Plattform zur Bereitstellung sicherer und hochverfügbarer Speicherressourcen in der Cloud

Maxim Schnjakin, Christoph Meinel[1]

Kurzfassung:
Immer mehr Unternehmen sehen sich vor dem Problem rasant wachsender Datenmenge, die sie verwalten müssen. Obwohl die Speicherkosten pro GB immer weiter sinken, müssen Firmen weiterhin regelmäßig in den Ausbau ihrer Rechenzentren, neuer Server, aktuelle Kühlung und möglichst niedrigen Stromverbrauch investieren. Mit Cloud-Computing können Unternehmen von den Vorteilen spezialisierter Dienstleister profitieren. Allerdings fürchten sich viele IT-Verantwortliche die Kontrolle über eigene Daten aus der Hand zu geben. In dieser Arbeit werden einige Probleme vorgestellt, die in Verbindung mit externer Datenaufbewahrung auftreten. Dabei ist die Frage der Zuverlässigkeit und des Risikos in eine Abhängigkeit von einem Dienstleister zu geraten zentral.
In unserer Lösung setzen wir RAID-ähnliche Techniken ein, um Daten der Anwender zu fragmentieren und unter Einhaltung nutzerspezifischer Anforderungen auf unabhängige Cloud-Ressourcen zu verteilen. Bei der Verteilung der Daten wird sichergestellt, dass kein Anbieter in vollständigem Besitz der Daten einzelner Anwender ist. Das Vorgehen erlaubt den Ausfall eines oder mehrerer Dienstleister ohne Datenverlust zu tolerieren, reduziert das Lock-in Risiko sowie auch die Gefahr eines möglichen Datenmissbrauchs seitens der Dienstleister.

Stichworte: Sicherheit, Cloudcomputing, Speicherdienste

1. Einführung

Cloud Computing beschreibt einen internet-zentrierten Ansatz zur Bereitstellung und Nutzung von IT-Ressourcen als elektronisch verfügbare Dienste. Das Modell ermöglicht Unternehmen die Wirtschaftlichkeit der IT zu verbessern und bietet gleichzeitig genügend Flexibilität zur Nutzung neuer Wachstumschancen und Trends. Denn Cloud Computing ermöglicht eine elastische Skalierbarkeit von Ressourcen, so dass die Infrastruktur stets an Veränderungen und aktuellen Bedarf angepasst werden kann.

Die Ressourcen werden von Anbietern (Provider) über festgelegte Schnittstellen als Dienste (Services) über das Internet bereitgestellt und nach ihrem tatsächlichen Verbrauch abgerechnet. Dabei brauchen sich Anwender nicht um die darunter liegende Technologie zu kümmern.

Aufgrund dieser Vorteile sagen zahlreiche Marktforschungsunternehmen Cloud Computing eine große Zukunft voraus. So rechnen Analysten von IDC für das Jahr 2012 mit einem Marktvolumen für Cloud Computing von annähernd 12 Milliarden US-Dollar weltweit [3]. Angesichts der stetig wachsenden Informationsbestände geniesst Cloud Speicher eine sehr hohe Popularität unter den Anwendern.

[1] Hasso Plattner Institut, Potsdam

Markforscher von iSupply rechnen damit, dass der Umsatz in dem Online-Speicher Segment bereits im Jahr 2013 auf rund 5.8 Milliarden US-Dollar (von derzeit 1.6 Milliarden) anwachsen wird [2].

Trotz wirtschaftlicher Vorteile zögern viele Unternehmen, interne Daten an externe Anbieter zu übertragen. Besonders wenn es sich dabei um vertrauliche Daten wie z.B. Kundeninformationen, Buchhaltung oder juristische Dokumente handelt.

Außer dem nachvollziehbaren Eigeninteresse der Unternehmen, interne Informationsbestände vor Außenstehenden zu schützen, gibt es bei bestimmten Daten klare gesetzlichen Vorschriften, wie diese zu verarbeiten und zu verwalten sind [6]. Branchenstandards und Richtlinien wie HIPPA5, Datenschutz, Payment Card Industry Data Security Standard oder Statement on Auditing Standards 70 legen eindeutige und messbare Sicherheitsbestimmungen fest, mit denen Unternehmen nachweisen müssen, wie Daten verarbeitet und gelagert werden. Im Falle einer so genannten Auftragsverarbeitung muss ebenfalls festgelegt werden, wo und von wem die Daten verarbeitet werden. Cloud-Anbieter versuchen zwar, eine geschützte Umgebung zu liefern, doch die Verantwortung für den sicheren Umgang mit Daten liegt bei den Unternehmen, die den Dienst nutzen.

