Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Werkzeuge zur Messung der datenschutzkonformen Einhaltung des Verarbeitungsstandorts in der Cloud

Bernd Jäger[1] Reiner Kraft[2] Annika Selzer[3] Ulrich Waldmann[4]

Abstract: Automatisiert nutzbare Datenquellen zur Bestimmung des aktuellen Verarbeitungsstandorts können Benutzer eines Cloud-Dienstes darin unterstützen, ihrer Kontrollpflicht nachzukommen und das Vertrauen in eine datenschutzkonforme Verarbeitung ihrer Daten zu stärken. Dazu können Standortmetriken definiert werden, für die geeignete Datenquellen gefunden und sinnvoll kombiniert werden müssen, um sich dem komplexen Datenschutzziel wie dem zulässigen Verarbeitungsort anzunähern. Diese Ausarbeitung beschreibt die Vorgehensweise und prototypische Umsetzung einer solchen Standortmetrik. Jede ausgewählte Messquelle bietet allerdings nur Indizien, mit denen sich bestenfalls die Zuverlässigkeit einer Gesamtaussage erhöhen lässt. Mit Hilfe eines Prototyps sollen aussagekräftige und zuverlässige Datenquellen und Indikatoren für die Standortmetrik identifiziert werden.

Keywords: Automatisierte Kontrolle, Datenschutz, Messquellen, Metrik, Verarbeitungsstandort.

1 Herausforderung und Ziel

Für Unternehmen, die personenbezogene Informationen zu Kunden, Mitarbeitern oder anderen Personen in der Cloud verarbeiten lassen wollen, ist sehr wichtig, dass diese Daten dort hinreichend geschützt sind.[5] Ein Grund hierfür ist, dass die Unternehmen als Cloud-Nutzer für die in ihrem Auftrag in der Cloud verarbeiteten Daten im Rahmen der sogenannten Auftragsdatenverarbeitung verantwortlich bleiben. Um dieser Verantwortlichkeit gerecht werden zu können, legt der Gesetzgeber dem Auftraggeber einer Auftragsdatenverarbeitung[6] als verantwortliche Stelle die Pflicht auf, den Auftragnehmer[7] regelmäßig datenschutzrechtlich zu kontrollieren.[8] Die Kontrollpflicht im Rahmen der Auftragsdatenverarbeitung betrifft die durch den Auftragnehmer zum Schutz der von ihm verarbeiteten personenbezogenen Daten getroffenen technischen und organisatorischen Maßnahmen gem. § 9 BDSG, zum Beispiel Maßnahmen zum Zutritts- und Zugangsschutz für Räume und/oder Systeme, auf welchen personenbezogene Daten verarbeitet werden, Maßnahmen

[1] Colt Technology Services, Herriotstraße 4, 60528 Frankfurt, bernd.jaeger@colt.net
[2] Fraunhofer SIT, Schloss Birlinghoven, 53754 Sankt Augustin, reiner.kraft@sit.fraunhofer.de
[3] Fraunhofer SIT, Rheinstraße 75, 64295 Darmstadt, annika.selzer@sit.fraunhofer.de
[4] Fraunhofer SIT, Rheinstraße 75, 64295 Darmstadt, ulrich.waldmann@sit.fraunhofer.de
[5] Der Beitrag entstand im Projekt VeriMetrix, gefördert vom BMBF im Programm IKT 2020 – Forschung für Innovationen, Förderkennzeichen: 16KIS0053K. Teile des Beitrags wurden zudem von der Europäischen Union aus dem Europäischen Fonds für regionale Entwicklung und vom Land Hessen kofinanziert.
[6] Im Rahmen der Nutzung von Cloud-Diensten ist dies der Cloud-Nutzer.
[7] Im Rahmen der Nutzung von Cloud-Diensten ist dies der Cloud-Anbieter.
[8] Vgl. § 11 Abs. 2 Satz 2 und 4 Bundesdatenschutzgesetz – kurz: BDSG.

zur Sicherstellung der Verfügbarkeit personenbezogener Daten und Maßnahmen zum Eingabeschutz, die nachträglich überprüfbar machen, ob und von wem personenbezogene Daten in Datenverarbeitungssysteme eingegeben, verändert oder entfernt wurden. Eine solche Kontrolle ist jedoch im Cloud-Umfeld aufgrund des verteilten Rechnens und der oft großen geographischen Entfernungen problematisch. [Bo12], [Se13]

Ein Lösungsansatz besteht in der Anwendung von automatisch überprüfbaren Datenschutzmetriken, also Kennzahlen zur Beurteilung datenschutzrelevanter Eigenschaften auf Basis von vertrauenswürdigen Messdaten, mit welchen sich der Umsetzungsgrad von Datenschutzmaßnahmen des Cloud-Anbieters kontinuierlich anhand von interpretierbaren Indikatoren überprüfen lässt. Solche Überprüfungen könnten sowohl den Prüfumfang der technischen und organisatorischen Maßnahmen gem. § 9 BDSG als auch weitere datenschutzrechtlich relevanten Eigenschaften einschließen. In einer Umfrage bei über 200 kleinen und mittleren Unternehmen aus dem Gesundheitswesen, der Rechtsberatung und dem Finanzsektor[9] erwiesen sich neben der Kontrolle der technischen und organisatorischen Maßnahmen – hier insbesondere die Kontrolle des Zugriffsschutzes einschließlich der Verschlüsselung personenbezogener Daten sowie die Datentrennung bzw. Mandantentrennung – u. a. die Löschung bzw. die Sperrung[10] von Daten nach Beendigung des Vertrages mit dem Cloud-Anbieter sowie die Kontrolle der zulässigen Verarbeitungsstandorte als wichtige Inhalte für automatisierte Datenschutzkontrollen im Cloud-Umfeld.

