Anonymisierung und externe Speicherung in Cloud-Speichersystemen

Konrad Meier, Steffen Philipp

Albert-Ludwigs-Universität Freiburg
Professur für Kommunikationssysteme
Hermann-Herder-Str. 10
79104 Freiburg
konrad.meier@rz.uni-freiburg.de
steffenmatthiasphilipp@googlemail.com

Abstract: Cloud-Speichersysteme ermöglichen es, Daten kostengünstig zu speichern und ortsunabhängig auf die gespeicherten Daten zuzugreifen. Einer Nutzung solcher Systeme durch die öffentliche Verwaltung und durch private Unternehmen stehen bisher oft datenschutzrechtliche Regelungen entgegen. Die vorliegende Arbeit analysiert die juristischen Anforderungen bei der Verwendung von Cloud-Speichersystemen und beschreibt ein reversibles Anonymisierungsverfahren, das es ermöglicht, personenbezogene Daten in Cloud-Speichersystemen abzulegen. Die datenschutzrechtlichen Regelungen werden dabei nicht verletzt. Das Verfahren wurde in einem föderierten Speichersystem implementiert und evaluiert.

1 Einleitung

In den vergangenen Jahren haben es Cloud-basierte Speichersysteme ermöglicht, die stark wachsende Datenmenge zu speichern und flexibel zur Verarbeitung bereitzustellen. Dabei skalieren die Systeme sowohl bei den Datenmengen als auch bei der Anzahl der Nutzer. Um dem kurzfristigen Speicherbedarf von Forschungs- und Projektgruppen gerecht zu werden, ist es notwendig, projektbezogene Speichercontainer flexibel bereitstellen zu können. Das Speichern und Auslagern von Daten in externen Cloud-Speichersystemen ist somit interessant, um lokale, stark genutzte Ressourcen kosteneffizient zu entlasten. Bedarfsspitzen einzelner Nutzer können mit geringem Aufwand abgefangen werden, was zu Kosteneinsparungen führt. Wesentliche Hindernisse, die einer externen Speicherung in solchen Systemen bisher entgegenstehen, sind neben allgemeinen Sicherheitsbedenken vor allem datenschutzrechtliche Regelungen, die für personenbezogene Daten gelten. Die Vorteile dieser Systeme führen aber oft dazu, dass Benutzer datenschutzrechtliche Aspekte ignorieren oder bewusst eine Verletzung dieser in Kauf nehmen [1]. Diese Problematik zeigt sich auch im Hochschulumfeld.
Speziell für personenbezogene Daten müssen strenge datenschutzrechtliche Aspekte be-

[1] Beispiel: Verwenden von Dropbox für personenbezogene Daten.

achtet werden. Von zentraler Bedeutung ist hier die Regelung des § 4 Abs. 1 Bundesdatenschutzgesetz (BDSG), nach der jede Verarbeitung personenbezogener Daten, also auch deren Auslagerung in ein Cloud-System, entweder einer Erlaubnis durch Rechtsvorschrift oder einer wirksamen Einwilligung des Betroffenen bedarf. Da zum Zeitpunkt der Speicherung der Daten eine solche Einwilligung meist nicht gegeben ist und ein nachträgliches Einholen zu aufwändig ist, erscheint eine Auslagerung personenbezogener Daten in externe Cloud-Speichersysteme zunächst als nicht möglich. Für Daten ohne Personenbezug oder anonymisierte Daten trifft diese Einschränkung nicht zu. Die Praxis zeigt jedoch, dass eine Unterscheidung zwischen personenbezogenen Daten und Daten ohne Personenbezug oft nur schwer möglich ist. Gerade bei Benutzerdaten, beispielsweise Home-Laufwerke von Universitäsmitarbeitern und Studenten, ist eine Klassifikation der Daten meist unmöglich. In den meisten Fällen kann nur der Besitzer selbst angeben, ob personenbezogene Daten vorliegen.

Aufgrund dieser problematischen Ausgangslage beschreibt und evaluiert diese Arbeit ein auf Verschlüsselungs- und Fragmentierungstechniken basierendes Verfahren, mit dem Daten vor der externen Speicherung so umgewandelt werden, dass die externe Speicherung reversibel anonymisiert erfolgt. Somit fallen die extern gespeicherten Daten aufgrund der Anonymisierung nicht in den Anwendungsbereich geltender Datenschutzregelungen. Eine Auslagerung und Speicherung von Daten in externe Cloud-Speichersysteme wird somit auch für personenbezogene Daten ermöglicht.