Damit ist die Sicherheit im Cloud Computing nicht nur eine Frage der Technologie, sondern auch des Vertrauens. Denn einerseits müssen Dienstnutzer darauf vertrauen, dass Dienstanbieter ihre Daten vor Zugriffen unberechtigter Dritter ausreichend schützen. Andererseits müssen sie sich darauf verlassen, dass ihre Informationsbestände von Dienstanbietern nicht für eigene Zwecke missbraucht werden.

Darüber hinaus spielen Zuverlässigkeit und Verfügbarkeit bei externer Datenaufbewahrung auch eine wichtige Rolle. Beim Cloud Computing werden Netzwerk-, Verarbeitungs- und Speicherfunktionen auf eine sehr große Basis physischer und virtueller Ressourcen verteilt. Theoretisch soll dadurch eine wesentlich höhere Toleranz gegenüber einzelnen Hardware-Ausfällen erreicht werden. Allerdings kam es in letzter Zeit immer wieder zu Aufsehen erregenden Ausfällen diverser Online-Speicherdienste.

So kam es beispielsweise Anfang Oktober 2009 bei dem Mobiltelefon-Service Sidekick (T-Mobile USA) zu einen Zwischenfall bei dem Sidekick-Telefonkunden ihre persönlichen Daten einschließlich Kontaktnamen, Telefonnummern und digitale Fotos verloren haben[2].

Ein weiteres Problem tritt vor allem bei längerer Nutzung der selben Dienste auf und wird allgemein als Anbieter-Lock-in bezeichnet. Hier wächst die Abhängigkeit der Nutzer mit der Datenmenge, die an den externen Dienstleister übertragen wird.

[2] http://www.golem.de/0910/70525.html

Im Cloud Computing wird auf langfristige Verträge verzichtet. Sollte ein Dienstanbieter von heute auf morgen seine Preispolitik ändern, können Kunden nicht ohne Weiteres zu einem anderen Anbieter wechseln. Auf der einen Seite ist ein Wechsel des Anbieters mit signifikanten Kosten allein für die Übertragung der Daten verbunden. Denn sowohl für den eingehenden als auch für den ausgehenden Datentransfer werden in der Regel seitens der Dienstleister Gebühren erhoben. So kostet die Migration eines 50 Terabyte großen Archivs (von Rackspace nach Nirvanix) ca. 20 Tausend USD. Auf der anderen Seite müssen bestehende Anwendungen an die Schnittstellen neuer Anbieter angepasst werden, da jeder Provider seine Dienste über proprietäre APIs bereitstellt, welche ihren eigenen Besonderheiten und Einschränkungen unterliegen. Ferner gibt es auch keine Garantie, dass der neue Dienstleister seine Preisstruktur dauerhaft beibehält.

Eine Möglichkeit, diesem Problem entgegenzuwirken sowie die Zuverlässigkeit der Datenaufbewahrung zu erhöhen, besteht in der Verteilung der Daten auf mehrere Dienstanbieter. Dies führt allerdings zu zwei weiteren Problemen. Einerseits müssen in diesem Fall Unternehmen fortwährend den Markt beobachten, wobei die Auswahl eines geeigneten Speicherdienstes den Anwender/Entscheidungsträger immer wieder vor dieselbe Herausforderung stellt. Das Angebot ist unübersichtlich und vielfältig. Nach [7] gibt es bereits heute über 100 verschiedene Cloud-Speicher Anbieter. Andererseits müssten auch in diesem Fall die unterschiedlichen Technologien neuer Anbieter in die bestehende Infrastruktur integriert werden.

Bis heute gibt es unseres Wissens nach kein Framework, das den Anwender bei der Auswahl eines geeigneten Online-Speicherdienstes unter Berücksichtigung seiner individuellen Anforderungen unterstützt. Im Umfeld von Web Services wird die Problematik der Anbieterauswahl mit der Definition und Auswertung von Dienstgüteparametern angegangen. Beschrieben werden dabei beispielsweise Vereinbarungen über Ressourcenzuteilung, Verfügbarkeiten oder Reaktionszeiten. Im Cloud Computing Umfeld sind die Vereinbarungen über die Dienstleistungen noch in einem frühen Entwicklungsstadium und genügen kaum den Geschäftsanforderungen [4], [8]. Außerdem sind die Service Level Agreements (SLA) für gewöhnlich in natürlicher Sprache formuliert [1] und können somit nicht automatisch ausgewertet werden.

