Datenschutz durch Technik

Anonymisierung und Pseudsonymisierung

Moritz Klammler*

23. Juni 2016

Betreuer:

PD Dr iur Oliver Raabe Prof Dr Marc Strittmatter
Ass iur Mieke Lorenz

*Studiengang: Informatik (Master), Fachsemester: 4, Matrikel-Nummer: XXXXXXX,
Kontakt: moritz.klammler@student.kit.edu.

Dieses Dokument und die zugehörige Präsentation sind zusammen mit dem Quellcode über die Webseite des Autors verfügbar.

`http://klammler.eu/data/jura/kit/anon-pseudo/`

Inhaltsverzeichnis

1 Motivation

Das Datenschutzrecht schützt (ausschließlich) personenbezogene Daten. Für seine Anwendbarkeit ist es daher entscheidend, ob ein gegebenes Datum einen Personenbezug aufweist. Die Unterscheidung in personenbezogene und nicht personenbezogene Daten und die daraus resultierenden Rechtsfolgen trifft das Datenschutzrecht nach gängiger Auffassung in absoluter Weise, ohne dabei Zwischenmöglichkeiten von „mehr oder weniger personenbezogenen" Daten in Erwägung zu ziehen [10, § 1]. Ziel dieser Arbeit soll es allerdings genau sein, hier eine nähere Differenzierung vorzunehmen, und zu untersuchen, inwiefern und mit welcher Begründung verschiedene Arten von Daten datenschutzrechtlich einzuordnen sind.

Pseudonyme Daten beschreiben personenbezogene Daten, bei denen die Zuordenbarkeit zu natürlichen Personen durch den Verarbeitungsschritt der *Pseudonymisierung* zwar entfernt wurde, zumindest die Stelle, die die Pseudonymisierung vorgenommen hat, den Personenbezug jedoch wieder herstellen kann.

Anonyme Daten beschreiben zwar Informationen zu Sachverhalten, die natürliche Personen betreffen, jedoch wurden diese durch einen Verarbeitungsschritt – die *Anonymisierung* – derart verändert, dass sie von jedermann nicht oder nur mit unverhältnismäßigem Aufwand einer natürlichen Person zugeordnet werden können.

Verschlüsselte personenbezogene Daten sind ein für sich genommen bedeutungsloses Chiffrat, das durch ein kryptographisches Verfahren – die *Verschlüsselung* – aus personenbezogenen Daten gewonnen wurde, und nach aktuellem Stand der Wissenschaft nur mithilfe eines (geheimen) Schlüssels mit realistischem Aufwand wieder in eine Form überführt werden kann, die Aufschluss über die enthaltenen Informationen gibt.[1]

Sachdaten (bzw nicht personenbezogene Daten) sind schließlich Daten, die Informationen zu Sachverhalten beschreiben, die bereits ihrem Wesen nach keinen Bezug zu natürlichen Personen aufweisen, etwa meterologische Daten.

Sachdaten sind aus datenschutzrechtlicher Sicht nicht weiter interessant, auch wenn sie sensible Informationen, etwa in Form von Betriebsgeheimnis-

[1] Sowohl nach juristischem als auch nach mathematischem Verständnis bezeichnen *Daten* und *Informationen* unterschiedliche Konzepte. Daten sind eine Folge von Zeichen; Information ist deren Aussagegehalt. Im juristischen Sprachgebrauch wird jedoch häufig schlampigerweise nur von „Daten" gesprochen.

sen darstellen können. Ebenso soll es im Rahmen dieser Arbeit nicht um unstrittig und offensichtlich personenbezogene Daten gehen. Anonyme, pseudonyme und verschlüsselte personenbezogene Daten werden durch einen eigenen Verarbeitungsschritt aus personenbezogenen Daten gewonnen. Anstatt uns sofort der Frage zu widmen, ob es sich danach immer noch um personenbezogene Daten handelt, wollen wir uns zunächst der Frage zuwenden, welche Gründe dafür sprechen könnten, diese Schritte vorzunehmen. Klar ist, dass alle diese Maßnahmen den Informationsgehalt der Daten reduzieren – im Falle der Pseudonymisierung und Verschlüsselung jedoch in reversibler Weise. Welche Gründe könnten dafür sprechen, einen solchen Schritt vornehmen zu wollen?

Einerseits könnte der Betroffene selbst ein Interesse daran haben, dies zu tun. Das naheliegendste Beispiel wäre wohl, dass ein Benutzer anstatt seines Klarnamens ein Pseudonym verwendet. § 13 Abs 6 TMG sieht diese Option explizit vor. Wir werden uns diesem Fall im Folgenden jedoch nicht weiter widmen.

Für die Zwecke unserer Betrachtungen interessanter ist der Fall, in dem die Stelle, welche die Daten erhebt, den Schritt vornimmt. Auch dazu kann es gute Gründe geben. Einerseits gebietet es das Prinzip der Datensparsamkeit, dass nicht mehr personenbezogene Daten verwendet werden, als für den jeweiligen Zweck erforderlich. Hierzu verlangt § 3a BDSG (eigene Hervorhebung):

> Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten und die Auswahl und Gestaltung von Datenverarbeitungssystemen sind an dem Ziel auszurichten, so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen. *Insbesondere sind personenbezogene Daten zu anonymisieren oder zu pseudonymisieren*, soweit dies nach dem Verwendungszweck möglich ist und keinen im Verhältnis zu dem angestrebten Schutzzweck unverhältnismäßigen Aufwand erfordert.

Es ist denkbar, dass die verarbeitende Stelle diese Informationsreduktion entweder sofort bei der Erhebung vornimmt, oder aber, dass sie den Schritt durchführt, bevor sie die Daten (oder eine *Kopie* davon) an eine andere Stelle übermittelt, etwa um sie von dieser in ihrem Auftrag speichern oder verarbeiten zu lassen. Neben ethischen Erwägungen und der Wahrung des gesetzlich geforderten Grundsatz' der Datensparsamkeit könnte sich die Stelle davon auch erhoffen, für den Umgang mit den nunmehr „weniger brisanten" Daten schwächeren datenschutzrechtlichen Auflagen zu unterliegen.

2 Legaldefinitionen und Rechtsfolgen

2.1 Bundesdatenschutzgesetz

§ 3 Abs 6 BDSG definiert *Anonymisieren* als

> […] das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.

und § 3 Abs 6a BDSG definiert *Pseudonymisieren* als

> […] das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.

Knopp [13, § 2.1] stellt fest, dass die Anforderungen an eine wirksame Pseudonymisierung („wesentlich erschweren") schwächer sind, als jene an eine wirksame Anonymisierung. Er lässt die Frage offen, ob dies den Willen des Gesetzgebers ausdrückt, oder einem Versehen geschuldet ist, und verweist darauf, dass etwa das Landesdatenschutzgesetz Schleswig-Holsteins für beide Verfahren dieselben Anforderungen stellt.