Aus datenschutzrechtlicher Sicht ist – neben der Kontrolle der technischen und organisatorischen Maßnahmen – die Kontrolle des Verarbeitungsstandorts besonders bedeutsam, da hierfür zum Teil strenge Vorgaben existieren. Eine Verarbeitung personenbezogener Daten durch einen Auftragsdatenverarbeiter ist innerhalb des Europäischen Wirtschaftsraumes, kurz: EWR, [11] datenschutzrechtlich privilegiert und nach Abschluss eines Auftragsdatenverarbeitungsvertrages und der Durchführung regelmäßiger Kontrollen in der Regel zulässig. Die Zulässigkeit der Verarbeitung personenbezogener Daten außerhalb des EWR muss hingegen anhand einer zweistufigen Prüfung bewertet werden: Zunächst muss die Verarbeitung durch eine gesetzliche Erlaubnisnorm oder durch die Einwilligung aller Betroffenen legitimiert sein (erste Zulässigkeitsstufe), sodann muss – etwa durch umfangreiche Datenschutzverträge oder internationale Datenschutzabkommen – sichergestellt werden, dass die personenbezogenen Daten auch außerhalb des EWR ausreichend geschützt werden (zweite Zulässigkeitsstufe).[12] Diese Unterscheidung geht auf den Umstand zurück, dass personenbezogene Daten innerhalb des EWR einem hohen und weitestgehend einheitlichen Schutz unterliegen, der sich aus der Umsetzung der Europäischen

[9] Vgl. http://www-wordpress.sit.fraunhofer.de/verimetrix/wp-content/uploads/sites/11/2014/05/Verimetrix-Fragebogen.pdf

[10] Die Sperrung tritt an die Stelle einer Löschung, wenn einer Löschung z. B. gesetzliche Aufbewahrungsfristen entgegenstehen oder eine Löschung wegen der besonderen Art der Speicherung nicht oder nur mit unverhältnismäßig hohem Aufwand möglich ist. Vgl. §§ 20 Abs. 3, 35 Abs. 3 BDSG

[11] Der Europäische Wirtschaftsraum ist eine Freihandelszone zwischen der Europäischen Union, Island, Liechtenstein und Norwegen.

[12] Vgl. Düsseldorfer Kreis, Datenübermittlung in Drittstaaten erfordert Prüfung in zwei Stufenhttp://www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/DuesseldorferKreis/12092013DatenuebermittlungInDrittstaaten.pdf

Datenschutzrichtlinie in den einzelnen Staaten ergibt.[13] Im Gegensatz dazu unterscheiden sich die Datenschutzsysteme sogenannter Drittstaaten – also Staaten außerhalb des EWR – sehr stark. [Se14]

Aufgrund der durchgängigen Virtualisierung der IT-Systeme in den unter Umständen weltweit verteilten Rechenzentren der Cloud-Anbieter bedeutet die zweifelsfreie Identifikation des Verarbeitungsstandorts eine hohe Herausforderung für die Entwicklung automatisierter Datenschutzkontrollen. In diesem Beitrag wird ein Lösungsvorschlag für diese Aufgabe beschrieben. Schwerpunkte sind dabei die Darstellung der Messverfahren zur Gewinnung valider Standortinformationen sowie die Skizze eines Architekturmodells für die sichere Erhebung, Verarbeitung und Speicherung der Messdaten. Abschließend werden einige Aspekte beschrieben, die bei der Weiterentwicklung der dargestellten Verfahren zu einem umfassenden Werkzeug zur kontinuierlichen Bewertung der Datenschutzqualität von Cloud-Angeboten zu berücksichtigen sind.

2 Verwandte Arbeiten

Die AG Rechtsrahmen des Technologieprogramms Trusted Cloud hat sich intensiv mit dem Problem der Kontrollpflicht beim Cloud-Computing auseinandergesetzt und schlägt in ihrem rechtspolitischen Thesenpapier [Bo12] eine vereinheitlichte Testatlösung zur Erfüllung der Kontrollpflicht vor. In dem Papier wird die Meinung vertreten, datenschutzrechtliche Kontrollen beim Cloud-Computing aufgrund des benötigten technischen und rechtlichen Spezialwissens von speziell zertifizierten Auditoren durchführen zu lassen, die für die Auditierung auch haften sollen. Zudem wird vorgeschlagen, die Kontrolle auf Basis einheitlicher, europaweit geltender Prüfkriterien durchzuführen. In [Se13] und [KNW13] wird dieser Vorschlag um eine automatisierte, kontinuierliche Datenschutzkontrolle auf Basis sicherer Log-Daten erweitert. Hierdurch wird es möglich, Datenschutzvorfälle zeitnah zu erkennen und auf diese zu reagieren.

Die Entwicklung von Metriken zur kontinuierlichen und weitgehend automatisierten Überprüfung der Einhaltung von Datenschutzanforderungen, ist eine Aufgabe, die sich überwiegend auf einem theoretischen und praktischen Neuland bewegt. Vorhandene Ansätze für Datenschutzmetriken – etwa das Mix-Modell von David Chaum oder andere Anonymitätsmetriken – sind auf das Problem der Anonymisierung personenbezogener Daten fokussiert, haben also keine umfassende Sicht auf den Datenschutz und wurden darüber hinaus bislang nur in geringem Umfang in der Praxis angewendet. Daneben gibt es vereinzelt Versuche, Datenschutzkennzahlen für organisatorische Sachverhalte zu erproben, etwa zur Anzahl und Bearbeitungsdauer von Datenschutzanfragen und -vorfällen.[14] Eine Reihe an Vorschlägen gibt es auch zu Informationssicherheitskennzahlen (siehe dazu z. B. die Publikationen [IS09, CI10, Ja07, NI08, So11]). Insofern Fragestellungen der IT- und Informationssicherheit betroffen sind, können diese auch in Datenschutzmetriken

[13] In Deutschland wurde die Europäische Datenschutzrichtlinie u. a. durch entsprechende Anpassungen im BDSG umgesetzt.