Fest steht, dass für jeden potentiellen Cloud Nutzer die Auswahl eines geeigneten Anbieters sowohl kostspielig aus auch zeitaufwendig ist. Denn bevor Anwender einem Anbieter ihre Daten anvertrauen, ist eine sorgfältige Prüfung unerlässlich, wobei eine individuelle Due Diligence höchst ineffizient ist.

Aus diesem Grund ist es nur schlüssig, die Überprüfungsfunktion an eine vertrauenswürdige dritte Partei zu übertragen, welche auf die Auswahl der passenden Provider anhand individueller Anforderungen der Anwender spezialisiert ist und nicht im direkten Besitz der Daten ist.

In unserer Arbeit integrieren wir sorgfältig ausgesuchte Cloud-Speicheranbieter in einer einheitlichen Plattform und stellen damit eine Metaebene zwischen Anwendern und diversen Cloud-Speicherdiensten bereit. Zum besseren Schutz der Inhalte, werden einzelne Datenobjekte vor ihrer Übermittlung fragmentiert und einmalig permutiert. Die Fragmente werden unter Einhaltung nutzerspezifischer Anforderungen auf verschiedene, von einander unabhängige Dienste verteilt.

Unser System überprüft die Einhaltung der Anwenderanforderungen und garantiert, dass kein Dienstanbieter im alleinigen Besitz der Anwenderdaten ist. Die Verteilung der Daten auf mehrere Anbieter erhöht nicht nur die Sicherheit und die Zuverlässigkeit der Daten bei externer Aufbewahrung, sondern verringert deutlich die Gefahr des Anbieter Lock-ins.

2. Ansatz

Im Grunde schlagen wir mit unserer Lösung die Brücke zwischen Risiken und Vorteilen externer Datenaufbewahrung. Um dies zu erreichen, sollen Unternehmensdaten auf mehrere unabhängige Quellen verteilt werden, wobei ein Großteil des Entscheidungsprozesses zur Ermittlung geeigneter Dienstanbieter automatisiert erfolgen soll. Ferner sollen Anwender ihre individuellen Anforderungen an das Hosting eigener Datenbestände festlegen können, ohne sich dabei um die Administration sowie Leistungskontrolle kümmern zu müssen. Die vorgestellte Architektur besteht im Wesentlichen aus drei Komponenten:

Nutzer-Schnittstelle:

Über diese Systemkomponente erhalten Anwender einen vollständigen Überblick über ihre Datenbestände sowie verfügbare technischen Features. Damit können Nutzer ihre Daten verwalten sowie Anforderung an deren Verwahrung festlegen (z.B. in Form von Dienstleistungsparametern wie Reaktionszeit oder Verfügbarkeit). Es besteht die Möglichkeit neue Daten hochzuladen oder bestehende Inhalte zu verändern. Darüber hinaus können noch Anforderung bezüglich der Sicherheit, geographischer Lage und Kosten festgelegt werden.

Ressourcen-Management (RM) Modul:

Diese Komponente ist für eine intelligente Zuweisung von Daten an Cloud-Ressourcen verantwortlich und wird dabei von folgenden Diensten unterstützt:

- *Register- und Matching Service*: bestimmt Cloud-Speicher-Provider basierend auf Nutzeranforderungen. Darüber hinaus überwacht der Dienst die Leistung der Anbieter und stellt sicher, dass diese nicht gegen die garantierten Vereinbarungen verstoßen.
- *Ressourcenmanagement Service*: trifft alle operativen Entscheidungen bezüglich der Datenaufbewahrung sowie deren Verwaltung.

- *Task Scheduler*: bietet die Möglichkeit Arbeitsaufträge zu ordnen, zu priorisieren, sowie zeitversetzt auszuführen. Damit können Aufgaben in festgelegten Zeitintervallen oder außerhalb von Stoßzeiten kostengünstig ausgeführt werden.
- *Load Balancer: ermöglicht eine Verteilung der Arbeitslast auf verfügbare* Speicherdienste unter Berücksichtigung der Anwenderanforderungen.