Die Arbeit ist wie folgt strukturiert: Im Abschnitt 2 werden die Erkenntnisse verwandter Arbeiten untersucht. Im Abschnitt 3 erfolgt eine Analyse der rechtlichen Anforderungen, die bei der Speicherung in externen Systemen beachtet werden müssen. Im Abschnitt 4 wird die technische Umsetzung einer reversiblen Anonymisierung vorgestellt und im Abschnitt 5, im Rahmen eines föderierten Speichersystems, umgesetzt. Die Erkenntnisse der Arbeit werden im Abschnitt 6 zusammengefasst.

2 Verwandte Arbeiten

Cloud-Speichersysteme werden von vielen kommerziellen Anbietern als Dienste angeboten. Dies hat dazu geführt, dass zahlreiche Arbeiten zu den Themen vendor lock-in, Verfügbarkeit und Datenschutz veröffentlicht worden sind, um problematische Eigenschaften dieser Systeme zu beheben.

Eine Gegenüberstellung unterschiedlicher Arbeiten zum Thema verteilte Cloud-Speichersysteme sowie dem Cloud of Clouds-Ansatz ist die Arbeit von Slamanig und Hanser [SH12]. Dabei werden die betrachteten Arbeiten unter anderem hinsichtlich ihrer Eigenschaften Vertraulichkeit, Integrität, Verfügbarkeit sowie Zugangskontrolle analysiert. Die Autoren kommen zu dem Ergebnis, dass Datenschutzaspekte zwar prinzipiell als wichtige Eigenschaften angesehen werden, jedoch wird dieser Aspekt in den meisten Arbeiten nicht ausreichend berücksichtigt. Insbesondere der Aspekt der access privacy wird in keiner der Arbeiten beachtet.

In der Arbeit von Sheng et al [SMGL11] wird eine Bit-interleaving-Technik verwendet,

um Daten aufzuteilen und um sie dann anschließend bei Cloud-Anbietern zu speichern. Es wird argumentiert, dass alleine durch das Aufteilen der Daten ein ausreichender Datenschutz gewährleistet werden kann, da ein möglicher Angreifer keinen Zugriff auf alle Daten besitzt. Einen ähnlichen Ansatz verfolgt auch die Arbeit von Abu-Lidbeh et al. [ALPW10]. Hier werden RAID-Techniken verwendet, um das Problem des vendor lock-ins zu verhindern. Dabei werden die Daten so auf mehrere Anbieter aufgeteilt, dass auch der Ausfall eines Anbieters die Verfügbarkeit der Daten nicht gefährdet.

Der Cloud of Clouds-Ansatz wird auch in der Arbeit von Bessani et al. [BCQ+11] beschrieben. Auch hier werden die Daten auf unterschiedliche Cloud-Anbieter verteilt. Die Daten werden jedoch zusätzlich verschlüsselt. Das System soll die Verfügbarkeit, Vertraulichkeit und Verfügbarkeit der abgelegten Daten verbessern.

Ein dezentrales Speichersystem, basierend auf Erasure Codes, wird in der Arbeit von Yao et al. [YXH13] beschrieben. Das System soll dabei die Benutzerdaten über Verschlüsselung schützen. Sowohl die Daten als auch der Schlüssel werden dezentral gespeichert, um zu verhindern, dass ein nicht vertrauenswürdiger Speicherserver die Daten entwendet.

Alle zuvor aufgeführten Arbeiten bieten zwar einen Ansatz für ein dezentrales Speichersystem. Für den konkreten Anwendungsfall der personenbezogenen Daten ist jedoch eine juristische Analyse der Datenschutzproblematik für Deutschland notwendig. Nur so kann sichergestellt werden, dass eine technischen Lösungen auch gesetzlichen Regelungen genügt. Die vorliegende Arbeit bietet genau dies, indem sie, basierend auf einer juristischen Anforderungsanalyse, eine technischen Lösung präsentiert.

3 Rechtliche Anforderungen

Im Zusammenhang mit der Datenverarbeitung in Cloud-Systemen ist eine ganze Reihe datenschutzrechtlicher Vorschriften zu beachten. Insbesondere sind dies Gesetzesvorschriften im Telekommunikationsgesetz (TKG), im Telemediengesetz (TMG) und im Bundesdatenschutzgesetz (BDSG). Speziell bei der Auslagerung personenbezogener Daten in Cloud-Speichersysteme sind jedoch in der Regel nur die Vorschriften des BDSG anwendbar[2]. Im Folgenden wird die datenschutzrechtliche Problematik einer Auslagerung personenbezogener Daten in öffentliche Cloud-Speichersysteme anhand der Regelungen des BDSG erläutert.