[14] Beispiele für solche Kennzahlen finden sich in der Telematikinfrastruktur der elektronischen Gesundheitskarte. Siehe `http://www.gematik.de/`.

einfließen. Allgemein anwendbare Metriken zur Überprüfung und Bewertung von Datenschutzeigenschaften existieren allerdings noch nicht.

Zur datenschutzfreundlichen Überprüfung des Verarbeitungsstandorts können die Ereignisse und Operationen der betreffenden virtuellen Maschinen (VM) analysiert werden. Die VM-Standorte sind dazu beispielsweise mittels Authentisierung der Rechenzentren bestimmbar [Ma11], was allerdings die aktive Mitwirkung der Cloud-Anbieter notwendig macht. VM-Standorte können mit anderen Verfahren auch betreiberunabhängig bestimmen werden. Beispielsweise können die Nutzer von VMs mittels aktiver Messungen von Paketumlaufzeiten zu einem Netzwerkknoten und wieder zurück („Round Trip Time", RTT) die virtuellen Koordinaten von umgebenen Netzwerkknoten und damit die relativen geographischen Positionen und Positionsänderungen von Cloud-VMs bestimmen. [Ri11] Zur Analyse von Standortdaten können verschiedenen Messdaten zu sogenannten Fingerprints zusammengefasst und mit statistischen Tools (z. B. mittels Programmiersprache R) und maschinellen Lernverfahren [Fl12] analysiert werden. Dabei werden aktuell gemessene Fingerprints mit vorliegenden Referenzdaten bekannter Standorte verglichen, ohne dass alle Teilinformationen der Fingerprints direkt verstanden oder interpretiert werden müssten. Solche heuristisch ermittelten Fingerprints wurden bereits zur Identifizierung von VM-Standorten eingesetzt. [Ri09], vgl. auch [JSW15]

3 Metriken und Messpunkte

Zur Entwicklung aussagekräftiger und realistischer Datenschutzmetriken sind diese zunächst aus den datenschutzrechtlichen Anforderungen abzuleiten. Im Anschluss an diese Top-Down-Betrachtung ist bottom-up zu prüfen, welche Datenquellen zur Verfügung stehen, um diese Metriken zu befüllen. Darauf aufbauend sind abschließend die Berechnungsverfahren der Metriken zu beschreiben. [JSW15]

3.1 Top-Down-Ableitung der Metriken

Im ersten Schritt zur Entwicklung von Metriken geht es darum, die bestehenden rechtlichen Anforderungen zu identifizieren und zu diesen passende Metriken zu finden. [LH15] Rechtliche Anforderungen ergeben sich u. a. aus dem Bundesdatenschutzgesetz, wo beispielsweise geregelt ist, dass eine privilegierte Auftragsdatenverarbeitung, bei der keine gesetzliche Erlaubnisnorm oder die Einwilligung der Betroffenen benötigt wird, geographisch auf den EWR beschränkt ist.[15] Eine Übermittlung personenbezogener Daten an Cloud-Anbieter außerhalb des EWR unterliegt demgegenüber wie oben beschrieben weiteren Anforderungen.

Um die beschriebene rechtliche Anforderung zu konkretisieren, benötigt es einer Darstellung der Maßnahmen, mit denen sie erfüllt werden können. Im Falle der geographischen Beschränkung des Auftragsdatenverarbeitungsprivilegs wäre eine solche Maßnahme z. B.,

[15] Vgl. § 11 BDSG i. V. m. § 3 Abs. 8 Satz 3, Abs. 4 BDSG.

die Datenverarbeitung personenbezogener Daten außerhalb des EWR zu unterbinden. Aus diesen Maßnahmen ergibt sich wiederum das sogenannte Konstrukt, das eine abstrakte, nicht direkt messbare Beschreibung des Messziels – z. B. die Kenntnis aller Verarbeitungsstandorte – darstellt. [LH15] Unter Berücksichtigung dieses Ansatzes lässt sich z. B. die folgende Metrik definieren:

$$\text{Prozentzahl zulässiger Standorte (PzS)} = \frac{\text{Anzahl zulässiger Standorte}}{\text{Gesamtanzahl der Standorte}} \times 100$$

In die oben angeführte Standortmetrik fließen in doppelter Weise kundenspezifische Anforderungen ein: Zum einen bestimmen diese, wann ein Standort als zulässig gilt, zum anderen legen sie auch fest, wie genau die Standortbestimmung erfolgen muss, um Verletzungen der Datenschutzanforderungen eines Kunden mit relativ hoher Wahrscheinlichkeit erkennen zu können.

Übliche Kundenanforderungen können beispielsweise sein, dass sich Daten nur innerhalb Deutschlands, des EWR oder eines Landes, das nach Einschätzung der Europäischen Kommission ein vergleichbares Datenschutzniveau bietet, befinden dürfen. Wieder andere Kunden erlauben darüber hinaus auch eine Datenverarbeitung in so genannten Drittstaaten[16], solange ein vergleichbares Datenschutzniveau durch entsprechende Verträge oder internationale Abkommen[17] sichergestellt ist. Im Extremfall ist eine Verarbeitung bei einem externen Cloud-Anbieter – wenn überhaupt – nur in einem räumlich streng festgelegten und sowohl physisch als auch netztechnisch vom übrigen Teil des Rechenzentrums abgetrennten Bereich zulässig.

Diese Beispiele machen deutlich, dass die Verfahren zur Standortermittlung mehrere Stufen und Kategorien bedürfen, um den verschiedenen Arten an Standortanforderungen gerecht zu werden. Neben der regionalen Verortung ist auch der Betreiber des Rechenzentrums zu bestimmen, das als Ort der Datenverarbeitung ermittelt wurde. Da bei hohen Anforderungen jeglicher Standortwechsel ausgeschlossen sein kann, muss ferner ein solcher Wechsel mit hoher Wahrscheinlichkeit detektiert werden können.

