeit weiß, was mit seinen Daten geschieht (Transparenz). Ein wichtiger Punkt zur Erreichung dieser Ziele ist die Analyse und der kontrollierte Eingriff in Datenströme, um die Verwendung der Daten je nach Nutzungssituation kontrollieren zu können. Beispielsweise müssen einzelne Datenfelder je nach Datenempfänger oder Nutzungssituation eingeblendet oder ausgeblendet werden. Dies ermöglicht es Daten nutzerspezifisch oder geschäftsmodellspezifisch bereitzustellen und zu verarbeiten.

Das Sicherheitsframework IND²UCE (Integrated Distributed Data Usage Control Enforcement) macht Datennutzungskontrolle für die praktische Anwendung nutzbar. Der Dateneigentümer kann durch Sicherheitsrichtlinien (engl. Policies) die gewünschte Datennutzung präzise und feingranular kontrollieren. Dabei kann er einstellen, welche seiner Daten unter welchen Bedingungen wie oft gelesen, verändert, kopiert oder weitergeleitet werden dürfen. Es bestehen Möglichkeiten, spezielle (personenbezogene) Daten automatisiert zu anonymisieren, Nutzungen nur auf bestimmten Geräten oder Geräteklassen (z.B. Dienstgeräte des Dateneigentümers) zu erlauben und die Örtlichkeit bei der Datennutzung einzuschränken (z.B. nur innerhalb eines bestimmten Gebäudes oder innerhalb der Landesgrenzen). Des Weiteren können ausgewählte Daten nach einer genau definierten Anzahl von Tagen gelöscht bzw. unbrauchbar gemacht werden.

Abbildung 2 | IND²UCE Framework

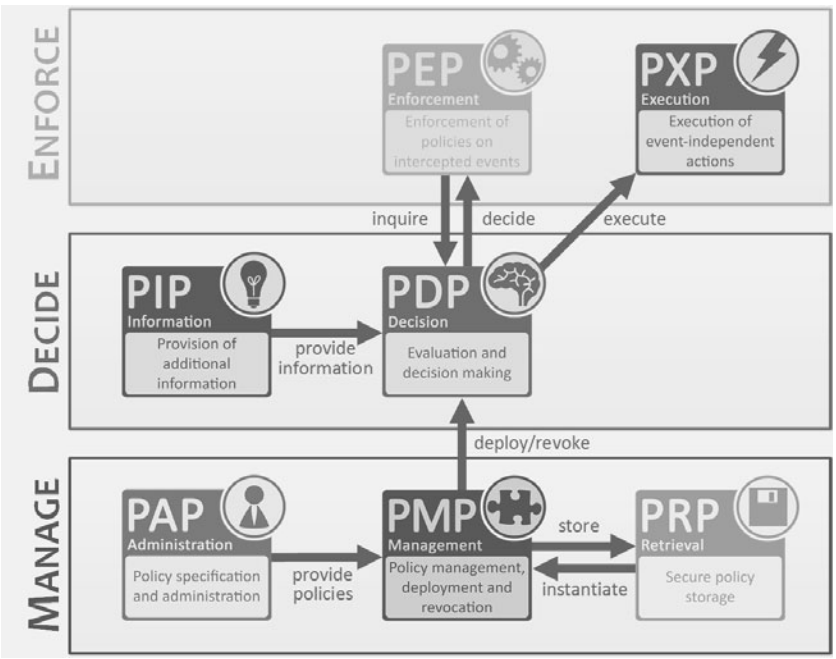


Abbildung 2 zeigt das IND²UCE Framework mit seinen drei Schichten „Manage", „Decide" und „Enforce" sowie den jeweils zugehörigen Komponenten. Im Zentrum des Frameworks steht eine generische Entscheidungskomponente (Policy Decision Point, PDP), welche anhand von Sicherheitsrichtlinien über die Legitimität von sicherheitsrelevanten Ereignissen (bspw. Datenoperationen) entscheidet. Diese Sicherheitsrichtlinien basieren auf dem Event-Condition-Action-Paradigma und erlauben zusätzlich die Verwendung der Obligation Specification Language (OSL) [11]. Mit Hilfe von OSL können zukünftige Verbindlichkeiten (engl. Obligation) spezifiziert werden (bspw. „Personenbezogene Daten müssen innerhalb von 14 Tagen gelöscht werden" oder „Ohne Genehmigung des Vorgesetzten dürfen nur 10 Akten pro Stunde geöffnet werden").