Das BDSG „belohnt" eine Pseudonymisierung jedenfalls nicht direkt. Tatsächlich kommt das Wort im ganzen Gesetz genau zweimal – nämlich in den oben zitierten § 3 Abs 6a und § 3a – vor. Allerdings schreibt das BDSG an einigen Stellen eine Abwägung der schutzwürdigen Interessen des Betroffenen gegen jene der verarbeitenden Stelle vor. In dieser Abwägung kann berücksichtigt werden, dass die Verarbeitung pseudonymer Daten die Interessen des Betroffenen in der Regel schwächer beeinträchtigen wird. [10, S 524]

Eine Definition für *Verschlüsseln* findet sich im BDSG nicht. Jedoch wird in der Anlage zu § 9 Satz 1 klargestellt, dass „insbesondere die Verwendung von dem Stand der Technik entsprechenden Verschlüsselungsverfahren" zu den regelmäßig zu ergreifenden technischen und organisatorischen Maßnahmen zählt.

2.2 Datenschutz-Grundverordnung

Die zukünftige EU-Datenschutz-Grundverordnung (Datenschutz-GVO) definiert *Pseudonymisierung* deutlich konkreter als

> [...] die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden[.]
>
> Art 4, Ziff 5 der Datenschutz-GVO [1, S 33]

Die Verordnung erwähnt Pseudonymisierung an zahlreichen Stellen als standardmäßig einzusetzendes Mittel in der Datenverarbeitung. Demgegenüber erwähnt sie Anonymität lediglich in Erwägungsgrund 26 [1, S 5] und dort nur, um klarzustellen, dass die Verordnung den Umgang mit solchen Daten nicht regeln soll.

Eine „Belohnung" für das Pseudonymisieren kennt die Datenschutz-GVO ebenfalls nicht. Im Gegenteil macht Erwägungsgrund 26 klar, dass pseudonyme Daten personenbezogene Daten im Sinne der Datenschutz-GVO sind (so auch Karg [10, S 524]).

> Einer Pseudonymisierung unterzogene personenbezogene Daten, die durch Heranziehung zusätzlicher Informationen einer natürlichen Person zugeordnet werden könnten, sollten als Informationen über eine identifizierbare natürliche Person betrachtet werden.
>
> Erwägungsgrund 26 der Datenschutz-GVO [1, S 5]

Sinnvollerweise wird sie dahingehend zu verstehen sein, dass der Einsatz von Pseudonymisierung wo immer möglich zu den Grundsätzen der Datensparsamkeit zählt, und somit grundsätzlich durchgeführt werden muss, ohne dass daraus eine Minderung der Sorgfaltspflichten abzuleiten wäre.

3 Wann sind Daten anonym?

Gelingt es, den Personenbezug von Daten durch Anonymisierung zu entfernen, können diese weitestgehend wie Sachdaten behandelt werden. Es stellt sich jedoch die Frage, ab wann davon ausgegangen werden kann, dass kein Personenbezug mehr vorliegt. Insbesondere in der deutschsprachigen Literatur haben sich dabei die Schulen der *absoluten* und *relativen* Bestimmbarkeit herauskristallisiert. Siehe dazu etwa die Ausführungen des BGH [2, Rn 23 ff] und die darin zitierte Literatur.

3.1 Absolute versus relative Bestimmbarkeit

Die Lehre der absoluten Bestimmbarkeit geht davon aus, dass ein personenbezogenes Datum vorliegt, wenn es *irgendeine* Stelle gibt, die in der Lage ist, die betroffene Person zu bestimmen. Diese Auffassung scheint in Europa außerhalb Deutschlands die vorherrschende Meinung zu sein (siehe Stiemerling und Hartung [17, II.1.b] und diverse Stellungnahmen vor dem EuGH [3, Rn 31 ff]).

Wie unter anderem Stiemerling und Hartung [17, II.1.b] feststellen, scheint Erwägungsgrund 26 der Europäische Datenschutzrichtlinie [4, S 33] eine absolute Auffassung nahezulegen, indem dort die Formulierung „alle Mittel […], die vernünftigerweise entweder von dem Verantwortlichen für die Verarbeitung oder von einem Dritten eingesetzt werden könnten, um die betreffende Person zu bestimmen“ verwendet wird. Derselbe Erwägungsgrund der Datenschutz-GVO [1, S 5] spricht von „alle[n] Mittel[n] […], die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren“. Es darf davon ausgegangen werden, dass der Gesetzgeber die Zusätze „oder einem Dritten“ beziehungsweise „oder einer anderen Person“ nicht verwendet hätte, wenn er diese Auslegung nicht gewünscht hätte.

Interessenterweise erkennt etwa das LG Berlin diese Formulierung zwar, argumentiert ihre Bedeutsamkeit allerdings in kaum nachvollziehbarer Weise weg [5, Rn 135].

Von den deutschen Gerichten wird bislang noch überwiegend – aber keineswegs ausschließlich (siehe etwa die vom LG Berlin zitierten Entscheidungen [5, Rn 132 ff]) – die Ansicht vertreten, dass ein Personenbezug nur dann vorliegt, wenn die verarbeitende Stelle selbst mit den ihr technisch und rechtlich zur Verfügung stehenden Mitteln in der Lage ist, die betroffene Person zu bestimmen. Eine höchstrichterliche Entscheidung in dieser Sache gibt es bisweilen nicht. Auch in seiner Vorlage an den EuGH zur Speicherung dynamischer IP-Adressen [2] hat der BGH diese Frage zumindest nicht explizit gestellt.

Nach dem relativen Verständnis können dieselben Daten für die eine Stelle personenbezogen sein, für eine andere dagegen nicht. Das führt zu der etwas kuriosen und meines Erachtens unzulänglichen Konsequenz, dass ein Datum für die verarbeitende Stelle zwar kein personenbezogenes Datum darstellt, ihr aber dennoch aufgegeben ist, es nicht an Dritte weiterzugeben, in deren Händen es den Personenbezug wieder entfalten könnte (siehe dazu etwa das LG Berlin [5, S. 135–136]). Es ist schwer einzusehen, wie ein solcher Schwebezustand von „aktuell nicht aber in Zukunft möglicherweise wieder

personenbezogenem“ Datum geeignet ist, die Rechte des Betroffenen ausreichend zu schützen. Insbesondere träfen die verarbeitende Stelle ja nicht die Sorgfaltspflichten für den Umgang mit personenbezogenen Daten, sodass nicht ersichtlich ist, wie eine (auch nur versehentliche) Weitergabe an eine Stelle, die den Personenbezug herstellen kann, wirksam verhindert werden sollte. Schließlich ist es auch zu befürchten, dass durch ein nach diesem Verständnis erlaubtes Austauschen von Daten, die für die beteiligten Stellen für sich genommen jeweils keinen Personenbezug haben, Kombinationen von Daten entstehen, die eine Bestimmung des Betroffenen wieder erlauben.

Karg [10, S 525] erwähnt außerdem noch einen dritten, „vermittelnden“ Ansatz zwischen den Extremen des absoluten und relativen Verständnis von Bestimmbarkeit und vertritt die Ansicht, dass diese vom Wortlaut der Datenschutz-GVO nahegelegt würde, was plausibel scheint.

3.2 Realistischer Aufwand

Weiter stellt sich die Frage, wie viel Aufwand berücksichtigt werden soll, um eine Person zu bestimmen. Die entsprechenden Rechtsnormen sprechen dabei von „verhältnismäßigem Aufwand an Zeit und Kosten“ (§ 3 Abs 6 BDSG), oder Mitteln, die „vernünftigerweise eingesetzt werden können“ (Erwägungsgrund 26 der Datenschutz-RL [4, S 33]) oder „nach allgemeinem Ermessen wahrscheinlich genutzt werden“ (Erwägungsgrund 26 der Datenschutz-GVO [1, S 5]).