3.2 Bottom-Up-Bestimmung der Datenquellen

In diesem Schritt geht es darum, in der Infrastruktur eines Cloud-Dienstes aussagekräftige Datenquellen und daraus abgeleitete Indikatoren und Messverfahren zu finden, anhand derer die Top-Down abgeleiteten Metriken und damit der Umsetzungsgrad der Datenschutzanforderungen berechnet werden können. Bei der Auswahl dieser Datenquellen und Messverfahren ist darauf zu achten, dass sie sich nicht nur funktional eignen und vertrauenswürdige Daten liefern, sondern auch tatsächlich einsetzbar sind – hierfür kann es technische, organisatorische und rechtliche Barrieren geben. Beispielsweise kann die Nutzung einer Datenquelle sich verbieten, weil sie datenschutzrechtlich bedenklich ist, oder aber

[16] Als Drittstaaten werden im Datenschutzrecht diejenigen Staaten bezeichnet, die nicht dem EWR angehören.
[17] Hierzu zählen u. a. die EU-Standardvertragsklauseln, Binding-Corporate-Rules und das Safe Harbor Abkommen.

an der mangelnden Akzeptanz eines Cloud-Anbieters scheitern, der Störungen in seiner Cloud-Infrastruktur befürchtet.

Das Bottom-Up-Vorgehen zur Identifikation von Indikatoren und Messverfahren und deren Zuordnung zu den top-down formulierten Metriken soll nachfolgend am Beispiel der oben genannten Metrik zum Standort der Datenverarbeitung veranschaulicht werden. Diese Metrik erfordert konsequenterweise eine kontinuierliche Messung des Verarbeitungsstandorts der Daten eines Cloud-Nutzers, um die Wahrscheinlichkeit zu erhöhen, selbst zeitweilige Verstöße gegen dessen Vorgaben zu entdecken. Dies bedeutet im Cloud-Umfeld, dass insbesondere der physische Standort des Hosts identifiziert werden muss, auf dem die virtuellen Maschinen (VM), welche die Nutzerdaten verarbeiten, gestartet wurden. Dieser „physische Standort" der VMs ist besonders interessant, weil mit dem heutigen Stand der Technik zwar Daten auf dem Transportweg oder im dem Moment, da sie auf der virtuellen Festplatte gespeichert werden, verschlüsselt werden können, zur Verarbeitung durch den virtuellen Prozessor der VM jedoch im Klartext vorliegen müssen. Am „physischen Standort" der VM besteht also ein erhöhtes Risiko, dass auf sensible Daten unberechtigt zugegriffen wird.

Da die für Cloud-Anwendungen charakteristische Virtualisierung eine eindeutige Identifizierung des physischen Standorts signifikant erschwert, ist eine Mehrzahl an Indikatoren und Messverfahren erforderlich, um diesen mit einer zumindest sehr hohen Wahrscheinlichkeit bestimmen zu können. Diese Vielzahl erhöht darüber hinaus auch die Manipulationssicherheit und Zuverlässigkeit der Messungen. Zum einen wird es so für einen Cloud-Anbieter aufwändiger und damit unattraktiv, die Datenbasis in ihrer Gesamtheit zu manipulieren, zum anderen kann so der Ausfall eines Messverfahrens leichter kompensiert werden.

Eine wesentliche Leitidee bei einem Großteil dieser Messverfahren ist es, zu prüfen, ob die ermittelten Informationen zur Umgebung der virtuellen Maschinen mit bekannten und in einer Datenbank eingetragenen Mustern übereinstimmen, die für einen bestimmten Standort charakteristisch sind. Diese Muster bilden einen sogenannten „Fingerprint" eines Standorts.[18] Um den „Fingerprint" mit der VM zu verknüpfen, werden die Messprogramme („Agenten") innerhalb der gleichen VM gestartet, auf der auch reguläre Anwendungen Nutzerdaten verarbeiten werden (Datenverarbeitungs-VM). Um Messungen, die nicht aus einer VM selbst heraus durchgeführt werden, mit dieser zu verknüpfen, kann eine externe Messinstanz z. B. eine eindeutige Agenten-ID abfragen (etwa durch einen Netzwerk-Trace von „außen" mit Übertragung der Agenten-ID). Eine Verknüpfung kann aber auch indirekt über geeignete Umgebungsparameter, die der VM und der externen Mess-Instanz bekannt sind, hergestellt werden.

Geeignete Umgebungsparameter zur Zusammenstellung von Fingerprints sind in verschiedenen Bereichen typischer Cloud-Plattformen zu finden und können innerhalb und außerhalb der Infrastruktur des Cloud-Anbieters angesiedelt sein. Es sind sowohl aktive Messungen mit Diagnosewerkzeugen (Security Scannern etc.) als auch passive Messungen

[18] Bei diesen handelt es sich allerdings nicht um kryptographische Hashwerte, selbst wenn sie eine ähnliche Funktion haben wie die Fingerprints, die für die Verifikation kryptographischer Schlüssel verwendet werden.

(z. B. von vorhandenen Datenströmen mit Hilfe eines Sniffers) möglich. Die Bereiche der virtuellen Maschinen und des VM-Managements mit der darunter liegenden Hardware bieten in ihren Implementierungsdetails hilfreiche Anhaltspunkte für eine Bestimmung des Verarbeitungsortes, insoweit der Cloud-Anbieter den Zugriff auf solche Daten zulässt.