3.1 Auftragsdatenverarbeitung

Das Auslagern personenbezogener Daten ist prinzipiell im Rahmen einer Auftragsdatenverarbeitung nach § 11 BDSG denkbar. Im Fall einer Auftragsdatenverarbeitung wird ein externer Auftragnehmer (Cloud-Anbieter) nicht als Dritter, sondern als verlängerter Arm

[2] Heidrich und Wegener [HW10], S. 803, (806)

der intern verantwortlichen Stelle behandelt, § 11 Abs. 1 S. 1 BDSG und § 3 Abs. 8 S. 3 BDSG. Das BDSG stellt für eine wirksame Auftragsdatenverarbeitung strenge Anforderungen hinsichtlich der Auswahl des Auftragnehmers und der vertraglich zu treffenden Regelungen, § 11 Abs. 2 S. 1 und S. 2 BDSG. Dadurch wird sichergestellt, dass der Auftragnehmer sich tatsächlich wie der verlängerte Arm der verantwortlichen Stelle verhält. Hieraus ergeben sich jedoch Pflichten für den Auftraggeber, die bei Cloud-Speicher-Anbietern nur schwer eingehalten werden können. So müssen Kontroll- und Dokumentationsvorschriften eingehalten werden. Bei Cloud-Speichersystemen besteht oft schon das Problem, den Speicherort zu bestimmen. Die Kontrolle von technischen und organisatorischen Maßnahmen zum Schutz der Daten stellt den Auftraggeber zusätzlich vor das Problem, dass er keinen Zugriff auf die Verarbeitungsprozesse und Verarbeitungsanlagen des Cloud-Anbieters hat.

Die mit der Auslagerung von Daten in öffentliche Cloud-Speichersysteme angestrebte Kostenersparnis ist nur realisierbar, wenn die Auslagerung nicht in den Anwendungsbereich der genannten Datenschutzvorschriften fällt. Der zusätzliche Aufwand für eine Auftragsdatenverarbeitung steht sonst im Widerspruch zur gewünschten Kostenersparnis. Damit die genannten Datenschutzvorschriften nicht zutreffen, muss sichergestellt werden, dass es sich bei den auszulagernden Daten nicht (mehr) um personenbezogene Daten handelt. Hierfür muss jedoch zunächst geklärt werden, wie personenbezogene Daten definiert werden.

3.2 Personenbezogene Daten

Zu den personenbezogenen Daten zählen Informationen, die einer natürlichen Person zugeordnet sind oder leicht einer Person zugeordnet werden können, wie beispielsweise eine Wohnadresse, eine Telefonnummer oder eine den Inhaber benennende E-Mail-Adresse [3]. Im Einzelnen ist oft schwer einzuschätzen, ob personenbezogene Daten vorliegen oder nicht. Auch der Begriff der personenbezogenen Daten selbst ist (wohl auch aufgrund der erheblichen Rechtsfolgen, die damit verknüpft sind) sehr umstritten. Das BDSG definiert personenbezogene Daten in § 3 Abs. 1 wie folgt:

> Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener).

Jedoch ist im BDSG nicht explizit geregelt, auf welche Stelle und auf welche Mittel es für das Kriterium „bestimmbar“ ankommen soll. Dazu gibt es zwei prinzipiell verschiedene Ansichten.
Nach der „relativen“ Ansicht sind für das Kriterium „bestimmbar“ nur die verantwortliche Stelle und deren Mittel relevant. „Verantwortliche Stelle“ ist jede Person oder Stelle, die personenbezogene Daten für sich selbst erhebt, verarbeitet oder nutzt oder dies durch andere im Auftrag vornehmen lässt, § 3 Abs. 7 BDSG [4].

[3] Siehe BDSG § 3 Abs. 1
[4] Siehe hierzu auch Gola u. a. [GKKS12], § 3 Rn. 10

Nach der „objektiven“ Ansicht sind für das Kriterium „bestimmbar“ auch Dritte und deren Mittel relevant. Ein Dritter in diesem Sinne ist außer dem Betroffenen jede Person oder Stelle außerhalb der verantwortlichen Stelle, § 3 Abs. 8 BDSG[5].

Wird jedoch der Text der europäischen Datenschutzrichtlinie 95/46/EG [EU-95] hinzugenommen, kommt man zu dem Schluss, dass weder die „relative“ noch die „objektive“ Ansicht mit der Richtlinie 95/46/EG und dem BDSG vereinbar ist. Die systematische und richtlinienkonforme Auslegung führt vielmehr zu dem Ergebnis, dass es für das Kritierium „bestimmbar“ im Sinne des § 3 Abs. 1 BDSG darauf ankommt, ob die verantwortliche Stelle eine Person direkt oder indirekt, mit Mitteln der verantwortlichen Stelle oder mit Mitteln eines Dritten identifizieren kann, wobei alle Mittel zu berücksichtigen sind, die vernünftigerweise eingesetzt werden könnten.