Diese unbestimmten Rechtsbegriffe fordern allerdings einen gewissen Auslegungsaufwand. Karg [10, S 526] geht davon aus, dass im Allgemeinen eine „wirtschaftlich und rational handelnde verantwortliche Stelle“ zugrundegelegt werden kann und von anonymen Daten ausgegangen werden können soll, „[w]enn zwischen dem Aufwand der Re-Identifizierung und dem damit verfolgten Zweck ein offensichtliches, eklatantes Missverhältnis besteht“. Allerdings hält er ebenfalls fest [10, S 526], dass insbesondere staatlichen Stellen (vor allem Geheimdiensten) nicht immer wirtschaftlich rationales Handeln unterstellt werden kann. Weiter stellt er fest, dass für wissenschaftliche Zwecke gerade die Überwindung – auch mit *unverhältnismäßigen* Mitteln – vermeintlich sicher geglaubter Hindernisse das Ziel sein könnte. Das stets wieder massive öffentliche Interesse an der (vermutlich illegalen) Analyse geleakter Datenbestände, für die teils erheblicher Aufwand ohne erkennbaren wirtschaftlichen Nutzen betrieben wird, scheint diese Vermutung zu stützen.

Interessant ist, dass die Legaldefinitionen von Anonymisiereung und Verschlüsselung teilweise eng beieinander liegen (so etwa festgestellt von Knopp [12, § 2]).

4 Exkurs: Mathematische Sicherheitsbegriffe

Im Rest dieser Arbeit soll nun thematisiert werden, wie mathematische Konstrukte dabei helfen können, anonyme und pseudonyme Daten zu erzeugen, und wie sie rechtlich zu bewerten sind. Da die juristische Literatur leider nicht an Beispielen armt, die von einem erheblichen Missverständnis geprägt zu sein scheinen, wollen wir zunächst einige grundlegende Definitionen aus der Kryptographie einführen. Gute kryptographische Einführungen bieten etwa die Werke von Schneier [16] (eher praktisch und nicht mehr ganz aktuell) oder Katz und Lindell [11] (eher theoretisch).

Grundsätzlich ist es nicht möglich, mit rein mathematischen Verfahren Information zu gewinnen. Auch verschlüsselte Daten müssen daher also aus informationstheoretischer Sicht dieselbe Information enthalten, wie die unverschlüsselten Daten. Jedoch kann man die Daten „in zwei Teile aufteilen", die jeweils mindestens so lang wie die Originalnachricht sind, und die man zunächst willkürlich *Schlüssel* und *Chiffrat* nennt, sodass jeder Teil für sich genommen keine Information außer der Nachrichtenlänge mehr enthält, und die Information erst durch Rekombination der beiden Teile wiedergewonnen werden kann. Dieses Verfahren ist als *One-Time-Pad* bekannt, und mathematisch unmöglich zu brechen. Leider ist es in der Praxis weitestgehend nutzlos, denn während man die eine Hälfte (nennen wir sie das Chiffrat) bedenkenlos auch nicht vertrauenswürdnigen Dritten zugänglich machen kann, steht man nun vor dem Problem, die andere Hälfte (den Schlüssel) sicher verwahren beziehungsweise transportieren zu müssen, was einen zu dem Ausgangsproblem zurückführt. (Man rufe sich in Erinnerung, dass beide Teile mindestens so lang sind, wie der Klartext.) Interessant kann dieses Verfahren allenfalls dann sein, wenn ein sicherer aber langsamer und ein unsicherer aber schneller Transportweg zur Verfügung stehen. Man überträgt dann den Schlüssel *vorab* auf dem langsamen sicheren Weg und sendet das Chiffrat – so die Daten angefallen sind – auf dem schnellen unsicheren Weg zurück. Außerhalb eng abgegrenzter hoheitlicher Anwendungsbereiche wird dies selten der Fall oder zumindest nicht wirtschaftlich sein.

Interessante kryptographische Verfahren versuchen daher, die Länge des Schlüssels klein zu halten. Üblicherweise wird man einen Schlüssel konstanter Länge benutzen wollen, um Nachrichten unbegrenzter Länge verschlüsseln zu können. Ein solches Verfahren kann im mathematischen Sinne keine vollständige Sicherheit liefern. Denn während eine Nachricht aus n Zeichen über einem binären Alphabet prinzipiell für 2^n verschiedene Texte stehen kann, kann ein gegebenes Chiffrat der Länge n, das mit einem Schlüssel der Länge k verschlüssselt wurde, allenfalls für 2^k verschiedene Texte stehen. Wenn k deutlich kleiner ist als n, kann ein Angreifer theoretisch immer

durch Durchprobieren aller Schlüssel mit hoher Wahrscheinlichkeit die Originalnachricht finden.[2] Dieses Vorgehen ist allerdings denkbar ineffizient und selbst für vergleichsweise kleine Werte von k unpraktikabel. Man parametrisiert kryptographische Verfahren daher mit einem sogenannten Sicherheitsparameter (eben diesem k, das die Schlüssellänge angibt). Die Idee dabei ist, dass man durch entsprechende Wahl von k die Sicherheit des Verfahrens individuell einstellen und der zu erwartenden Mächtigkeit potentieller Angreifer anpassen kann.

Leider ist es nicht bekannt, ob es ein Verfahren gibt, bei dem das ineffiziente Durchprobieren aller möglichen Schlüssel den effektivsten Angriff darstellt. Allerdings kann man Verfahren V angeben, für die sich Aussagen der Form

> Angenommen, es gibt keine effiziente Möglichkeit, um Problem Π zu lösen. Unter dieser Annahme gibt es auch keinen effizienten Angriff auf das Verschlüsselungsverfahren V.

beweisen lassen. Das führt allerdings nur dazu, dass man nun in Bezug auf Π in derselben Situation ist, wie zuvor für V. Man hofft nun aber, Probleme Π einsetzen zu können, von denen es sehr naheliegend ist, dass sie nicht effizient gelöst werden können. Es gibt zwar eine sehr prominente Klasse solcher Probleme (die Klasse der sogenannten $\mathcal{NP}$-vollständigen Sprachen), allerdings ist die aktuell bekannte Mathematik aus sehr fundamentalen Gründen bisweilen nicht ausdrucksstark genug, um diese Probleme als Grundlage für kryptographische Verfahren zu nutzen. Man behilft sich stattdessen mit Problemen Π, für die man bestenfalls sagen kann, dass trotz allgemeinen Interesses noch niemand eine effiziente Möglichkeit gefunden hat, sie zu lösen.

Die vollständige Wahrheit ist, dass gängige Verfahren wie der *Advanced Encryption Standard* (AES) jegliche theoretische Beweisbarkeit zugunsten effizienter Implementierbarkeit opfern, und ihre Sicherheit ausschließlich darauf beruht, dass auch in Jahrzehnten öffentlicher Forschung bisweilen niemand eine wirksame Angriffsmöglichkeit finden konnte. Dies kann sich freilich jederzeit ändern.