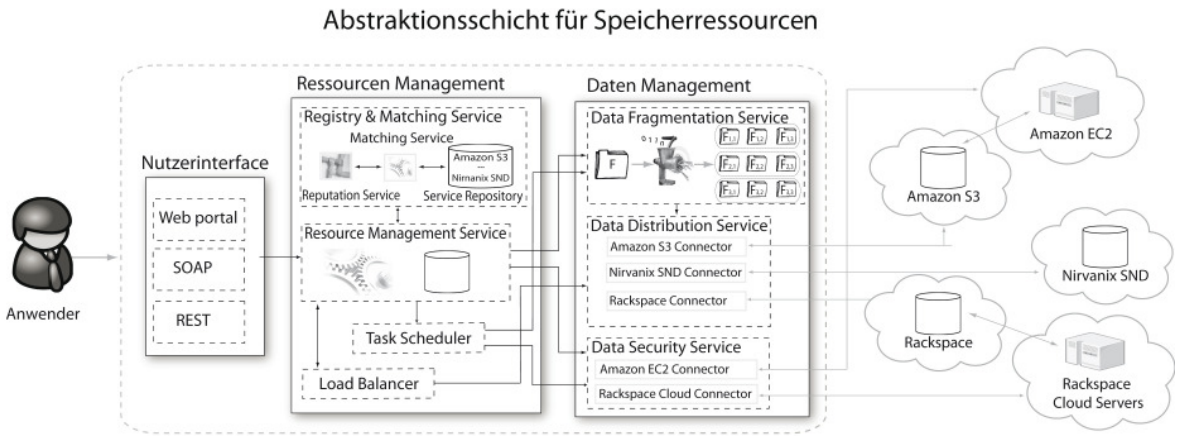

Abbildung 1: Metaebene für Cloud-Speicher-Ressourcen

Daten-Management (DM) Modul:

Die Systemkomponente wird von dem RM Modul angesteuert und ist für die physische Verteilung der Daten auf einzelne Cloud-Provider verantwortlich. Hierbei wird die Komponente von folgenden Diensten unterstützt:

- *Datenfragmentierungsservice:* ist für die Permutation und Fragmentierung einzelner Datenobjekte verantwortlich.
- *Datenverteilungsservice:* verteilt separate Fragmente auf verschiedene Speicheranbieter. Die Kommunikation mit jedem einzelnen Provider findet über so genannte "Storage-Adapter" statt. Diese kapseln die Funktionalität und vereinheitlichen die Kommunikation mit den proprietären Schnittstellen der Dienstanbieter.
- *Sicherheitsservice:* ist für die Durchsetzung festgelegter Sicherheitsrichtlinien verantwortlich.

Unser Ansatz verfolgt die Idee, bei Nutzung von Cloud-Speicherressourcen die Anwenderdaten nicht einem einzigen Anbieter anzuvertrauen sondern gleichmäßig auf mehrere zu verteilen, um somit sowohl die Verfügbarkeit, Zuverlässigkeit als auch die Sicherheit der Daten zu erhöhen. Im Grunde ist unser Herangehen mit einer Service-orientierten Ausführung der RAID-Technologie vergleichbar. RAID-Systeme verbinden mehrere physische Festplatten zu einem logischen Laufwerk zum Erreichen höherer Transfer- und Ausfallraten. So arbeiten RAID-Systeme ab Level 2 mit der Aufspaltung der Daten und der Verteilung einzelner Fragmente auf verschiedene Hardwareressourcen. In unserem System setzen wir das selbe Prinzip für Cloud-

Speicher-Ressourcen ein. Zur Aufteilung sowie Rekonstruktion der Daten werden Erasure Coding Techniken eingesetzt. Der große Vorteil bei diesem Vorgehen liegt in dem günstigen Verhältnis des Speicherbedarfs relativ zur dazu gewonnenen Datenverfügbarkeit. Das Vorgehen ermöglicht unserer Plattform, den Ausfall von einem oder sogar mehreren Online-Speicherdiensten zu tolerieren, ohne Daten der Anwender zu verlieren.

3. Aufbau der Plattform

In diesem Abschnitt soll die Funktionsweise der wichtigsten Komponenten vorgestellt werden.