Welche Mittel „vernünftigerweise“ eingesetzt werden könnten, wird mittelbar durch § 3 Abs. 6 BDSG konkretisiert. Dort werden, als Gegenstück zu personenbezogenen Daten, anonymisierte Daten definiert.

3.3 Anonymisierte Daten

Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person. Bestimmbar ist eine Person, wenn die verantwortliche Stelle direkt oder indirekt, mit Mitteln der verantwortlichen Stelle oder mit Mitteln eines Dritten, eine Person identifizieren kann und der Aufwand für den Einsatz dieser Mittel im Hinblick auf Zeit, Kosten und Arbeitskraft nicht unverhältnismäßig ist, § 3 Abs. 6 BDSG.
Um Daten im Sinne des Datenschutzgesetzes zu anoymisieren, muss der Aufwand zum Identifizieren der Person unverhältnismäßig hoch sein. Was als unverhältnismäßig angesehen wird, ist im Gesetz nicht definiert. Eine Verschlüsselung personenbezogener Daten kann in diesem Zusammenhang nicht als Anonymisierung betrachtet werden[6], da dies hauptsächlich einen unverhältnismäßig hoher Rechenaufwand (Aufwand an Zeit und Kosten) bedeutet. Ein unverhältnismäßig hohen Aufwand an Arbeitskraft ist nicht gegeben. Auch ist es möglich, dass zu einem späteren Zeitpunkt der hohe Zeitaufwand drastisch reduziert werden kann, wenn Schwachstellen im verwendeten Verschlüsselungsverfahren entdeckt werden[7]. Ein Angreifer könnte somit Daten entwenden und zu einem späteren Zeitpunkt das Verschlüsselungsverfahren brechen, wenn dieses über technische Mittel möglich geworden ist.

Ein technischer Lösungsansatz zur automatischen Anonymisierung kann darin bestehen, die Verschlüsselung durch Maßnahmen zu ergänzen, die sicherstellen, dass für einen Angreifer ein unverhältnismäßiger Aufwand an Arbeitskraft erforderlich ist und der erforder-

[5] Siehe hierzu auch Weichert in Däubler u. a. [DWWK10], § 3 Rn. 13

[6] Zu diesem Ergebnis kommt auch das Beratungsgremium, das von der europäischen Kommission auf Grundlage der Richtlinie 95/46/EG eingesetzt wurde, in [AA07a], S. 15, allerdings ohne nähere Begründung. Differenzierter war die Einschätzung des Gremiums noch in [AA07b], S. 24, wo abhängig von den Umständen noch von einer Anonymisierung ausgegangen wurde.

[7] So auch Spies [Spi11], Ziffer 2

liche Aufwand an Zeit und Kosten selbst dann unverhältnismäßig bleibt, wenn die Entzifferung für sich genommen nicht mehr mit unverhältnismäßigem Rechenaufwand verbunden ist. Nur so kann ein unverhältnismäßig hoher Aufwand für die drei Aspekte Zeit, Kosten und Arbeitskraft hergestellt werden. Dies ermöglicht eine gesetzeskonforme reversible Anonymisierung mit technischen Mitteln.

4 Technische Lösung: Reversible Anonymisierung

Eine reversible Anonymisierung und damit die Erfüllung datenschutzrechtlicher Vorgaben, kann erreicht werden, wenn Datensätze nach dem Stand der Technik verschlüsselt, in mehrere Fragmente zerlegt und die Fragmente anschließend in voneinander unabhängigen Speichersystemen mit unabhängigen Zugriffssicherungen nach dem Stand der Technik gespeichert werden. Die Fragmentierung kann zufällig bitweise oder blockweise, beipielsweise byteweise erfolgen.

Durch die Speicherung der Fragmente in unabhängigen Speichersystemen entsteht eine zusätzliche, in Beschaffungsaufwand bestehende Hürde, die sich qualitativ von den Sicherheitskomponenten Verschlüsselung und Fragmentierung unterscheidet. Sie stellt sicher, dass keine extern speichernde Stelle die Originaldaten alleine mit Rechenaufwand wiederherstellen kann. Die so hinzugefügte Sicherheitskomponente ist nicht durch eine Schwächung der verwendeten Verschlüsselungs- und Fragmentierungsverfahren betroffen und stellt in Kombination mit Verschlüsselung und Fragmentierung sicher, dass die Herstellung des Personenbezugs für jede andere als die auslagernde Stelle im Hinblick auf Zeit, Kosten und Arbeitskraft auf Dauer unverhältnismäßig ist.