Wenngleich diese Ausgangslage einigermaßen ernüchternd wirken mag, kann man festhalten, dass Kryptographie in der Praxis sehr erfolgreich ist, und es Verfahren gibt, deren (theoretisch natürlich unbewiesene) Sicherheit auch nach Jahrzehnten intensiver Anwendung in der Praxis immer noch ungebrochen ist. Dies mag aus mathematischer Sicht frustrierend sein, entspricht

[2] Tatsächlich fordert man von einem sicheren kryptographischen Verfahren, dass es nicht nur schwer sein soll, die Originalnachricht aus einem gegebenen Chiffrat zu berechnen, sondern dass es aussichtslos sein soll, *irgendeine* nützliche Information, außer der Nachrichtenlänge, zu gewinnen. Man nennt dies *semantische Sicherheit.*

aber durchaus dem, was man auch aus anderen Lebensbereichen (außerhalb der Informatik) kennt, wo es auch selten mathematische Beweise für die formale Korrektheit der geübten Praxis gibt.

5 Pseudonymisierung

Betrachten wir nun einige Verfahren zur Pseudonymisierung.

5.1 Zufällig generierte IDs

Eine naheliegende – und sichere – Möglichkeit, um Daten zu pseudonymisieren, besteht darin, eine Tabelle mit Zuordnungen von Klarnamen[3] und Pseudonymen zu verwenden. Für jeden neuen Klarnamen fügt man ein echt zufällig generiertes Pseudonym in die Tabelle ein. Derart gewälte Pseudonyme lassen prinzipiell keinen Rückschluss auf den Klarnamen zu. Problematisch ist allerdings, dass die Tabelle mit den Zuordnungen datenschutzkonform gespeichert werden muss. Weiß man, dass in Zukunft keine neuen Personen mehr hinzukommen werden, und ist eine De-Pseudonymisierung nicht erwünscht, kann man die Zuordnungstabelle nach getaner Arbeit auch wegwerfen.

Will man es vermeiden, eine explizite Tabelle mit den Zuordnungen von Pseudonymen zu Klarnamen speichern zu müssen, kann man auch eine Funktion definieren, die *on-demand* ein Pseudonym zu einem gegebenen Klarnamen berechnet. Natürlich soll die Funktion nicht einfach zu invertieren sein – ansonsten liefe die Pseudonymisierung ins Leere. Je nach Anwendung kann es jedoch gewünscht sein, dass eine privilegierte Stelle (die im Besitz eines geheimen Schlüssels ist) in der Lage ist, die Pseudonyme wieder Klarnamen zuzuordnen.

5.2 Symmetrische Verschlüsselung

Betrachten wir als nächstes die Möglichkeit, die Klarnamen mit einem (geheimen) Schlüssel zu verschlüsseln. Wer im Besitz des geheimen Schlüssels ist, kann die derart gewonnenen Pseudonyme leicht wieder Klarnamen zuordnen, ohne dass eine Tabelle gespeichert werden müsste. Für andere Stellen ist es – jedenfalls nicht ohne Zuhilfenahme von Weltwissen – nicht möglich, aus einem Pseudonym zurück auf den Klarnamen zu schließen.

Ein gutes Verschlüsselungsverfahren hat eine Eigenschaft, die man IND-CPA-Sicherheit nennt, und die äquivalent zu der in Fußnote 2 erwähnten semantischen Sicherheit ist. Sie bewirkt, dass ein Angreifer nicht effektiv zwi-

[3] Selbstverständlich muss der „Klarname“ kein bürgerlicher Name sein. Es kann sich um jedes identifizierende Merkmal handeln, das zu verbergen wünschenswert erscheint.

schen den Chiffraten zweier (auch selbst gewählter) Nachrichten unterscheiden kann. Entsprechend kann er also auch insbesondere nicht feststellen, ob sich ein Eintrag zu einem gegebenen Namen in den Daten befindet. Will man allerdings, dass die Stelle, die die pseudonymen Daten verarbeitet, in der Lage ist, Identitäten zu vergleichen, so wird man zumindest dasselbe Merkmal immer durch dasselbe Chiffrat ersetzen müssen. Das erlaubt es allerdings unter Umständen, Personen anhand von Weltwissen zu identifizieren.

Müller-Quade, Huber und Nilges [14] haben ein interessantes Verfahren entwickelt, vermittels dessen eine relationale Datenbank in einer für die Anwendung transparenten Weise so modifiziert werden kann, dass aus der Datenbank alleine – die Sicherheit des verwendeten Verschlüsselungsverfahrens unterstellt – keine Rüchschlüsse über die gespeicherten Beziehungen der Merkmale gewonnen werden können. Wenngleich in der Publikation nicht explizit vorgesehen, ließe sich das Schema zumindest für textuelle Attribute, auf denen man nicht rechnen will, durch zusätzliche Verschlüsselung dieser Attribute ergänzen. Nachdem es immer noch eine Stelle gibt, die die Daten entschlüsseln kann – ansonsten wäre das Verfahren reichlich nutzlos – dies für den Betreiber des Datenbank-Servers dagegen praktisch ausgeschlossen ist, müssten die Daten wohl als pseudonymisiert betrachtet werden.

5.3 Kryptographische Hash-Funktionen

Eine kryptographische Hash-Funktion ist eine mathematische Funktion, die beliebige Eingaben auf Zeichenfolgen – sogenannten *Hashes* – einer festen (durch den Sicherheitsparameter k gegebenen) Länge abbildet. Diese Abbildung hat die Eigenschaft, dass sie zwar effizient berechenbar ist, es aber nicht effizient möglich ist, zu einem gegebenen Hash-Wert eine Zeichenfolge zu berechnen, bei deren Eingabe die Hash-Funktion den gegebenen Hash ausgeben würde. Für eine gute kryptographische Hash-Funktion fordert man noch eine stärkere Eigenschaft, nämlich, dass es schwierig sein soll, zwei unterschiedliche Eingaben zu finden, die denselben Hash-Wert generieren. Diese Eigenschaft bezeichnet man als *Kollisionsresistenz*.

Es ist unbekannt, ob solche Funktionen überhaupt existieren, geschweigedenn ist eine konkrete Funktion bekannt, von der man diese Eigenschaft mathematisch beweisen könnte. Wie auch bei den Verschlüsselungsverfahren gibt es jedoch populäre Verfahren, die sich in der Praxis bewährt haben. Allerdings wurden in letzter Zeit einige einstmals als sicher geglaubte Hash-Funktionen (namentlich MD5 und SHA-1) theoretisch und im Fall von MD5 auch praktisch zerstört.

Eine kryptographische Hash-Funktion kann als Einweg-Verschlüsselung benutzt werden. Etwa könnte man die Namen in einer Datei durch ihre Hash-

Werte ersetzen. Die Kollisionsresistenz der Funktion stellt dabei sicher, dass es extrem unwahrscheinlich ist, dass dabei zwei Hashes kollidieren, sodass die Daten eindeutig bleiben. Es ist aber (für jedermann) nicht mehr möglich, aus den Hash-Werten zurück auf die Namen zu schließen. Allerdings kann man – wenn man bereits weiß, nach welchem Namen man suchen möchte – diesen hashen und sodann den Hash-Wert in der Datei nachschlagen. Da die Hash-Funktion deterministisch ist, funktioniert dies.

6 Anonymisierung

Sämtlichen Personenbezug so aus den Daten zu entfernen, dass es dauerhaft und für jedermann unmöglich ist, die Betroffenen Personen zu bestimmen, ist ein ambitioniertes Vorhaben.