Nutzerinterface

Grundsätzlich stehen einem Anwender zwei Möglichkeiten zur Interaktion mit dem System zur Verfügung: maschinenlesbare APIs (SOAP und REST) und die nutzerorientierte Schnittstelle. Die standardisierten REST- und SOAP-Schnittstellen sollen Entwickler bei der Durchführung häufiger und komplexer Aufgaben unterstützen und befinden sich momentan im Entwicklungsstadium. Zum gegenwärtigen Zeitpunkt ist das System nur über das grafische Web-Interface nutzbar. Hier stehen zwei Funktionalitäten im Vordergrund: Datenverwaltung und Anforderungsspezifikation. Damit stehen Anwendern grundlegende Funktionen wie das Lesen, Löschen, Kopieren oder das Hochladen neuer Objekte zur Verfügung. Darüberhinaus können über das Nutzerinterface auch Anforderungen bezüglich der Datenverwahrung festgelegt werden. Insgesamt stehen folgende Optionen zur Verfügung:

- *Geografische Lage:* Mit der Einstellung können Anwender den Verwahrungsort ihrer Daten geografisch eingrenzen. Grundsätzlich sind die Standorte der Anbieter in geografische Regionen und politische Zonen aufgeteilt. Zum gegenwärtigen Zeitpunkt stehen folgende Optionen zur Verfügung: Nordamerika, Europa und EU.
- *Sicherheit:* Mit der Option können Nutzer Sicherheitsrichtlinien bezüglich der Datenaufbewahrung bestimmen (z.B. Verschlüsselung sowie die Verschlüsselungsstärke).
- *Dienstgüte:* An dieser Stelle können Leistungserwartungen in Form von Dienstgüteparametern angegeben werden (Verfügbarkeit, Bandbreite, Reaktionszeit, etc.).
- *Budget:* Mit der Budget-Option können Kosten für die grundsätzliche Datenaufbewahrung sowie erweiterte Funktionalität festgelegt werden.

Die festgelegten Einstellungen können sich gegenseitig beeinflussen. So kann die Auswahl der Verschlüsselungsfunktionalität eine Verzögerung bei der Verfügbarkeit der Datenbestände verursachen, da die angeforderten Daten vor der Übertragung erst

entschlüsselt werden müssen. Für das Ver- und Entschlüsseln der Daten fallen Extrakosten an. Dies ist damit zu erklären, dass zusätzlich zu den Basisgebühren für die Übertragung und das Hosting von Daten noch weitere Kosten für die Inanspruchnahme der Rechenressourcen anfallen.

Prinzipiell wird die Erhöhung der Verfügbarkeit durch redundante Speicherung der Datensätze erreicht. Für kostenbewusste Nutzer wird die Redundanz der Aufträge mittels Erasure Coding, einer komplexen Form der Datenspeicherung, erreicht. Der Nachteil besteht allerdings in der aufwendigen Kodierung und Dekodierung der Fragmente und äußert sich in einer Verzögerung bei der Bereitstellung der Inhalte. Mit höherem Budget kann die Performanz beim Zugriff auf die gehosteten Inhalte gesteigert werden. Denn ein größerer Finanzrahmen ermöglicht den Einsatz naiver Replikation. Es können beispielsweise so viele Datenfragmente (ohne vorheriger Kodierung) auf verschiedene Online-Ressourcen repliziert werden, bis eine festgelegte Grenze pro Gigabyte (GB) erreicht ist.

Somit variieren die Gesamtkosten abhängig von den festgelegten Nutzeranforderungen. Mit der Auswahl einzelner Features wird ihre direkte Auswirkung auf die Kosten abgeschätzt und visuell dargestellt.

Ressourcenmanagement Service

Diese Systemkomponente überwacht die Verteilung der Daten und ist außerdem für diverse Verwaltungsaufgaben verantwortlich:

- *Datenmanagement:* Der Dienst verfügt über eine MySQL Datenbank in der alle wichtigen Informationen über die aktuellen Datenbestände gespeichert sind. Dazu gehören: Nutzerdaten, verwendete Speicherdienste sowie entsprechende Zugriffsberechtigungen, Metadaten zur Rekonstruktion verteilter Fragmente. Vergleichbar einer Datei im lokalen Dateisystem kann ein Datenobjekt gelesen und/oder geschrieben werden. Da dies konkurrierend von mehreren Anwendern geschehen kann, muss die Konsistenz der Daten gesichert werden. Der Dienst überwacht Änderungen an Datenobjekten und propagiert diese auf alle Kopien betroffener Objekte.
- *Monitoring:* Die Komponente überwacht die Kommunikation mit den beteiligten Dienstanbietern hinsichtlich der Einhaltung vereinbarter SLAs. Verstöße werden protokolliert und wirken sich negativ auf die Reputation der Dienste aus. Dies kann dazu führen, dass die Managementkomponente die entsprechenden Speicherdienste sowohl bei der Verteilung der Daten als auch bei der Wiederzusammensetzung übergeht. Beim Zugriff auf redundant vorhandene Einträge werden Quellen mit besserer Reputation bevorzugt. Somit wird das Monitoring in erster Linie nicht zur finanziellen Entschädigung der Nutzer im Falle eines Verstoßes eingesetzt. Es dient vielmehr dazu bei, dass das System langfristig eigene Entscheidungen über die Zuverlässigkeit der Dinsteanbieter treffen kann.

- *Zeitzuteilung:* Der Management-Dienst ist außerdem für die Priorisierung und das Scheduling nicht zeitkritischer Aufgaben zuständig. Das betrifft vor allem die Vervielfältigung der Inhalte auf zusätzliche Speicherquellen und die Sicherheitsfunktionalität. Verschiedene Anbieter bieten Preisnachlässe auf freie Kapazitäten. Das bedeutet, dass die aktuellen Preise sich abhängig von Angebot und Nachfrage in regelmäßigen Abständen ändern. Wir wollen den Anwendern die Möglichkeit geben, von diesen Nachlässen zu profitieren. Das System beobachtet die gegenwärtigen Preise (über die Schnittstelle zum Register-Service) und nutzt günstige Intervalle zur Ausführung zeitlich flexibler Aufgaben (in einem von dem Anwender festgelegtem Zeitabschnitt).
- *Budgetierung:* Darüber hinaus stellt der Management-Service sicher, dass das vom Anwender festgelegte Budget zu keinem Zeitpunkt überschritten wird. Das System reagiert automatisch auf Preisänderungen bei beteiligten Anbietern. Abhängig von Nutzereinstellungen können beispielsweise bei Preisverfall die Daten der Anwender bis zum Erreichen eines festgelegten Betrags auf zusätzliche Quellen repliziert werden. Bei Preiserhöhung werden dagegen die redundanten Einträge unter der Berücksichtigung der Nutzereinstellungen entfernt.

Register- und Matching Service

Der Dienst ist für die Auswahl geeigneter Speicher-Anbieter verantwortlich. Hierzu müssen die Nutzeranforderungen mit den technischen Möglichkeiten der Dienstanbieter abgeglichen werden. Allerdings unterscheiden sich SLAs verschiedener Anbieter in ihrer Struktur und den zugesicherten Leistungen. Zur automatischen Selektion geeigneter Anbieter wurden die Qualität und der Umfang der Leistungen in eine standardisierte Form gebracht und in einer Datenbank abgelegt. Dabei beschränken sich die Beschreibungen nur auf die zur Auswahl notwendigen Parameter. Mit der Erweiterung der Funktionalität auf weitere Dienstleister muss die Beschreibung der Dienstgüte sowie weiterer Provider bezogener Details (z.B. physische Lage der Rechenzentren) manuell in die Datenbank eingepflegt werden.

Prinzipiell hängt die Anzahl der Dienstanbieter zur Verteilung der Daten von den Anforderungen der Nutzer an die Sicherheit, Verfügbarkeit, geografische Lage, Reaktionszeit etc. ab. Zum gegenwärtigen Zeitpunkt unterstützt das System drei Cloud-Speicheranbieter: Amazon S3, Rackspace Cloud Files und Nirvanix SND.

Datenfragmentierungsservice

Im Wesentlichen ähnelt das Datenmodell unseres System dem des Amazon S3. Alle Datenobjekte werden in so genannten "Buckets" abgelegt. Jedes Bucket besitzt einen eindeutigen Erkennungsschlüssel und wird systemweit nur ein Mal vergeben. Eine Schachtelung von Buckets ist nicht möglich. In einem Bucket können beliebig viele

Datenobjekte abgelegt werden. Allerdings dürfen einzelne Objekte eine Gesamtgröße von 5 GB nicht überschreiten.