Mögliche Messungen im Bereich der Infrastruktur eines Rechenzentrums betreffen auch die nicht-öffentlichen Schnittstellen und zusätzliche von außen nicht sichtbare Informationen, die aus gezielten Abfragen von u. a. Management-Schnittstellen, internen Switches und Netzwerkkomponenten gewonnen werden können. Mögliche Messdaten, die als Parameter in die Berechnung eines Fingerprints einfließen können, sind u. a. auch IP-Adressen, Paketlaufzeiten und Paketrouten sowie Software-Versionen von Gerätetreibern, Virtualisierungskomponenten oder deren Management-Schnittstellen (insoweit verfügbar). Beispielsweise kann in einigen Fällen aus einer gemessenen IP-Adresse per „Reverse DNS Lookup“ der Hostname einer Cloud-Anbieter-Netzwerkdienstinstanz ermittelt und als Indikator für die globale Region zur Metrikenentwicklung herangezogen werden. [JSW15]

Die Messdaten werden von einer externen Komponente ausgewertet und zu charakteristischen Fingerprints zusammengestellt mit dem Ziel, sie bestimmten Regionen und Standorten zuzuordnen und als Indikatoren für die Standortmetrik zu interpretieren. Ist der Standort eines Fingerprint mit hoher Wahrscheinlichkeit bestimmt, kann er in Form von Referenzdaten zur Standortbestimmung anderer Fingerprints zugelassen werden.

3.3 Beschreibung der Berechnungsverfahren

Das Zusammenwirken der verschiedenen Einzelindikatoren in einer Metrik wird mit Hilfe eines sogenannten analytischen Modells beschrieben. Dieses Modell liefert die Berechnungsvorschrift für die Metrik und drückt aus, welche Relevanz jedes einzelne Messverfahren für die Approximation eines betrachteten Konstrukts hat. Damit legt es auch fest, wie aussagekräftig die Metrik insgesamt ist. Nachfolgend wird ein solches Modell für die angeführte Standort-Metrik beschrieben, vgl. Beitrag [LH15] im Rahmen dieses Workshops.

Diese Metrik ergibt sich (logischerweise) als Funktion der Einzelbewertungen zu den insgesamt betrachteten Standorten. Für die Prüfung der Zulässigkeit eines einzelnen Standorts können die folgenden, mit Messinstrumenten innerhalb und außerhalb der virtuellen Infrastruktur des Cloud-Kunden (bzw. des Cloud-Anbieters) gewonnenen Indikatoren dienen:[19]

- Ind-1: mithilfe von DNS-Anfragen und der Abfrage von Standortdatenbanken ermittelte Standorte öffentlicher Server,
- Ind-2: mithilfe von Traceroute-Abfragen, die innerhalb der VM abgesetzt wurden, ermittelte Knotenpunkte und Zielstandorte von Datenpaketen,

[19] Die genannten Indikatoren und Messmethoden sollen das Verfahren veranschaulichen. Für die Standortbestimmung können zahlreiche weitere Indikatoren und Messverfahren sinnvoll sein.

- Ind-3: mithilfe von ping-Anfragen, die außerhalb der VM abgesetzt wurden, ermittelte Zielstandorte von Datenpaketen,
- Ind-4: mithilfe eines Netzwerk-Sniffers zusammengestellte Informationen über die Geräte in der Umgebung einer virtuellen Maschine,
- Ind-5: mithilfe spezialisierter Detection-Tools[20] zusammengestellte Informationen über die verwendete Virtualisierungssoftware.

Für die Bestimmung des Standorts einer virtuellen Maschine werden diese Indikatoren in zwei unterschiedliche Analyseverfahren gruppiert: Zum einen in solche, die sich auf einer direkten Messung des Standorts stützen (Ind-1, Ind-2 und Ind-3), zum anderen in solche, bei denen ein Standort indirekt mithilfe einer Fingerprint-Datenbank und statistischen Klassifikationsverfahren[21] bestimmt wird (Ind-4 und Ind-5). Alle Verfahren unterstützen zunächst einmal die Standortidentifikation. Für die Zulässigkeit des gefundenen Standorts ist dieser, wie oben beschrieben, zusätzlich noch an den kundenspezifischen Anforderungen zu spiegeln.

Je größer die Übereinstimmung der Ergebnisse aus Direktmessungen und Fingerprint-Vergleichen ist, desto eindeutiger ist auch die Standortbestimmung. Im Idealfall – positive Ergebnisse bei allen Einzelindikatoren – kann der Standort mit einer hohen Wahrscheinlichkeit korrekt bestimmt werden und hat auch die daraus abgeleitete Bewertung seiner Zulässigkeit einen hohen Grad an Plausibilität. Umgekehrt gilt dies auch, wenn alle Einzelindikatoren darauf hinweisen, dass der Standort einer Virtuellen Maschine nur unzureichend bestimmt werden kann.

Schwieriger fällt eine solche Festlegung immer dann, wenn die Einzelindikatoren unterschiedliche Ergebnisse liefern. Sind die Ergebnisse aus der Direktmessung negativ, kann mit sehr hoher Wahrscheinlichkeit ein korrekt bestimmter Standort vermutet werden. Ist hingegen der Fingerprint-Vergleich negativ, kann bei gleichzeitigem positiven Befund aus den Direktmessungen der Konfidenzwert[22] des verwendeten Klassifikationsverfahrens eine Entscheidung unterstützen.

Idealerweise sind die Berechnungsverfahren zur Standortbestimmung so offen gestaltet, dass Änderungen in den für einen Standort charakteristischen Angaben unmittelbar nachvollzogen werden können. Darüber hinaus sollten sie flexibel an neue Verfahren (z. B. einen besseren Algorithmus zum Fingerprint-Vergleich) angepasst werden können.

4 Architektur

Anwender des VeriMetrix-Systems sind insbesondere Cloud-Kunden, die kontinuierlich prüfen möchten, ob und inwieweit ihre Datenschutzanforderungen aktuell erfüllt werden.

[20] z. B. Scoopy

[21] Statistikwerkzeuge wie R oder SPSS bieten hierfür ein breites Spektrum an Methoden an.