6.1 Simulierte Daten

Eine sehr sichere Möglichkeit, um alle Personenbezüge aus den Daten zu entfernen, wäre es, die gesamten Daten dauerhaft zu löschen. Dies für sich genommen mag zwar sicher, aber auch denkbar wenig hilfreich sein. Allerdings kann man die Daten, bevor man sie löscht, verwenden, um statistische Zusammenhänge zu extrahieren, auf deren Grundlage man neue Daten *simuliert*. Da die derart erzeugten Daten das Ergebnis eines Zufallsprozess' sind, enthalten sie grundsätzlich keinen Personenbezug. Man wird jedoch vorsichtig sein müssen, was die Größe der Stichprobe betrifft. Ist sie klein im Vergleich zur Anzahl der untersuchten Merkmale, werden auch die simulierten Daten Rückschlüsse auf die ursprünglichen Betroffenen zulassen, da in diesem Fall die statistischen Methoden den Datensatz im Wesentlichen *auswendig gelernt* haben werden, anstatt echte statistische Zusammenhänge zu extrahieren. Wenngleich die Verwendung simulierter Daten sehr datenschonend ist, und sich etwa für Beispiele in der Lehre hervorragend eignet, sind die aus den simulierten Daten gewonnenen Ergebnisse im Vergleich zu Analysen auf Originaldaten natürlich mit einer gewissen Vorsicht zu genießen.

6.2 Verschlüsselung

Die juristische Literatur ist sich uneins darüber, ob verschlüsselte Daten als anonym betrachtet werden können. Zunächst ist klar, dass bei einem absoluten Verständnis von Bestimmbarkeit verschlüsselte Daten niemals anonym sein können, weil die Stelle, die im Besitz des (geheimen) Schlüssels ist, die Daten jederzeit wieder entschlüsseln und damit den Personenbezug offenkundig wiederherstellen kann.

Selbst wenn man von einem vermittelnden oder relativen Ansatz ausgeht, ist klar, dass auch mit den verschlüsselten Daten nicht völlig bedenkenlos umgegangen werden kann. Für die Stelle, welche den geheimen Schlüssel innehat, bleiben die verschlüsselt gespeicherten Daten selbstverständlich personenbezogen, sodass sie insbesondere nicht von ihren Auskunfts- oder Löschungspflichten entbunden ist. Das setzt zumindest geeignete vertragliche Abmachungen mit einem Dienstleister voraus, der die Daten im Auftrag speichert, um diesen Pflichten gewissenhaft nachkommen zu können. Diese Ansicht, der soweit fraglos zuzustimmen ist, findet sich etwa bei Knopp [12, § 3].

Eine interessante Frage ist jedoch, ob verschlüsselte Daten als anonym betrachtet werden können, wenn der geheime Schlüssel zuverlässig gelöscht wurde. Bejaht man diese Frage, eröffnet sich dadurch eine interessante Möglichkeit zur einfachen und zuverlässigen Löschung großer Datenbestände. Ob dieses Verständnis nach aktuell herrschender Meinung den Anforderungen des Datenschutzrechts genügt, scheint jedoch nicht klar zu sein.

Etwa fordert die Verordnung über die Durchführung von Erhebungen zum Zwecke wissenschaftlicher Forschung in Schulen des Saarlands in der Fassung vom 2. Dezember 2015,

> - dass die Daten auch in verschlüsselter Form Dritten nicht zugänglich gemacht werden,
> - dass die Daten zu anonymisieren sind, sobald dies ohne Beeinträchtigung des Erfolgs der Untersuchung möglich ist [und]
> - dass auch die in verschlüsselter Form gespeicherten Daten nach ihrer Auswertung zu löschen sind,
>
> § 4 SchulErhV SL

sodass man klar am Wortlaut der Norm vorbei argumentieren müsste, um die oben vertretene Auffassung zu legitimieren.

Ein (echtes) Problem, das sich aus dem „Löschen durch Vernichten des Schlüssels“ ergibt, ist, dass es dadurch für einen Angreifer natürlich in keiner Weise schwieriger wird, sich daran zu versuchen, die nach wie vor vorhandenen Daten unerlaubt zu „entschlüsseln“. Im Gegenteil arbeitet die Zeit und der damit einhergehende technische Fortschritt bezüglich der ihm realistischerweise zur Verfügung stehenden Rechenleistung, für den Angreifer.

Pordesch und Steidle [15, §§ 3.1 ff] stellen fest, dass ein gewisses Dilemma besteht, indem das BDSG einerseits keine zeitliche Befristung für die Schutzbedürftigkeit von personenbezogenen Daten kennt, andererseits aber der (sich stetig ändernde) Stand der Technik bestimmt, was eine zureichende Sicherung ist. Knopp [12, Fn 16] verweist dagegen auf eine Ansicht, dass

der Personenbezug zumindest über den Tod hinaus abnehme. Zumindest die Datenschutz-GVO unterstützt diese Ansicht – der tendenziell zuzustimmen ist – auch im Wortlaut:

> Diese Verordnung gilt nicht für die personenbezogenen Daten Verstorbener.
>
> Erwägungsgrund 26 Satz 1 der Datenschutz-GVO [1, S 5]

Knopp [12] betont weiter, dass das Strafrecht auch solche Daten schütze, die in der Vergangenheit mit einem Verfahren verschlüsselt wurden, das nicht mehr dem aktuellen Stand der Technik entspricht.

> § 202a — Ausspähen von Daten
>
> (1) Wer unbefugt sich oder einem anderen Zugang zu Daten, die nicht für ihn bestimmt und die gegen unberechtigten Zugang besonders gesichert sind, unter Überwindung der Zugangssicherung verschafft, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
> (2) Daten im Sinne des Absatzes 1 sind nur solche, die elektronisch, magnetisch oder sonst nicht unmittelbar wahrnehmbar gespeichert sind oder übermittelt werden.
>
> § 202a StGB

Bei Daten, die langfristig gespeichert werden sollen, wird man auch antizipieren müssen, dass es nötig werden wird, das verwendete Verschlüsselungsverfahren einem sich ändernden Stand der Technik anzupassen. Das ist freilich nur dann wirkungsvoll, wenn der Storage-Provider sicherstellen kann, dass die alten – nach neuem Verständnis unsicher verschlüsselten – Daten zeitnah zuverlässig gelöscht werden.

Knopp [12] argumentiert, und scheint dabei durchaus dem Zeitgeist zu entsprechen, dass Verschlüsselung generell nicht als eine Maßnahme zur Anonymisierung angesehen werden sollte, sondern als „Schutzmaßnahme und Zugriffsschutz“. Prima facia ist dem wenig entgegen zu halten. In seiner Konsequenz überseugt sein Fazit jedoch nicht. Es ist ein Fakt, dass die Sicherheit der modernen Welt maßgeblich auf dem Vertrauen in Kryptographie beruht. Sei es bei der Übertragung von Daten über das Internet oder der Zugriffskontrolle in vernetzten Systemen. Weshalb man ausgerechnet im Verhältnis zum Storage-Provider einmalig hohe Anforderungen stellen sollte, ist nicht ersichtlich. Allenfalls zeichnet sich diese Konstellation dadurch aus, dass der Storage-Provider die Daten lange über das vertraglich vereinbarte Ende hinaus speichern könnte, sodass die verwendete Verschlüsselung nicht nur den Angriffsmöglichkeiten der Gegenwart, sondern auch jenen der Zukunft ausgesetzt ist. Es ist jedoch nicht ersichtlich, weshalb ein (selbst unredlicher)