Für jeden Anwender werden *n* Speicherbehälter eingerichtet. Diese repräsentieren Buckets, die später auf verschiedene Cloud-Storage Anbieter verteilt werden. Vor der Datenübertragung werden die einzelnen Datenobjekte einmalig permutiert. Obwohl einzelne Anbieter nicht über vollständige Datenobjekte, sondern nur über deren Blöcke verfügen, wären sie ohne diesen Schritt in der Lage, die anvertrauten Inhalte bis zu einem gewissen Maß zu interpretieren (z.B. bei Audio- oder Videomaterial). Der Permutationsschritt schafft eine initiale Hürde indem er die Dateninterpretation durch Dritte erschwert.

Die Fragmentierung der Datenobjekte erfolgt mittels Erasure Coding Techniken. Hierbei wird ein Datenobjekt in *m* Blöcke unterteilt, die wiederum in *n* Fragmente kodiert werden. Dabei gilt stets $n > m$ [5]. Die Fragmente werden anschließend auf verschiedene Speicherquellen verteilt. Der Vorteil bei dieser Form der Kodierung liegt darin, dass jeder Datenblock mit beliebigen *m* der *n* Fragmente zu rekonstruieren ist. Die Kodierungsrate ergibt sich aus dem Verhältnis *m/n*. Der zusätzliche Speicherbedarf ergibt sich aus dem umgekehrten Verhältnis *n/m*. Bei Verteilung der Daten auf 9 Anbieter, wäre das System in der Lage den Verlust einer Quelle zu tolerieren. Der zusätzliche Speicherbedarf würde sich dabei auf lediglich 10 Prozent erhöhen. Aktuell unterstützt das System nur drei Dienstleister, wobei die Ausdehnung der Funktionalität auf weitere Anbieter geplant ist. Abhängig von Nutzeranforderungen kann auf aufwendige Kodierung zu Gunsten der Performanz verzichtet werden. Bei naiver Replikation steigen allerdings die Kosten proportional zur Anzahl verteilter Kopien.

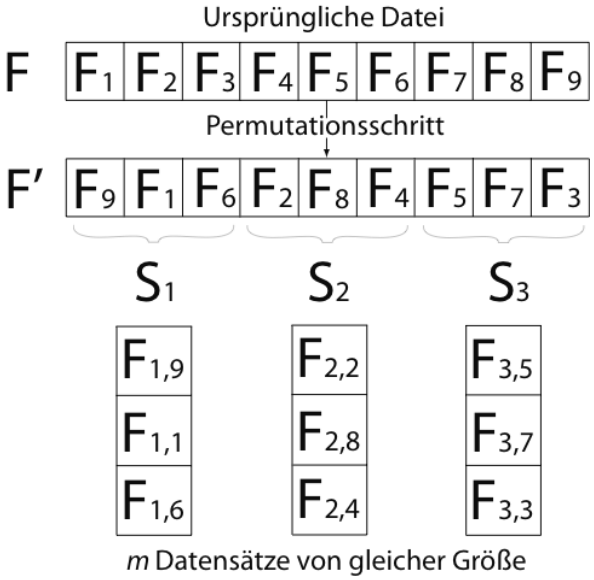

Abbildung 2: Vorgehen bei Datenfragmentierung

Datenverteilungsservice

Die Kommunikation mit jedem einzelnen Cloud-Provider findet über so genannte "Storage-Adapter" statt. Jeder Adapter unterstützt die grundlegende Funktionalität zur Datenmodifikation (*put, get, delete:* object) und übersetzt diese in das jeweilige Format der Anbieter. Der Abstraktionsschritt kapselt die technische Komplexität im Umgang mit den proprietären Schnittstellen verschiedener Dienstanbieter und macht das System flexibel erweiterbar. Zur Ausdehnung der Funktionalität auf zusätzliche Dienstleister ist lediglich die Implementierung eines entsprechenden Adapters erforderlich.

Security

Das initiale Sicherheitsniveau wird durch die Fragmentierung, Permutation und das Verteilen initialer Datenobjekte auf unterschiedliche Cloud-Provider gewährleistet. Allerdings schließt der Transformationsschritt das Durchsickern wichtiger Informationen nicht aus. Besonders wichtige Inhalte sollten daher verschlüsselt aufbewahrt werden. Um dies zu ermöglichen, nutzt das System das Angebot der Cloud-Anbieter, die zusätzlich zu Speicherressourcen auch Rechenressourcen als Dienst anbieten.