[22] Dieser Wert, ein Maß für die Vertrauenswürdigkeit einer Aussage, wird üblicherweise von den statistischen Werkzeugen geliefert

Eine Architektur zur Messung des Verarbeitungsstandorts sollte dies berücksichtigen und ferner, dass der Cloud-Anbieter und der Cloud-Kunde in Bezug auf die Messungen unterschiedliche Interessen haben können. Der Betreiber hat nur geringes Interesse, unbearbeitete Messdaten aus seiner Cloud-Infrastruktur an externe auswertende Komponenten zu übertragen, und fürchtet deren unkontrollierte Verbreitung. Zudem könnte er daran interessiert sein, kritische Daten auszufiltern oder gar zu manipulieren. Der Cloud-Kunde möchten unverfälschte Informationen über die beauftragte Datenverarbeitung erhalten, aber nutzerspezifische Daten möglichst nicht mit anderen Kunden des Cloud-Betreibers teilen.

Abbildung 1 zeigt die wichtigsten Komponenten des VeriMetrix-Systems, links die Komponenten auf Seiten des Cloud-Betreibers. Dieser betreibt kundenspezifische VMs, in denen die eigentliche Datenverarbeitung stattfindet. Messdaten sogenannter Referenz-VMs (links unten), deren Standorte bekannt sind, dienen dem VeriMetrix-System zum Vergleich mit aktuellen Messdaten der Kunden-VMs. In der Mitte und rechts zeigt die Abbildung einige Komponenten des Kunden und ggf. auch eines unabhängigen VeriMetrix-Betreibers, der mit der Erfassung und Auswertung von Messdaten beauftragt werden könnte („Trust-as-a-Service“). Auf Seiten des Cloud-Betreibers sind in den Kunden-VMs und Referenz-VMs betreiberunabhängige Messmodule integriert, welche die in Kapitel 3.2 beschriebenen Umgebungsparameter messen. Die zentrale Komponente zur Berechnung der Metriken ist der Kollektor in der Mitte des Bildes. Er unterhält weitere Messmodule, die außerhalb der Infrastruktur des Cloud-Anbieters Standortparameter erfassen können und greift zur Berechnung der Metriken auf eine Referenzdatenbank zu, die regelmäßig aktualisiert und erweitert wird. Der Kollektor bietet eine Schnittstelle für Auditoren, die z. B. Ergebnisse aus vor-Ort-Prüfungen manuell eintragen können, und Schnittstelle für den Kunden und seine Anwendungen.

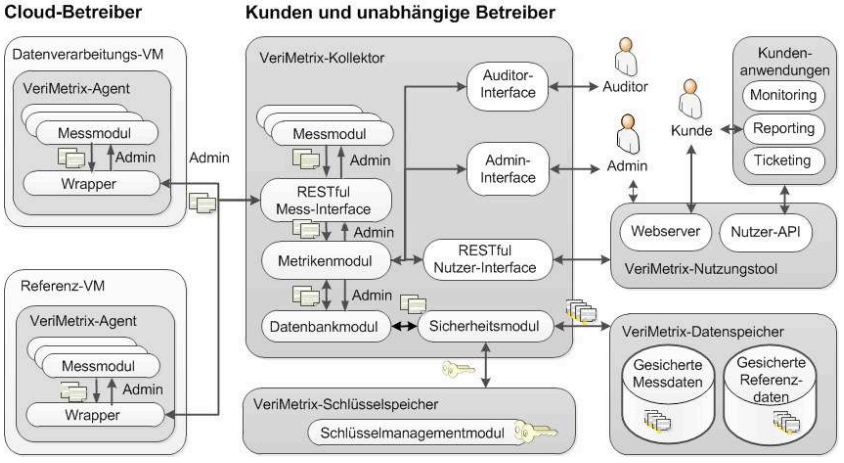

Abb. 1: VeriMetrix-Architektur

Ein Ziel dieses VeriMetrix-Systems ist, mutmaßliche oder tatsächliche Verstöße gegen Datenschutzanforderungen, etwa unzulässige Änderungen des Verarbeitungsstandorts, zuverlässig zu erkennen und ausreichend belegen zu können. Die Möglichkeit zur Konfiguration entsprechender Meldungen und zur Einsichtname in die einer Meldung zugrunde liegende Datenbasis sind daher wichtige Features der Nutzerschnittstelle.[23] Eine Administrationsschnittstelle auf Seiten des Auditors konfiguriert und steuert die VeriMetrix-Messmodule, beispielsweise den Umfang und die Häufigkeit von Messungen.

Die Frage, ob und wie die Messdaten geschützt werden müssen, betrifft nicht nur den Datenschutz, sondern auch die allgemeine IT-Sicherheit, da bekannt gewordene VM-Standortdaten auch als Grundlage für Angriffe dienen können.[24] [BS14] Wenn beispielsweise Umgebungsinformationen bestimmter VMs in die Hand von Angreifern fielen, könnten diese Angriffs-VMs mit höherer Erfolgsaussicht in derselben Umgebung starten. [Ri09] Die Architektur sieht daher Sicherheitskomponenten vor, welche die VM-spezifischen Messdaten in ein gesichertes Format überführen. In der potenziell unsicheren Umgebung des Cloud-Anbieters setzt das in Abbildung 1 gezeigte Sicherheitsmodul ein kryptographisches Verfahren ein, mit dem die Authentizität, Integrität und Vertraulichkeit der Messdaten geschützt werden. [KNW13]