wirtschaftlich handelnder Akteur einen erheblichen Anteil seiner Speicherkapazität dafür verschwenden sollte, Daten zu speichern, die er aktuell nicht lesen kann, nur um in Zukunft die sehr hypothetische Möglichkeit zu haben, eventuell in der Lage zu sein, unter Einsatz massiver Rechenkapazitäten (und damit Kosten) und unter Begehung einer Straftat Informationen von fragwürdigem Wert zu erlangen. Karg [10, S 526] wendet zwar zurecht ein, dass staatliche Stellen auch solch wirtschaftlich unsinnige Unterfangen in Angriff nehmen könnten, wenn sie der Meinung sind, dass es der nationalen Sicherheit dienlich sei, allerdings hält unter dieser Annahme den Staat auch niemand davon ab, Telekommunikationsdaten für eine mögliche Auswertung in der Zukunft zu speichern, sodass sich eine Sonderbehandlung für bei Dritten gespeicherte Daten daraus nicht rechtfertigen lässt. Im Gegenteil sieht das geltende deutsche Recht vor, dass die Behörden zur Gefahrenabwehr standardmäßig die Telekommunikation abgreifen drürfen, während sie sich zu den beim Storage-Provider gespeicherten Daten erst Zugang verschaffen müssten. Schließlich wäre der Gesetzgeber vielleicht besser beraten, die Arbeit der Nachrichtendienste in einer Art und Weise zu regeln, die einer Demokratie würdig ist, anstatt der Wirtschaft aufzubürden, sich gegen irrationale Angreifer zu wappnen.

Generell scheint es mir, als ob die juristische Literatur ein meiner Meinung nach in diesem Maß unangebrachtes Misstrauen gegenüber der Sicherheit von Kryptographie hegte. Zwar basiert die Sicherheit keines der aktuell verfügbaren kryptographischen Verfahren auf einer ohne Annahmen beweisbaren mathematischen Grundlage, ähnliches ließe sich jedoch auch für Risiken in anderen Lebensbereichen sagen. Die wichtigsten populären Verschlüsselungsverfahren sind (zumindest für praktische Belange) seit Jahrzehnten strukturell ungebrochen. Die stetig (und aktuell immer noch exponentiell) zunehmende Rechenleistung stellt zwar eine Herausforderung dar, diese ist jedoch mit vertretbarem Risiko kalkulierbar und durch geeignete Wahl der Sicherheitsparameter handhabbar. Es stimmt, dass in den 1980er Jahren mittels des damals üblichen *Data Encryption Standards* (DES) verschlüsselte Daten heute praktisch schutzlos sind. Dass dies so kommen würde, war allerdings auch damals bereits bekannt. Für Daten, deren langfristige Vertraulichkeit bereits damals als wichtig erkannt worden war, konnte dem durch den Einsatz von Double- oder Triple-DES wirkungsvoll begegnet werden. Mittels Triple-DES verschlüsselte Daten sind auch nach heutigem Ermessen und auf absehbare Zukunft sicher, wenngleich theoretische Angriffe existieren, die es unratsam machen, DES weiterhin einzusetzen.

Gänzlich hahnebüchen ist es jedoch, auf eigene Faust ein vermeintlich sicheres Krypto-Verfahren entwerfen zu wollen, wie Pordesch und Steidle [15, § 4.1] es versuchen, und dabei bei einer Lösung landen, die nicht nur unprakti-

kabel, sondern auch denkbar unsicher ist. Ihr Vorschlag, ein One-Time-Pad zu verwenden, und die beiden Teile bei unterschiedlichen Storage-Anbietern abzulegen tauscht das sehr hypothetische Risiko einer radikalen Erkenntnis auf dem Gebiet der Kryptoanalyse mit dem sehr realen Risiko, dass zwei Storage-Provider miteinander kommunizieren, oder eines der beiden Unternehmen das andere aufkauft.

Das Risiko, dass Verschlüsselung durch fehlerhafte Implementierung oder Fehlverhalten der Benutzer unsicher wird, ist ungleich höher, als jenes, dass bahnbrechende mathematische Erkenntnisse oder Quantensprünge in der Entwicklung der Rechenleistung bisherige Annahmen auf den Kopf stellen. Dieses Risiko menschlicher Fehler ist jedoch allem menschlichen Handeln zueigen, und bestünde auch, wenn andere Verfahren als Kryptographie zum Einsatz kämen. Der Kryptographie eigen ist jedoch, dass in keiner anderen wissenschaftlichen Disziplin Staaten aktiv daran arbeiten, die öffentliche Forschung zu unterminieren und Unwissen und unsichere Verhaltensweisen zu propagieren. Wie der Gesetzgeber diesem untragbaren Zustand abhelfen könnte, liegt indes auf der Hand, und ist eine rein politische Frage.

Mein größter Kritikpunkt an der gegenwärtigen einschlägigen juristischen Literatur ist jedoch, dass sie stets nur den *einfachen* Fall betrachtet, dass ein gesamter Datensatz am Stück verschlüsselt wird, und als Gefährdung nur direkte Angriffe auf die Verschlüsselung in Betracht zieht. In der Praxis ist dieses Szenario – abgesehen von wenigen Ausnahmen wie etwa Backups – jedoch nicht relevant. Niemand will Terabytes an Daten verschlüsselt bei einem Storage-Anbieter ablegen, nur um jedes Mal, wenn er darauf zugreifen oder sie verändern möchte, den kompletten Datenbestand herunter und gegebenenfalls wieder erneut hochladen zu müssen. Stattdessen wird eine Lösung benötigt, die es erlaubt, zumindest in begrenztem Umfang in den verschlüsselten Daten zu suchen, und Aktualisierungen vorzunehmen. Sollen diese Aktionen unterstützt werden, eröffnen sich aber ganz neue Angriffsmöglichkeiten, die deutlich schwerer handzuhaben sind. Speichert ein Anwaltsbüro etwa die Akten jedes Mandaten in einer eigenen verschlüsselten Datei, so ist anhand der Dateien, auf die sie am Verhandlungstag zugreift, unschwer festzustellen, welche Datei zu welchem Mandanten gehört. Greift sie später auf dieselbe Datei wieder schreibend zu, liegt die Vermutung nahe, dass dem Mandanten ein neues Verfahren bevorsteht. Diese Erkenntnis definitiv personenbezogener Information konnte gewonnen werden, ohne dass auch nur eine einzige Datei entschlüsselt werden musste. Einen solchen Angriff bezeichnet man als *Seitenkanalangriff*, weil die Information auf einem dafür „nicht vorgesehenen“ Weg über Meta-Daten erlangt wurde. Da sich der Seitenkanal eben gerade dadurch auszeichnet, dass man ihn üblicherweise nicht als Teil des zu schützenden Systems erkennen wird, ist es besonders schwierig, solche An-

griffe vorherzusehen und zu unterbinden.

Brauchbare theoretische Sicherheitskonzepte für solche Konstellationen zu entwickeln, ist eine noch lange nicht abgeschlossene wissenschaftliche Herausforderung.

6.3 Homomorphe Kryptographie

Unter *homomorpher Kryptographie* versteht man, sehr grob gesagt, verschlüsselte Daten, mit denen man rechnen kann. Der Vorteil liegt auf der Hand: Anstatt die verschlüsselten Daten herunterladen, lokal mit ihnen zu rechnen, und die Ergebnisse wieder verschlüsseln und hochladen zu müssen, oder zu unsicheren Lösungen, wie am Ende des vorangehenden Abschnitts skizziert, greifen zu müssen, übergiebt man der Partei, die die Daten speichert, ein Programm, und bittet sie, es auf den Daten auszuführen. Alle Zwischenergebnisse und das Endresultat sind selbst verschlüsselte Daten, die für denjenigen, der das Programm ausführt, keine Bedeutung haben.