Durchsetzungskomponenten, so genannte Policy Enforcement Points (PEPs), sind Kontrollpunkte, welche in bestehende Systeme integriert werden, um Informationsflüsse gemäß den spezifizierten Sicherheitsrichtlinien kontrollieren zu können. PEPs erfassen relevante Ereignisse auf verschiedenen Systemebenen und lassen sie je nach Sicherheitsvorgabe zu, modifizieren oder verwerfen sie. Hierbei können Modifikationen, wie beispielsweise Anonymisierungen oder Aggregationen von Daten, sehr feingranular und situationsbedingt gesteuert werden.

Policy Execution Points (PXP) sind ereignisunabhängig und können daher zusätzliche Aktionen wie das Löschen von Daten, das Protokollieren von Operationen oder das Versenden von Benachrichtigungen durchführen. Die Minimalkonfiguration des IND²UCE-Frameworks zur Durchsetzung von Sicherheitsrichtlinien erfordert einen PDP für die Entscheidungsfindung und einen PEP zum Durchsetzen der Entscheidung.

Der Policy Retrieval Point (PRP) bietet einen gesicherten Speicher für Sicherheitsrichtlinien. Dieser muss gegen bösartige oder versehentliche Veränderung geschützt werden. Die einzigen Komponenten mit Zugriff auf diesen Speicher sind der PDP, um Sicherheitsrichtlinien zu beziehen, und der Policy Management Point (PMP), um Sicherheitsrichtlinien zu verwalten.

Wir unterscheiden für die Spezifikation und Verwaltung der Sicherheitsrichtlinien zwei Komponenten: den Policy Management Point (PMP) und den Policy Administration Point (PAP). Der PAP ist eine Mensch-Maschine-Schnittstelle, die dazu dient, Sicherheitsrichtlinien auf benutzerfreundliche Art und Weise zu spezifizieren, welche schlussendlich in eine maschinenlesbare Form überführt werden. PAPs müssen an den Wissenstand der Endanwender und die Sicherheitsbedürfnisse der Anwendungsdomäne angepasst werden.

Der PMP übernimmt die Verwaltung der spezifizierten Sicherheitsrichtlinien. Dazu gehört das Aushandeln, Aktivieren, Modifizieren und Zurückziehen von Sicherheitsrichtlinien. Die Komponenten Da der PDP die Sicherheitsrichtlinien vom PRP lädt, müssen diese vom PMP dort vorab hinterlegt werden. Zudem übernimmt der PMP die Verwaltung der einzelnen Komponenten. Jede Komponente registriert sich beim PMP mit Informationen über die eigene Funktion und zur Verfügung gestellten Kommunikationsschnittstellen. Dies ermöglicht dem Framework ein dynamisches Laufzeitverhalten.

Als letzte Komponente ist der Policy Information Point (PIP) zu nennen. Diese Komponente stellt zusätzliche Informationen bereit, welche für die Entscheidungsfindung im PDP benötigt werden und im Systemereignis nicht vorliegen. Zusätzliche Informationen können Daten über Informationsflüsse (engl. Information Flows) oder kontextabhängige Daten, wie etwa die aktuelle Lokation oder Wi-Fi-Konnektivität eines Endgeräts, sein. Kontextsensitivität erlaubt es Sicherheitsmechanismen nur dann scharf zu schalten, wenn diese in der Situation angebracht sind. Dies ermöglicht beispielsweise, allgemeine Verbote aufzulockern, zu welchen Unternehmen sonst gezwungen sind. Aus dem allgemeinen Verbot „Smartphones sind im Unternehmen verboten, da Fotos von geheimen Informationen gemacht werden können" kann eine kontextsensitive Sicherheitsrichtlinie entstehen, wie zum Beispiel „Fotos, welche mit einem Smartphone innerhalb des Unternehmens aufgenommen wurden, dürfen auch nur dort angesehen werden".