Zum gegenwärtigen Zeitpunkt werden von dem System zwei solcher Dienstanbieter unterstützt: Amazon Elastic Compute Cloud (EC2) und Rackspace Cloud Server. Mit EC2 bietet Amazon die Möglichkeit, virtuelle Maschinen in der Cloud zu konfigurieren und zu starten. Dies erfolgt über so genannte Amazon Maschine Images (AMIs). Aus Nutzerperspektive sieht die angebotene Infrastruktur wie eine physische Hardware aus. Zur Verwaltung der Infrastruktur werden spezielle Webservices zur Verfügung gestellt. Innerhalb nur weniger Minuten können beliebig viele Instanzen (virtuelle Server) gestartet und wieder entfernt werden. Mit Rackspace Cloud Servers steht ein ähnliches Angebot zur Verfügung. Die Konfiguration der Instanzen erfolgt über eine spezielle von Rackspace bereitgestellte Schnittstelle. Die Server verfügen über einen vollständigen Root-Zugang und können individuell eingerichtet werden.

Zur Bereitstellung der Verschlüsselungsfunktion sind bei jedem Anbieter spezielle Instanzen eingerichtet worden, die je nach Bedarf beliebig skaliert werden können. Ähnlich wie bei der Verteilung der Daten erfolgt die Kommunikation mit den jeweiligen Instanzen über spezielle "Security-Service-Adapter". Im Auftrag des Ressourcen Management Services können gehostete Inhalte providerseitig ver- und entschlüsselt werden.

4. Diskussion

In dieser Arbeit haben wir ein System vorgestellt, das eine Metaebene zwischen Anwendern und Anbietern von Cloud-Speicherressourcen bereitstellt. Bei der Übertragung der Daten an die Plattform werden die Datensätze der Anwender fragmentiert und auf verschiedene Dienstleister verteilt. Einzelne Cloud-Ressourcen

werden sorgfältig nach den benutzerdefinierten Anforderungen an die Leistungsfähigkeit, geografische Lage, sowie etwaige technische Eigenschaften ausgewählt. Dabei wird sichergestellt, dass kein Anbieter in vollständigem Besitz der Anwenderdaten ist. Das Vorgehen erhöht die Zuverlässigkeit bei externer Datenlagerung, reduziert das Lock-in Risiko sowie auch die Gefahr eines möglichen Datenmissbrauchs seitens der Dienstleister.

Der Einsatz von Erasure Coding erlaubt ein gutes Verhältnis zwischen Wirtschaftlichkeit und Effizienz. Denn im Gegensatz zu naiver Replikation, bei der der Speicherplatzbedarf und damit auch die Kosten linear zur Anzahl verteilter Kopien wachsen, ist unser System in der Lage, die selbe Datenverfügbarkeit mit einem Mehraufwand von lediglich 10 Prozent zu erreichen. Für externe Aufbewahrung besonders sensibler Inhalte besteht außerdem die Möglichkeit, Informationsbestände zu verschlüsseln. Die Plattform abstrahiert die technische Komplexität im Umgang mit den proprietären Schnittstellen verschiedener Dienstanbieter und ermöglicht seinen Nutzern einen einfachen Zugriff auf Cloud-Speicherressourcen.

Literaturhinweise:

[1] Amazon. Amazon ec2 service level agreement. Online, 2009.

[2] Jeffrey Burt. Future for cloud computing looks good, report says. Online, 2009.

[3] http://blogs.idc.com/ie/?p=543. Idc's new it cloud services fore-cast: 2009-2013. Online, 2009.

[4] A. Keller and H. Ludwig. The WSLA framework: Specifying andmonitoring service level agreements for web services. Journal of Network and Systems Management, 2004.

[5] D. Patterson, G. Gibson, and R. Katz. The case for RAID:Redundant arrays of inexpensive disks. ACM SIGMOD, 1988.

[6] Thomas Smedinghoff. Information Security: The Emerging Standard for Corporate Compliance. IT Governance Pub., 2008.

[7] Anthony T.Velte, Toby J. Velte, and Robert Elsenpeter. CloudComputing: A Practical Approach. Mc Graw Hill, 2009.

[8] Srikumar Venugopal, Xingchen Chu, and Rajkumar Buyya. A negotiation mechanism for advance resource reservation using the alternate offers protocol. Proceedings of the 16th Int. Workshop on Quality of Service, IWQoS, June 2008.