In § 5.2 wurde bereits eine Arbeit vorgestellt [14], mithilfe derer – wenn auch unter theoretischen Abstrichen – eine verschlüsselte relationale Datenbank implementiert werden kann. Dieses Schema könnte durchaus als Beispiel für homomorphe Kryptographie angesehen werden.

Idealerweise wünscht man sich jedoch, dass man auf den Daten beliebige Programme – nicht lediglich Datenbanktransaktionen – ausführen kann. Dazu sind zumindest eine Addition und Multiplikation erforderlich. Seit 2009 steht ein solches Verfahren – das lange Zeit für unmöglich gehalten wurde – tatsächlich auch zur Verfügung [9]. Leider ist es sehr aufwändig und damit zwar eine spannende theoretische Erkenntnis, jedoch mit bisweilen fragwürdigem praktischen Wert. Wie Müller-Quade, Huber und Nilges [14, S 532] betonen, macht es wirtschaftlich keinen Sinn, ein Verfahren zu benutzen, das Berechnungen sicher zu einem Dienstleister auslagert, wenn es günstiger und mindestens ebenso sicher wäre, eigene Hardware anzuschaffen.

Beispiele, in denen der hohe Rechenaufwand keine Rolle spielt, sind reichlich konstruiert. Etwa könnte man sich vorstellen, dass ein Stromliefervertrag einen variablen Strompreis vorsieht. Der Kunde möchte jedoch nicht offenbaren, wann er wie viel Strom verbraucht hat, und der Lieferant möchte seine Kalkulation nicht offenlegen müssen. Hier könnten beide Parteien ihre Geheimnisse verschlüsseln, das Skalarprodukt auf den verschlüsselten Daten berechnen, und das Ergebnis anschließend wieder von der anderen Partei entschlüsseln lassen. In der Praxis werden wohl beide Vertragsparteien nicht sonderlich erbaut darüber sein, nicht nachvollziehen zu können, wie der Endpreis zustande kam, den sie nun bezahlen beziehungsweise bekommen sollen.

7 Fallstudien

Zum Abschluss wollen wir noch einige konkrete Beispiele betrachten, und aus Sicht des Datenschutzes bewerten.

7.1 Dynamische IP-Adressen

Seit Jahren läuft in Deutschland ein Rechtsstreit darüber, ob eine dynamische IP-Adresse ein personenbezogenes Datum sei. Das LG Berlin ist – der Theorie der relativen Bestimmbarkeit folgend – davon ausgegangen, dass eine dynamische IP-Adresse zusammen mit einem Zeitstempel für sich genommen kein personenbezogenes Datum darstelle, und daher von einem Webseiten-Betreiber ohne weitere Auflagen gespeichert werden könne, um mögliche Angriffe auf seine Seite erkennen und abwehren sowie gegebenenfalls die Angreifer strafrechtlich verfolgen zu können [5]. An der Paradoxie, dass das Ziel der Verwendung im Zuge einer möglichen Strafanzeige offenkundig im Widerspruch zur Annahme steht, dass die betroffene Person nicht bestimmbar sei, nimmt das Gericht mit schleierhaften Argumenten keinen Anstoß. Auch wenn der Seitenbetreiber die Zuordnung nicht selbst vornimmt, so handelt es sich nicht leugenbar um eine Identifizierung, die ein Dritter (wenn auch nur unter bestimmten gesetzlich vorgegebenen Voraussetzungen) vernünftigerweise legal durchführen kann und wird. Der BGH hat das Revisionsverfahren hierzu ausgesetzt, und die Frage dem EuGH zur Vorabentscheidung vorgelegt [2]. Der EuGH hat in dieser Sache noch nicht entschieden, jedoch plädiert der Generalanwalt in seinem Schlussantrag in nachvollziehbarer und zu unterstützender Weise dafür, das Vorliegen eines personenbezogenen Datums zu bejahen [3].

Knopp [13, S 528] argumentiert ebenfalls, dass eine IP-Adresse nicht einmal ein Pseudonym sei. Unter anderem, da „es sich um kein Verfahren handle, das auch nur die Erschwerung der Zuordnung zum Ziel habe“. Wenngleich sich mir nicht erschließt, weshalb auf das Ziel und nicht die Wirkung eines Verfahrens abzustellen sein sollte, halte ich diese Einschätzung im Ergebnis für zutreffend. Die bisherige Argumentation aller beteiligten Gerichte scheint mir am Kern der Sache vorbeizugehen. Bestimmbarkeit im Sinne des BDSG kann nicht ausschließlich dahingehend ausgelegt werden, den bürgerlichen Namen und womöglich die Anschrift einer natürlichen Person zu ermitteln. Vielmehr ist die Person bereits dadurch bestimmt, dass sie anhand ihrer IP-Adresse wiedererkennbar und in ihrem Online-Verhalten verfolgbar ist. Dadurch kann ihr Verhalten unmittelbar beobachtet und beeinflusst werden. Der Schutzzweck des Datenschutzrechts würde verfehlt, wenn man den Schutzbereich alleine deshalb nicht als eröffnet ansehen wollte, weil ihr bürgerlicher Name – der letztendlich wie die IP-Adresse auch nur ein anderes (wenn

auch langlebigeres) Identifikationsmerkmal darstellt – nicht bekannt ist. Die Datenschutz-GVO scheint diese Ansicht zu stützen.

> [Im Sinne dieser Verordnung bezeichnet der Ausdruck] „personenbezogene Daten“ alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann;
>
> Art 4, Ziff 1, Datenschutz-GVO, [1, S 33]

allerdings fehlt es an einer expliziten Klarstellung in Erwägungsgrund 30.

7.2 Elektronische Kennzeichenerfassung

Die Bayrische Polizei setzt zur Fahndung nach zur solchen ausgeschriebenen Fahrzeugen automatisierte sowohl stationäre als auch mobile Systeme ein, die die Kennzeichen aller vorbeifahrenden Fahrzeuge unbemerkt elektronisch erfassen, und mit den Fahndungslisten abgleichen. Sofern dabei ein Treffer festgestellt wird, wird der Vorfall zur weiteren Bearbeitung gespeichert. Anderenfalls wird das Kennzeichen sofort wieder aus dem Speicher des Systems gelöscht. Allerdings wird in jedem Fall ein MD5-Hash des Kennzeichens in einer Protokolldatei festgehalten. MD5 ist eine (inzwischen definitiv nicht mehr empfohlene) kryptographische Hash-Funktion. Ein Bayer klagte gegen das System, weil er sich in seinem Grundrecht auf informationelle Selbstbestimmung verletzt sah; die Klage wurde jedoch in allen Instanzen abgewiesen [6, 7, 8]. Während die Urteile in Bezug auf den Abgleich mit den Fahndungslisten nicht zu beanstanden sind, verkennt der Bayrische VGH [7, Rn 78] die Unwirksamkeit der Pseudonymisierung, die durch das Speichern der MD5-Hashes vorgenommen wird. Zwar ist es trotz der (inzwischen sowohl theoretisch wie auch praktisch) gebrochenen Kolloisionsresistenz von MD5 nach wie vor nicht praktikabel, MD5-Hashes in großem Stil zu invertieren, allerdings gibt es in diesem Fall eine deutlich schnellere Methode, um ans Ziel zu kommen. Während es