Das IND²UCE-Framework eignet sich besonders zum Einsatz in geschlossenen Systemen (z.B. Android, Unternehmensinfrastruktur). Je nach Einsatzszenario können Technologien wie TPM (Trusted Platform Module) eingesetzt werden, um eine Attestierung hinsichtlich der Integrität von IND²UCE vornehmen zu können. Dieser Ansatz wäre beim Einsatz in einem Rechenzentrum eines Unternehmens oder in Virtualisierungsumgebungen [10] denkbar. Am Beispiel von Android wurden zwei Versionen entwickelt: In der ersten Version wurde das Android-System angepasst und ein eigenes Android-Image erzeugt [9]. IND²UCE ist in diesem Fall Bestandteil das mobilen Betriebssystem und profitiert von dessen Sicherheitseigenschaften. Eine zweite Version ist weniger invasiv und ist in Form einer mobilen Anwendung erstellt worden. Diese verfügt zwar über Sicherheitsmerkmale wie „Device Administrator"-Funktionen, ist aber vom Sicherheitsniveau ganz anders einzustufen. Zusammenfassend kann man festhalten, dass sich je nach Nutzungsszenario verschiedene Möglichkeiten des Einsatzes für IND²UCE ergeben.

Ein weiterer wichtiger Punkt ist die Absicherung der Kommunikation zwischen den einzelnen Komponenten und deren Berechtigungen. Die Kommunikation zwischen den Komponenten kann verschlüsselt erfolgen. Im Fall einer Verschlüsselung findet eine gegenseitige Authentisierung mit Hilfe von digitalen Zertifikaten statt. Aktuelle Arbeiten befassen sich zusätzlich mit der Abbildung von Framework-Berechtigungen als Teil der Zertifikatsausstellung. Beispielsweise könnte ein spezieller PXP die Berechtigung bzw. Rolle bekommen Sicherheitsrichtlinien zu aktivieren oder zu deaktivieren, was üblicherweise dem PMP vorbehalten ist. Sicherheitsrichtlinien könnten damit nicht nur die Datennutzung kontrollieren, sondern auch dynamisch Sicherheitseinstellungen verändern.

Abschließend lässt sich festhalten, dass das Sicherheitsframework IND²UCE als Ergänzung oder Erweiterung bestehender Systeme wie Zugriffskontrolle konzipiert wurde. Es ergänzt also bestehende Systeme, um flexible Datenschutzanforderungen zu meistern und Datennutzungskontrolle praktisch umzusetzen, ersetzt diese jedoch nicht.

6 Ausblick

Datenschutz, ist ein Thema, welches immer wieder im öffentlichen Raum engagiert diskutiert wird. Vereinfacht gesagt treffen dabei zwei Meinungen aufeinander: Die einen sehen Datenschutz als Hemmnis an, den „Rohstoff" Daten effizient nutzen zu können, um Gewinne zu erzielen und Arbeitsplätze zu schaffen. Die anderen sehen den Schutz der Daten als ein Grundrecht und sehen die informationelle Selbstbestimmung als nicht verhandel-

bar an. Das hierbei teilweise ein sehr emotionaler Meinungsaustausch stattfindet, bleibt nicht aus.

Die in weiten Bereichen eher schwammig formulierte Rechtslage, die zum einen Datenschutz fordert, zum anderen aber auch zahlreiche Ausnahmen erlaubt, trägt das ihre dazu bei. Es ist beispielsweise eher unwahrscheinlich, dass ein Unternehmen von sich heraus in hohem Maße in Datenschutz investiert, wenn ein Gesetz dies nur im Rahmen des wirtschaftlich vertretbaren Aufwands verlangt. Datenschutzmaßnahmen werden so zu einem individuell gestaltbaren Vorgehen, welches ein Unternehmen gegebenenfalls aus seiner Philosophie, nicht aber aus äußeren Anforderungen heraus umsetzt europäische Datenschutzgrundverordnung (DSGVO).