Ein Angreifer müsste somit die beteiligten externen Speichersysteme identifizieren, deren Zugriffssicherungen überwinden, alle Fragmente kopieren, die Fragmente zusammensetzen und die Verschlüsselung brechen.

Werden Datensätze wie beschrieben verschlüsselt, fragmentiert und gespeichert, sind die Daten in den Fragmenten anonymisiert. Die Übermittlung der Fragmente an externe Speicher-Systeme und deren Speicherung in externen Speichersystemen fällt nicht mehr in den Anwendungsbereich des BDSG.

Fragmente, die nach dem zuvor beschriebenen Verfahren erzeugt wurden, können, ohne Beachtung der für personenbezogene Daten einzuhaltenden rechtlichen Vorgaben, an externe Speichersysteme übermittelt und dort gespeichert werden. Die intern verantwortliche Stelle muss die Vorschriften des BDSG jedoch weiterhin beachten. Denn die ausgelagerten Daten sind, wie zuvor dargelegt, nicht für die intern verantwortliche Stelle anonymisiert, sondern nur für jede andere Stelle. Bei der Auswahl der externen Speicheranbieter sollte beachtet werden, dass eine Kooperation der Anbieter unwahrscheinlich ist.

Weitere Anforderungen ergeben sich aus der Regelung in § 30 Abs. 1 BDSG, die schon im Zusammenhang mit der Bestimmung des Begriffs der personenbezogenen Daten relevant geworden war. Sie legt fest, dass die für eine Identifizierung erforderlichen Zusatzinformationen separat zu speichern sind. Dies bedeutet im Fall der reversiblen Anonymisierung, dass die Informationen zum Wiederherstellen der Daten nicht zusammen mit den Frag-

menten ausgelagert werden dürfen. Daher werden diese Daten im lokalen Speichersystem abgelegt.

5 Föderiertes Speichersystem

Das föderierte Speichersystem verbindet mehrere Cloud-Speichersysteme und deren jeweilige Systeme zu einem Speicherverbund [MWS13]. Mit diesem System ist es beispielsweise möglich, die Speicherressourcen an mehreren Universitäten im Verbund zu nutzen. Das System basiert auf dem Ansatz eines Object-Storage-Systems mit frei wählbarem Datenspeicherort und wird durch eine Abstraktions- und Verwaltungsschicht über der lokalen Speicherverwaltung ermöglicht. Die Abstraktionsschichten und die Kommunikation zwischen ihnen ist in Abbildung 1 dargestellt. Die Abbildung zeigt für die Rechenzentren A und B das vorgeschlagene Schichtenmodell. Die Kommunikation zwischen den Rechenzentren erfolgt über die Schicht der föderierten Speicherverwaltung. Die Rechenzentren C und D sind äquivalent aufgebaut. Wie in der Abbildung zu sehen ist, setzt die föderierte

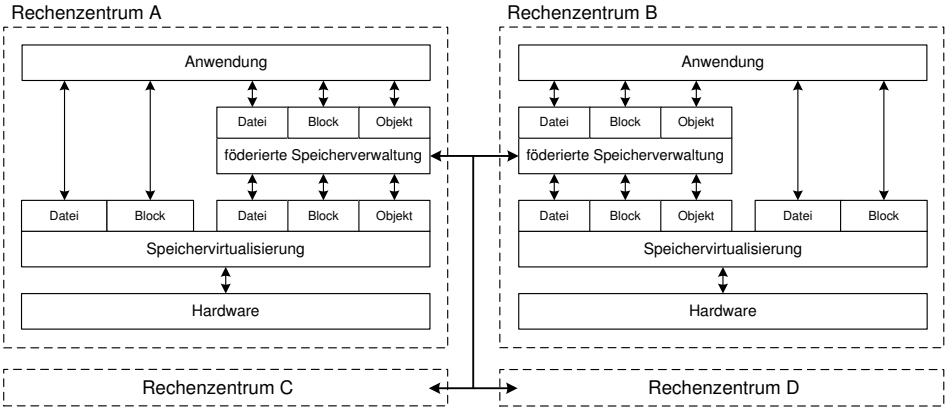

Abbildung 1: Schichtenmodell des föderierten Speichersystems mit Kommunikationsbeziehungen

auf eine lokale Speicherverwaltung auf und nutzt die bereitgestellten Schnittstellen, um auf Speichercontainer zuzugreifen. Applikationen greifen über die föderierte Speicherverwaltung auf die Speichercontainer zu. Eine detaillierte Beschreibung der einzelnen Schichten ist in der früheren Arbeit [MWS13] zu finden.
Jede Anwendung ist in der Lage, mit Hilfe von Metadaten zu definieren, welche Anforderungen sie an den Datenschutz ihrer gespeicherten Daten hat. Die föderierte Speicherverwaltung stellt anschließend sicher, dass die Datenschutzdefinition eingehalten wird, auch wenn die Daten den lokalen Standort verlassen und an einen weiteren Standort innerhalb der Föderation ausgelagert werden. Die Datenschutzdefinition umfasst aktuell die Möglichkeit, die Daten nach mehreren Sicherheitsleveln zu klassifizieren. Personenbezogene Daten werden dabei mit dem höchsten Sicherheitslevel abgespeichert.