Die Messdaten werden auf Basis eines von Schneier und Kelsey definierten Verfahrens gesichert, das Forward Integrity selbst in Umgebungen ermöglicht, denen nicht vorbehaltlos vertraut wird [SK99]. Unter dieser Voraussetzung können die Daten des VeriMetrix-Datenspeichers wiederum beim Cloud-Betreiber gespeichert werden, ohne dass der Cloud-Betreiber diese lesen oder verändern kann. Das Verfahren sieht vor, dass jeder neue Messwert mit dem vorigen Eintrag über Verschlüsselung, Hash-Werte und Prüfsummen verkettet und in eine aktuelle Messdatei geschrieben wird. Jedes Integritätsmodul handelt mit dem VeriMetrix-Schlüsselspeicher für jede neue Messdatendatei zwei symmetrische Startschlüssel aus. Das Sicherheitsmodul leitet daraus bzw. aus den Nachfolgeschlüsseln in jeder Runde für den aktuell zu erstellenden Messdateneintrag zwei symmetrische Arbeitsschlüssel für den zu erstellenden Eintrag ab: Durch Hashen der vorigen Arbeitsschlüssel zusammen mit den Zugriffsrechten für den neuen Eintrags werden daraus ein individueller Schlüssel zur Verschlüsselung der originären Messdaten und ein Schlüssel zur Berechnung einer Prüfsumme erzeugt. Nach jeder Verwendung und dem Erzeugen neuer Arbeitsschlüssel werden im Sicherheitsmodul die alten Schlüssel gelöscht. Die Startschlüssel stellen die Sicherheitsanker jeder gesicherten Datei dar und verlassen niemals den Schlüsselspeicher. Die Arbeitsschlüssel gesicherter Einträge sind nirgendwo gespeichert und können ausschließlich im Schlüsselspeicher aus den Startschlüsseln und den kodierten Zugriffsrechten rekonstruiert werden. Der Schlüsselspeicher sendet auf Anfrage diese Arbeitsschlüssel an autorisierte externe Instanzen (z. B. an einen Auditor), damit diese die Messdaten verifizieren, entschlüsseln und auswerten können.

[23] Aus wettbewerbsrechtlicher Sicht ist bei der zur Verfügung Stellung der Kontrollergebnisse darauf zu achten, Schmähkritik und nicht beweiskräftig nachweisbare Tatsachenbehauptungen zu unterlassen, keine offensichtlich unrichtigen Ergebnisse zu veröffentlichen sowie offenzulegen, wenn eine Bewertung ausschließlich auf den Angaben eines Cloud-Anbieters beruht. Vgl. analog das Ritter Sport-Urteil des Oberlandesgerichts München vom 09.09.2014.

[24] Vgl. BSI: IT-Grundschutz-Kataloge, Gefährdung G 4.90 „Ungewollte Preisgabe von Informationen durch Cloud Cartography“.

Der Betreiber des VeriMetrix-Systems könnte den Schlüsselspeicher wiederum in eine Cloud auslagern, sogar zum betreffenden Cloud-Betreiber selbst – beispielsweise in Form eines HSM-Dienstes, der nicht vom Cloud-Betreiber, sondern nur vom Kunden (oder von einem Treuhänder) konfigurierbar ist.[25]

5 Reaktion auf unrechtmäßige Standorte

Sollte das Ergebnis der Metrikenmessung Auffälligkeiten aufdecken, so muss der Cloud-Nutzer – als verantwortliche Stelle der Auftragsdatenverarbeitung – das Messergebnis juristisch interpretieren und dem konkreten Verstoß entsprechend reagieren. Hierfür sollte er zunächst den Schweregrad des festgestellten Datenschutzverstoßes anhand des vorliegenden Einzelfalls bewerten. Kriterien hierzu sind u. a. der Schutzbedarf personenbezogener Daten, der den Betroffenen drohende Schaden, die Umstände des Verstoßes und die Gesamtheit der getroffenen Sicherheitsmaßnahmen. [KSS15] Je nach Ergebnis der Schweregrad-Bewertung kommen unterschiedliche Reaktionen auf einen festgestellten Datenschutzverstoß in Betracht. Unter Berücksichtigung aller Kriterien des Einzelfalls kommen zunächst die Aufforderung zur unverzüglichen Beseitigung des Verstoßes sowie die Kontrolle seiner Beseitigung in Betracht. Bis zur Beseitigung des Verstoßes sollten nach Möglichkeit keine weiteren personenbezogenen Daten an den Cloud-Anbieter weitergegeben werden. Bei erheblichen Verstößen oder wenn der Cloud-Anbieter den Mangel nicht beseitigt, sollte der Cloud-Nutzer den Vertrag kündigen.

Zumindest bei Cloud-Speicherdiensten bestehen für den Cloud-Nutzer weitere Möglichkeiten, aktiv auf einen erheblichen Verstoß zu reagieren: Da das Herunterladen und das Löschen von zuvor gespeicherten Daten zum typischen Funktionsumfang dieser Art von Cloud-Diensten gehört, ist es dem Cloud-Nutzer möglich, die Daten beim Cloud-Anbieter ersatzlos zu löschen, auf eigene Systeme zu migrieren und danach beim Cloud-Anbieter zu löschen oder zu einem anderen Cloud-Anbieter zu migrieren.[26] [KSS15]

6 Ausblick

Wie eingangs erwähnt, sind Anwender, die personenbezogene Daten in der Cloud speichern oder verarbeiten, dazu verpflichtet, die Verfahrensabläufe und technischen Sicherheitsmaßnahmen ihres Cloud-Anbieters regelmäßig zu kontrollieren, um die Einhaltung geltender datenschutzrechtlicher Anforderungen sicherzustellen. In diesem Beitrag wurden Verfahren und eine Systemarchitektur vorgestellt, mit denen ein wichtiger Teilaspekt dieser Aufgabe automatisiert erfüllt werden kann, nämlich die Überprüfung der Zulässigkeit der Verarbeitungsstandorte der Daten. Grundsätzlich ist hierfür eine Vielzahl an Messverfahren möglich. Mithilfe praktischer Tests wird in der Folge im Rahmen des hier dargestellten Projekts VeriMetrix erprobt, welche Kombination der vorstehend skizzierten Verfahren in besonderem Maße geeignet ist, Verarbeitungsstandorte mit einem hohen Grad