Betrachtet man die neue Datenschutzverordnung, ist abzusehen, dass dieser Umstand auch in Zukunft anhalten wird. Auch hier sind die Anforderungen aus Sicht der Technik interpretierbar. Und so fordert dann auch der Bitkom in einer aktuellen Stellungnahme[2] zur DSGVO eine Begrenzung des Aufwands auf das „absolut notwendige", was aus Sicht der Wirtschaft nachvollziehbar ist. Aus Sicht der Sicherheitsforschung ist diese Haltung natürlich unbefriedigend, da diese anstrebt, den Stand der Technik voranzutreiben und durch Innovation das technisch Umsetzbare zu erweitern.

Am Beispiel von Privacy by Design kann dies gut nachvollzogen werden: Derzeit ist es für ein Unternehmen nachvollziehbar unattraktiv, sich diesem Konzept in seiner Fülle von Forderungen bei gleichzeitig geringer technischer Handreichung zu verschreiben. Die Folgen wären schwer absehbare Kosten und gegebenenfalls auch eine Beeinträchtigung der Konkurrenzfähigkeit, wenn andere Unternehmen Daten effizienter und aussagekräftiger interpretieren können.

Die DSGVO wird in absehbarer Zukunft an dieser Situation wenig ändern. Unter anderem sind zahlreiche Anforderungen erneut mit Einschränkungen versehen oder können national individuell interpretiert werden. Gegebenenfalls kann hier die technische Sicherheitsforschung zumindest dabei helfen, den Stand der Forschung und das Potential zukünftiger Verbesserung aufzuzeigen. Beispielhafte Umsetzungen der Ansätze, die in diesem Beitrag aufgezeigt wurden, können hier als Blaupause für die Weiterentwicklung der datenschutzfreundlichen Analyse von personenbezogenen Daten dienen: Privacy by Design als leitendes Prinzip, die Methoden der datenschutzfreundlichen Datenweitergabe und -verarbeiten als Mechanismen und IND²UCE als Framework zur Datennutzungskontrolle.

2 https://www.bitkom.org/Presse/Presseinformation/Datenschutzverordnung-sollte-einheitlich-angewendet-werden.html

Literatur

[1] Ann Cavoukian, Privacy by Design – The 7 Foundational Principles, https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf, veröffentlicht 2009 (überarbeitet 2011)

[2] Dominik Herrmann, Jens Lindemann, Obtaining personal data and asking for erasure: Do app vendors and website owners honour your privacy rights?, Sicherheit 2016

[3] Danezis, George, Josep Domingo Ferrer, Marit Hansen, Jaap Henk Hoepman, Daniel Le Métayer, Rodica Tirtea, and Stefan Schiffner: Privacy and Data Protection by Design – from policy to engineering. ENISA, December 2014, ISBN 9789292041083.

[4] Cavoukian, Ann and Jeff Jonas: Privacy by design in the age of big data. Technical report, Information and Privacy Commissioner, Ontario, Canada,June 2012.

[5] Y.-A. de Montjoye: Just four bits of credit card data can identify most anyone (Update) (2015).

[6] L. Sweeney: K-anonymity: A Model for Protecting Privacy. In: International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10, 5 (2002), 557–570,

[7] Machanavajjhala, Ashwin, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam: ldiversity: Privacy beyond kanonymity.ACM Trans. Knowl. Discov. Data, 1(1), March 2007, ISSN 15564681. 31,32

[8] Rivest, Ronald L, Len Adleman, and Michael L Dertouzos: On data banks and privacy homomorphisms. Foundations of secure computation, 4(11):169–180, 1978. 26

[9] Christian Jung, Denis Feth, Christian Seise: Context-Aware Policy Enforcement for Android. Software Security and Reliability (SERE), 2013 IEEE 7th International Conference on, vol., no., pp.40, 49, 18-20 June 2013.

[10] Christian Jung, Andreas Eitel, Reinhard Schwarz: Enhancing Cloud Security with Context-aware Usage Control Policies. INFORMATIK2014: Big Data- Komplexität meistern. Workshop on Provisioning and Management of Portable and Secure Cloud-Services (CloudCycle 2014), 22-26 September, 2014.

[11] M. Hilty, A. Pretschner, D. Basin, C. Schaefer, T. Walter (2007). A policy language for distributed usage control. In Computer Security–ESORICS 2007 (pp. 531-546). Springer Berlin Heidelberg.