Die Definition sieht aktuell wie folgt aus:

```
SecurityLevel = 0|1|2|9
0: Daten nicht verschlüsseln
1: Daten beim Auslagern verschlüsseln
2: Daten lokal und beim Auslagern verschlüsseln
9: personenbezogene Daten
```

Der in Kapitel 4 beschriebene Ansatz zur reversible Anonymisierung wurde im Rahmen dieses föderierten Speichersystems beispielhaft implementiert. Das föderierte Speichersystems basiert auf dem Object-Storage System OpenStack Swift [8]. In Swift-Systemen werden Objekte in Containern gespeichert und Container in Accounts. Objekte umfassen Daten (Content) und Metadaten (Header). Die Metadaten werden unter anderem verwendet, um die definierten Sicherheitslevel zu speichern. Eine Sicherheitslevel-Definition wird als String gespeichert und kann wie folgt aussehen:

```
X-Object-Meta-Security = "9"
```

Eine strukturelle Übersicht der Komponenten der föderierten Speicherverwaltung ist in Abbildung 2 gegeben. Wie der Grafik entnommen werden kann, dient der Federated-Proxy

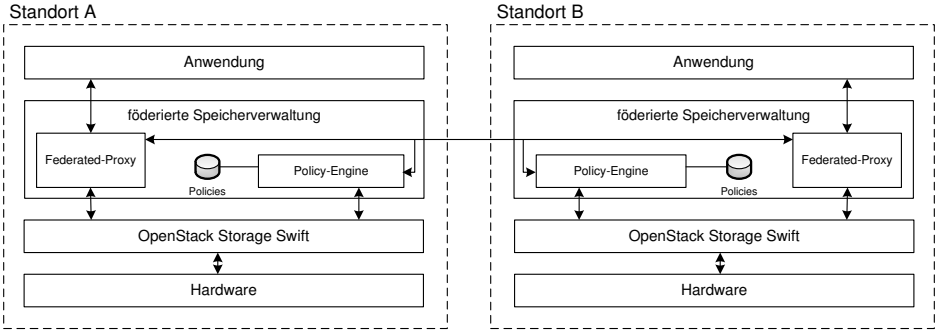

Abbildung 2: Komponenten der föderierten Speicherverwaltung

den Applikationen als Schnittstelle zum föderierten Speichersystem. Er sorgt dafür, dass Daten entsprechend ihrer Sicherheitsstufe abgelegt werden.

Das Policy-Tool liest Statistik-Daten aus dem Speichersystem aus und bestimmt anhand hinterlegter Policies immer wieder neu, in welchem Speichersystem Daten abzulegen sind. Policies sind automatische Regeln, die es ermöglichen, Daten zwischen den Standorten zu verschieben. Dies ermöglicht es, Daten, die nicht länger lokal benötigt werden, in einen Überlaufspeicher auszulagern. Das Verschieben der Daten an einen anderen Standort wird als ”export” bezeichnet. Die entsprechende Umkehrfunktion wird als ”import” bezeichnet.

[8] Object Storage System OpenStack Swift: `http://swift.openstack.org`

Der Syntax der Policies ist wie folgt definiert:

```
RULE ruleName
 EXPORT [source storage condition]
  FROM containerName [data condition]
  TO target [target storage condition]
```

```
RULE ruleName
 IMPORT [source storage condition]
  CONTAINER containerName [data condition]
```

Die Bedingungen in eckigen Klammern sind optional. Der Name der Policies, `ruleName`, ist notwendig, um die Regeln zu unterscheiden. Danach wird die Aktion angegeben und gegebenenfalls mit einer Bedingung verknüpft (beispielsweise mehr als 80 % der lokalen Speicher-Ressourcen ist aufgebraucht). Die Daten, die verschoben werden sollen, werden als `containerName` angegeben und können über die `data condition` eingeschränkt werden. So ist es beispielsweise möglich, nur Daten zu verschieben deren Änderungsdatum älter ist als ein bestimmtes Datum. Zuletzt wird das Ziel-Speichersystem `target` angegeben. Auch hier ist es wiederum möglich, eine Bedingung anzugeben (z. B. Größe des noch verfügbarer Speichers).