$$2^{128} = 340\,282\,366\,920\,938\,463\,463\,374\,607\,431\,768\,211\,456$$

(also *sehr viele*) mögliche MD5-Hashes gibt, ist die Zahl der weltweit jemals vergebenen KFZ-Kennzeichen offensichtlich *deutlich* geringer. Die Wahrscheinlichkeit, dass zwei Kennzeichen auf denselben Wert hashen, ist also prak-

tisch vernachlässigbar. Um die Protokolldatei zu de-pseudonymisieren, besorge man sich also eine Liste der Kennzeichen aller in Deutschland (oder auch Europa) zugelassenen Fahrzeuge (was für eine Sicherheitsbehörde, die mit der Fahrzeugfahndung beauftragt ist, sicherlich eine leichte Übung ist), berechne für jedes davon den MD5-Hash und speichere die Ergebnisse in einer Tabelle. Dieser Vorgang dauert auf einem handelsüblichen Computer (auch den im Jahr 2009 erhältlichen Modellen) maximal wenige Minuten. Sodann hat man die Möglichkeit, mithilfe der Tabelle beliebige Protokolldateien in Sekundenbruchteilen zu de-pseudonymisieren.

8 Fazit

Um die Bedeutung von Anonymität, Pseudonymität und die Rolle, die Kryptographie dabei spielen kann, wird in der juristischen Literatur aktuell noch heftig gerungen. Deutschland stellt im europäischen Vergleich einen Sonderfall dar, indem hierzulande noch das Konzept der relativen Bestimmbarkeit verfolgt wird, das sich in dieser Form im Rechtsverständnis anderer europäischer Staaten nicht wiederfindet.

Das Verständnis, dass Bestimmbarkeit einer Person die Kenntnis eines bürgerlichen Namens und einer ladungsfähigen Anschrift voraussetzt, ist meines Erachtens ein Anachronismus, der überwunden werden sollte. Der moderne Mensch ist durch seine virtuelle Identität nicht minder bestimmt als durch seine melderechtliche.

Das Verhältnis der juristischen Literatur zur Kryptographie ist durch ein erhebliches Misstrauen geprägt, das sich jedoch vor allem auf den meines Erachtens am wenigsten kritischen Bereich konzentriert, während die Gefahr von Seitenkanalangriffen nicht erkannt und thematisiert wird.

Grundsätzlich sind Pseudonymisierung, Anonymisierung und Verschlüsselung jedenfalls sehr wirkungsvolle Maßnahmen, um die Menge an verwendeten personenbezogenen Informationen klein zu halten, und diejenigen Daten, die verwendet werden, vor Missbrauch zu schützen, sodass alle Beteiligten aus eigenem rechtlichen und wirtschaftlichen Interesse, wie auch aus ethischen Gründen wo immer möglich davon Gebrauch machen sollten. Insbesondere die kommende Datenschutz-GVO lässt auch keinen Zweifel daran, dass der Gesetzgeber von den wirtschaftlichen Akteuren erwartet, dass sie von sich aus ernsthaft daran arbeiten, in jeder Stufe der Datenverarbeitung die Menge an personenbezogener Information so gering wie möglich zu halten, und dazu ohne eigene „Belohnung“ entsprechende Verfahren einsetzen.

Die konkrete Implementierung erfordert allerdings einiges an Überlegung, um tatsächlich ein sicheres System zu schaffen. Dabei muss der Stand der

Technik selbstverständlich berücksichtigt werden, unter dem in diesem Fall fortschrittliche aber etablierte Verfahren verstanden werden sollten, sodass auch eine gewisse Zukunftssicherheit gegeben ist. Umso mehr bei Daten, die einem Dritten zur Speicherung für eine längere Zeit übergeben werden sollen. Nicht-Kryptographen ist dringend davon abzuraten, eigene Verfahren ersinnen zu wollen.

Karlsruhe am 23. Juni 2016, Moritz Klammler

Literatur

[1] Amtsblatt der Europäischen Union (DE) vom 4. Mai 2016, L 119.

[2] BGH, EuGH-Vorlage vom 28. Oktober 2014 – VI ZR 135/13.

[3] EuGH, Schlussanträge des Generalanwalts Campos Sánchez-Bordona vom 12. Mai 2016 in der Rechtssache C-582/14, Celex-Nr. 62014CC0582.

[4] Amtsblatt der Europäischen Gemeinschaften (DE) vom 23. November 1995, L 281.

[5] LG Berlin, Urteil vom 31. Januar 2013 – 57 S 87/08.

[6] VG München, Urteil vom 23. September 2009 – M 7 K 08.3052.

[7] Bayerischer Verwaltungsgerichtshof, Urteil vom 17. Dezember 2012 – 10 BV 09.2641.

[8] BVerwG, Urteil vom 22. Oktober 2014 – 6 C 7/13.

[9] Craig Gentry. „A Fully Homomorphic Encryption Scheme". Diss. Stanford, CA, USA, 2009. ISBN: 978-1-109-44450-6.

[10] Moritz Karg. „Anonymität, Pseudonyme und Personenbezug revisited?" In: *Datenschutz und Datensicherheit* 39.8 (2015), S. 520–526. ISSN: 1862-2607. DOI: 10.1007/s11623-015-0463-z.

[11] Jonathan Katz und Yehuda Lindell. *Introduction to Modern Cryptography.* Champman & Hall/CRC, 2008. ISBN: 978-1-58488-551-1.

[12] Michael Knopp. „Muss die Wirkung von Verschlüsselung neu gedacht werden?" In: *Datenschutz und Datensicherheit* 39.8 (2015), S. 542–545. ISSN: 1862-2607. DOI: 10.1007/s11623-015-0467-8.

[13] Michael Knopp. „Pseudonym – Grauzone zwischen Anonymisierung und Personenbezug". In: *Datenschutz und Datensicherheit* 39.8 (2015), S. 527–530. ISSN: 1862-2607. DOI: 10.1007/s11623-015-0464-y.

[14] Jörn Müller-Quade, Matthias Huber und Tobias Nilges. „Daten verschlüsselt speichern und verarbeiten in der Cloud". In: *Datenschutz und Datensicherheit* 39.8 (2015), S. 531–535. ISSN: 1862-2607. DOI: 10.1007/s11623-015-0465-x.

[15] Ulrich Pordesch und Roland Steidle. „Entfernen des Personenbezugs mittels Verschlüsselung durch Cloudnutzer". In: *Datenschutz und Datensicherheit* 39.8 (2015), S. 536–541. ISSN: 1862-2607. DOI: 10.1007/s11623-015-0466-9.

[16] Bruce Schneier. *Applied Cryptography.* 2. Aufl. John Wiley & Sons, 1996. ISBN: 0-471-12845-7.

[17] Oliver Stiemerling und Jürgen Hartung. „Datenschutz und Verschlüsselung – Wie belastbar ist Verschlüsselung gegenüber dem Anwendungsbereich des Datenschutzes?“ In: *Computer und Recht* 28.1 (2012), S. 60–68. ISSN: 2194-4172. DOI: 10.9785/ovs-cr-2012-60.