[25] Derartige Funktionalität bietet auch der AWS CloudHSM-Service: `https://aws.amazon.com/de/cloudhsm`

[26] In allen drei Fällen empfehlen sich vertragliche Vereinbarungen mit dem Cloud-Anbieter über die Löschung von Backups.

an Zuverlässigkeit und Manipulationssicherheit zu bestimmen. In diesem Zusammenhang sind nicht nur die Schwellwerte zu bestimmen, ab denen Aussagen zu Standorten hinreichend plausibel sind, sondern ist auch auf eine möglichst breite und langfristige Anwendbarkeit der Messverfahren zu achten, also darauf, dass sie providerübergreifend funktionieren und flexibel an neuartige technische Gegebenheiten angepasst werden können.

Für eine umfassende Kontrolle und Bewertung der Datenschutzqualität eines Cloud-Dienstes sind weitere Indikatoren und Kennzahlen erforderlich, etwa solche zur Datentrennung, zum Zugriffsschutz oder auch zur Vollständigkeit der Datenlöschung nach Beendigung der Benutzung eines Dienstes. Auch hierfür lassen sich wichtige Kennzahlen automatisiert erheben, beispielsweise kann mithilfe von Wasserzeichen die Qualität der angewendeten Löschverfahren getestet werden. Automatisierte Kontrollen erlauben insbesondere eine zeitnahe Reaktion auf datenschutzrelevante Ereignisse bei einem Cloud-Anbieter.

Insbesondere für die Bewertung der organisatorischen und prozeduralen Sicherheitsvorkehrungen eines Cloud-Anbieters sind auch Indikatoren erforderlich, die entweder nur teilautomatisiert oder nur „händisch“ in unabhängigen Audits erhoben werden können. Automatisiert erhobene und auf anderen Wegen ermittelte Kennzahlen zusammengenommen können, bei geeigneter Vereinheitlichung und Standardisierung Grundlagen schaffen für eine höhere Transparenz und Vergleichbarkeit der Datenschutzqualität von Cloud-Anbieter und damit Hemmschwellen für die Benutzung von Cloud-Diensten beseitigen.

Literaturverzeichnis

[Bo12] Borges, G. et al.: Datenschutzrechtliche Lösungen für Cloud Computing. Kompetenzzentrum Trusted Cloud, Oct 2012.

[BS14] BSI: IT-Grundschutz-Kataloge: 14. Ergänzungslieferung Stand 2014. Bericht, Bundesamt für Sicherheit in der Informationstechnik, 2014.

[CI10] CIS: The CIS Security Metrics. The Center for Internet Security (Hrsg.), 2010.

[Fl12] Flach, P.: Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, 2012.

[IS09] ISO: ISO 27004:2009, Information Security Management – Measurement. International Organization of Standardization (Hrsg.), 2009.

[Ja07] Jaquith, A.: Security Metrics: Replacing Fear, Uncertainty, and Doubt. 2007.

[JSW15] Jäger, B.; Selzer, A.; Waldmann, U.: Die automatisierte Messung von Cloud-Verarbeitungsstandorten. Datenschutz und Datensicherheit – DuD, 01:26–30, 2015.

[KNW13] Kunz, T.; Niehues, P.; Waldmann, U.: Technische Unterstützung von Audits bei Cloud-Betreibern. Datenschutz und Datensicherheit – DuD, 37(8):521–525, 2013.

[KSS15] Kunz, T.; Selzer, A.; Steiner, S.: Konsequenzen festgestellter Verstöße bei der Auftragsdatenverarbeitungskontrolle. Datenschutz und Datensicherheit – DuD, 01:21–25, 2015.

[LH15] Luhn, S.; Hils, M.: Ein Rahmenwerk für Datenschutz-Metriken in der Cloud. In (Douglas Cunningham, Petra Hofstedt, Klaus Meer Ingo Schmitt, Hrsg.): INFORMATIK 2015, Lecture Notes in Informatics (LNI). Gesellschaft für Informatik, 2015.

[Ma11] Massonet, P.; Naqvi, S.; Ponsard, C.; Latanicki, J.; Rochwerger, B.; Villari, M.: A Monitoring and Audit Logging Architecture for Data Location Compliance in Federated Cloud Infrastructures. In: Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE International Symposium on. S. 1510–1517, may 2011.

[NI08] NIST: NIST 800-55: Performance Measurement Guide for Information Security. National Institute of Standards and Technology (Hrsg.), 2008.

[Ri09] Ristenpart, T.; Tromer, E.; Shacham, H.; Savage, S.: Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In: Proceedings of the 16th ACM conference on Computer and communications security. CCS '09, ACM, New York, NY, USA, S. 199–212, 2009.

[Ri11] Ries, T. et al.: Verification of data location in cloud networking. In: Fourth IEEE International Conference on Utility and Cloud Computing, S. 439–444, 2011.

[Se13] Selzer, A.: Die Kontrollpflicht nach §11 Abs. 2 Satz 4 BDSG im Zeitalter des Cloud Computing – Alternativen zur Vor-Ort-Kontrolle des Auftragnehmers durch den Auftraggeber. Datenschutz und Datensicherheit – DuD, 4:215–219, 2013.

[Se14] Selzer, A.: Datenschutz bei internationalen Cloud Computing Services. Datenschutz und Datensicherheit – DuD, 38(7):470–474, 2014.

[SK99] Schneier, B.; Kelsey, J.: Secure audit logs to support computer forensics. ACM Transactions on Information and System Security (TISSEC), 2(2):159–176, 1999.

[So11] Sowa, A.: Metriken, der Schlüssel zum erfolgreichen Security und Compliance Monitoring. 2011.