Der Transporter verlagert Daten entsprechend ihrem Sicherheitslevel und in Übereinstimmung mit datenschutzrechtlichen Anforderungen in und zwischen den Swift-Systemen. Er verfügt über einen Load-Balancer und mehrere Worker. Die Worker sind jeweils als separate Prozesse mit mehreren Threads implementiert und kommunizieren über Sockets mit dem Load-Balancer. Der Transporter implementiert das Auslagern von personenbezogenen Daten über die reversible Anonymisierung.

Werden Daten durch eine Applikation im föderierten Speichersystem abgelegt, geschieht dies über den Federated-Proxy. Dieser behandelt alle Daten entsprechend ihrer Metadaten. Wird kein Sicherheitslevel angegeben, behandelt der Federated-Proxy die Daten wie personenbezogene Daten.

1. Datensatz empfangen
2. Verschlüsseln, wenn SecurityLevel = 2 oder 9
3. Datensatz mit Metadaten lokal schreiben

Falls verschlüsselt wird, wird die symmetrische Blockchiffre AES verwendet. Dabei wird der Objekt-Header um die zusätzlichen Metadaten für die Verschlüsselung erweitert. Die Schlüssel für die Verschlüsselung werden in der lokalen Benutzerverwaltung gespeichert.

5.1 Datenschutzkonformes Auslagern von Daten

Der Export von Daten in externe Speichersysteme innerhalb der Föderation wird über ein Transport-Kommando des Policy-Tools ausgelöst und vom Transporter ausgeführt. Das in Kapitel 4 beschriebene reversible Anonymisierungsverfahren wird angewendet, wenn personenbezogene Daten (SecurityLevel 9) an externe Speichersysteme weitergegeben werden. Dabei wird folgender Ablauf angewandt:

1. Verschlüsselten Datensatz mit Metadaten lokal lesen
2. Verschlüsselten Datensatz fragmentieren
3. Fragmente extern speichern
4. Link-Informationen verschlüsseln

5. Verschlüsselte Link-Informationen lokal speichern

Die Fragmentierung erfolgt im aktuellen Prototyp in 8 Bit-Blöcken. Die Verteilung der Blöcke auf die Fragmente wird über ein Fragmentierungsmuster aus Zufallszahlen bestimmt. Sollte zu einem Zeitpunkt die verwendete Verschlüsselungsmethode unsicher geworden sein, bleiben die in den Fragmenten gespeicherten Daten anonymisiert. Da keines der Fragmente die verschlüsselten Daten enthält, kann eine Entzifferung erst nach der Defragmentierung erfolgen. Die Anzahl der Fragmente wird aktuell statisch festgelegt und stellt somit eine harte Anforderung an die benötigte Anzahl externer Speichersysteme. Dabei wird genau ein Fragment in einem externen Speichersystem abgelegt. Stehen eine unzureichende Zahl externer Speichersysteme zur Verfügung, können die Daten nicht ausgelagert werden. Bei der aktuell verwendeten Fragmentierungstechnik werden die Daten ohne Redundanz aufgeteilt und exportiert. Beim Zugriff auf die Daten müssen alle externen Speichersysteme verfügbar sein. Um diese Beschränkung zu umgehen, können RAID-ähnliche Fragmentierungstechniken eingesetzt werden, die über Redundanzen den Ausfall kompensieren können.

Um die Daten später wieder zusammensetzen zu können, werden Link-Informationen im lokalen Speichersystem abgelegt. Diese werden im Datenbereich (Content) des ursprünglichen Objekts gespeichert. Sie enthalten Metadaten, die auf den Speicherort, den Container und den Objektnamen der Fragmente im externen Speichersystem verweisen. Der externe Objektname wird über einen String gebildet, der aus dem internen Speicherzeitpunkt, der Kopie-Nummer, dem internen Objektnamen und der Fragment-Nummer besteht. Dadurch wird für externe Angreifer die Zuordnung der Fragmente zueinander erschwert.

5.2 Lesen von Daten

Der Zugriff auf die Daten erfolgt immer über den Federated-Proxy. Dieser liest die Metadaten des angeforderten Objekts, um zu unterscheiden, ob die Daten lokal oder extern abgelegt sind.
Sind die Daten lokal gespeichert, ist es möglich, dass die Daten verschlüsselt vorliegen (SecurityLevel 2 oder 9). In diesem Fall werden die Daten entschlüsselt und der Anwendung bereitgestellt.
Sind die Daten in ein externes Speichersystem ausgelagert worden, werden die lokalen Link-Informationen gelesen. Anschließend werden die Fragmente aus den externen Systemen gelesen, defragmentiert, entschlüsselt und an die Anwendung übertragen. Der Federated-Proxy stellt insgesamt sicher, dass der Nutzer bzw. die Applikation die zum Speichern an den Federated-Proxy übermittelten Daten in dieser Form auch zurückerhält.

6 Fazit und Ausblick

Der Verwendung von Cloud-Speichersystemen standen bisher meistens Datenschutzbedenken gegenüber. Diese Bedenken sind gerade bei personenbezogenen Daten absolut gerechtfertigt und lassen sich juristisch begründen. Da selbst eine Verschlüsselung personenbezogener Daten nicht als ausreichender Schutz beim Auslagern von Daten angesehen werden kann, schien bisher eine Verwendung solcher Systeme nur im Rahmen einer Auftragsdatenverarbeitung möglich.

Diese Arbeit zeigt einen alternativen Ansatz, um eine problematische Auftragsdatenverarbeitung zu umgehen. Das vorgestellte Verfahren der reversiblen Anonymisierung beschreibt, wie Daten derart verändert werden können, damit sie nach den Anforderungen des BDSG als anonymisiert betrachtet werden können. Die Daten werden dabei nach dem Stand der Technik verschlüsselt, in mehrere Fragmente zerlegt und die Fragmente anschließend in voneinander unabhängigen Speichersystemen mit unabhängigen Zugriffssicherungen nach dem Stand der Technik gespeichert. Die beispielhafte Umsetzung dieses Verfahrens in einem föderierten Speichersystem hat gezeigt, dass eine solche Methode praktisch umgesetzt werden kann. Das föderierte Speichersystem übernimmt die Aufgabe, die Daten entsprechend ihrer Klassifizierung zu schützen, selbst wenn die Daten den ursprünglichen Standort verlassen.

Literatur

[AA07a] Artikel-29-Arbeitsgruppe. Opinion 05/2012 on Cloud Computing, WP 196. *Drucksachen Europäische Kommision*, 2007.

[AA07b] Artikel-29-Arbeitsgruppe. Stellungnahme 4/2007 zum Begriff „personenbezogene Daten“, WP 136. *Drucksachen Europäische Kommision*, 2007.

[ALPW10] Hussam Abu-Libdeh, Lonnie Princehouse und Hakim Weatherspoon. RACS: A Case for Cloud Storage Diversity. In *Proceedings of the 1st ACM Symposium on Cloud Computing*, SoCC '10, Seiten 229–240, New York, NY, USA, 2010. ACM.

[BCQ+11] Alysson Bessani, Miguel Correia, Bruno Quaresma, Fernando André und Paulo Sousa. DepSky: dependable and secure storage in a cloud-of-clouds. In *Proceedings of the sixth conference on Computer systems*, EuroSys '11, Seiten 31–46, New York, NY, USA, 2011. ACM.

[DWWK10] Wolfgang Däubler, Peter Wedde, Thilo Weichert und Thomas Klebe. *Bundesdatenschutzgesetz - Kompaktkommentar zum BDSG und anderen Gesetzen*. Bund-Verlag, 2010.

[EU-95] EU-Parlament. Richtlinie 95/46/EG. *Amtsblatt*, (L 281):31–50, Oktober 1995. Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr.

[GKKS12] Peter Gola, Christoph Klug, Barbara Körffer und Rudolf Schomerus. *BDSG Bundesdatenschutzgesetz*. Gelbe Erläuterungsbücher. Beck C. H., 2012.

[HW10] Joerg Heidrich und Christoph Wegener. Sichere Datenwolken - Cloud Computing und Datenschutz. In *MultiMedia und Recht*, 2010.

[MWS13] Konrad Meier, Dennis Wehrle und Nico Schlitter. Ein Konzept zum Aufbau eines föderierten, dezentralen Speichersystems im Hochschulumfeld. In *6. DFN-Forum Kommunikationstechnologien*. GI, 2013.

[SH12] D. Slamanig und C. Hanser. On cloud storage and the cloud of clouds approach. In *International Conference for Internet Technology and Secured Transactions*, Seiten 649–655, 2012.

[SMGL11] Zhonghua Sheng, Zhiqiang Ma, Lin Gu und Ang Li. A privacy-protecting file system on public cloud storage. In *International Conference on Cloud and Service Computing (CSC)*, Seiten 141–149, 2011.

[Spi11] Axel Spies. Cloud Computing: Keine personenbezogenen Daten bei Verschlüsselung. *MMR-Aktuell*, 2011.

[YXH13] Chuan Yao, Li Xu und Xinyi Huang. A Secure Cloud Storage System from Threshold Encryption. In *5th International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, Seiten 541–545, 2013.