Cloud Storage und IT-Sicherheit: sichere und hochverfügbare Speicherressourcen in öffentlichen Clouds.

Maxim Schnjakin, Michael Goderbauer, Martin Krüger, Christoph Meinel

Kurzfassung:
Cloud Computing bietet die Möglichkeit, IT-Ressourcen auszulagern und über das Internet zugänglich zu machen. Auf der einen Seite haben Unternehmen hohe Erwartungen an das Konzept des *Cloud Computing,* auf der anderen Seite herrschen aktuell noch große Bedenken bezüglich der Sicherheit der Daten und der Einhaltung gesetzlicher Vorgaben wie z.B. dem Datenschutz. Diese Ausarbeitung betrachtet die beliebteste Cloud Anwendung — *Cloud Storage* — den allgemeinen rechtlichen Kontext bei der Nutzung und diskutiert die IT-Sicherheitsziele Verfügbarkeit, Authentifizierung, Autorisierung, Vertraulichkeit und Datenschutz. Nach der Diskussion bedeutender Herausforderungen wird ein Lösungsansatz präsentiert, welcher mit dem Einsatz RAID-ähnlicher Techniken und unter Einhaltung nutzerspezifischer Anforderungen, Nutzerdaten auf unabhängige Cloud-Ressourcen verteilt. Dabei 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. Zum Test des entwickelten Prototypen wurde ein Experiment durchgeführt, bei dem Daten zu verschiedenen Cloud Storage Anbietern im Zeitraum von mehreren Wochen (7x24) übertragen wurden. Im Rahmen der vorliegenden Arbeit werden ausgewählte Ergebnisse des Experiments vorgestellt, die über die Vor- und Nachteile des Forschungsansatzes hinaus, Aufschlüsse über das Verhalten der öffentlichen Cloud Speicher Anbieter liefern.

Stichworte: Sicherheit, Cloud Computing, Speicherdienste

1. Einführung

Cloud Computing stellt ein Modell dar, welches den bequemen, skalierbaren Netzwerkzugriff auf gemeinsame, konfigurierbare Ressourcen ermöglicht [1]. Eine dieser Ressourcen ist Speicher, der so genannte *Cloud Storage.* Durch Nutzung von Cloud-Storage-Diensten können Unternehmen ihre traditionell in eigenen Rechenzentren vorgehaltenen Daten an einen externen Dienstleister auslagern, der über festgelegte Schnittstellen über das Internet diese bereitgestellt und nach dem tatsächlichen Verbrauch abgerechnet. Die Beliebtheit und Relevanz von *Cloud Storage* für Unternehmen hat in den letzten Jahren vermehrt zugenommen [2]. Die Cloud als Datenspeicher gehört heute zu den beliebtesten Cloud Anwendungen [3]. Die Erwartungen dabei sind vielfältig: Bessere Skalierbarkeit, Kosteneinsparungen, Auslagerung nicht wertschöpfender Aktivitäten und sicheres Vorhalten der Daten. Trotz wirtschaftlicher Vorteile zögern jedoch 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.

Eine Möglichkeit, die Anbieter von Cloud Storage, *Cloud Storage Provider* (CSP), zu unterscheiden, ist diese in *Basic Storage Provider* (BSP) und *Advanced Storage Provider* (ASP) zu unterteilen [2]. BSPs betreiben eine eigene physikalische Infrastruktur zum Speichern der Daten. Üblicherweise bieten sie dem Endnutzer keine grafische Oberfläche zum Zugriff auf die Daten an. Stattdessen können die Daten über ein *Application Programming Interface* (API) programmatisch abgerufen werden. Zu den BSPs zählen u.a. Amazon S3[1], Google Storage[2] und Rackspace Cloud Files[3]. Im Gegensatz zu den BSPs betreiben ASPs keine eigene Infrastruktur zum Speichern der Daten und greifen hierfür auf die Dienste der erwähnten BSPs zurück. Dafür bieten sie dem Endnutzer in der Regel eine grafische Benutzeroberfläche zum Verwalten der Daten an, z.B. als Desktopanwendung oder als Web-Interface. Außerdem warten ASPs häufig mit Zusatzfunktionen auf, die über das reine Speichern von Daten hinausgehen. Dropbox[4] ermöglicht hier z.B. die Synchronisation ganzer Ordner über Rechnergrenzen hinweg. Neben Dropbox sind Google Drive5[5] und Apples iCloud[6] weitere Beispiele für ASPs. Das Angebot ist unübersichtlich und vielfältig. Nach [4] gibt es bereits heute über 100 verschiedene Cloud-Speicher Anbieter.

Bei der Nutzung von Cloud Ressourcen wird zunächst zwischen einer *Public* und einer *Private Cloud* unterschieden. Eine *Private Cloud* ist dadurch gekennzeichnet, dass die Ressourcen exklusiv für die Nutzung durch eine bestimmte Organisation bereitgestellt werden. Die Bereitstellung und Wartung übernimmt die Organisation entweder selbst oder ein Drittanbieter den Wünschen der Organisation entsprechend. Im Gegensatz dazu wird eine *Public Cloud* immer von einem Drittanbieter betrieben und mehreren Organisationen zur Nutzung angeboten. Diese müssen sich die bereitgestellten Ressourcen teilen [1].

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, Verfügbarkeit oder Reaktionszeit. Im Cloud Computing Umfeld sind die Vereinbarungen über die Dienstleistungen noch in einem frühen Entwicklungsstadium und genügen kaum den Geschäftsanforderungen [5], [6]. Außerdem sind die Service Level Agreements (SLA) für gewöhnlich in natürlicher Sprache formuliert und können somit nicht automatisch ausgewertet werden .

[1] https://aws.amazon.com/de/s3/
[2] https://developers.google.com/storage/
[3] http://www.rackspace.com/cloud/cloud_hosting_products/files/
[4] https://www.dropbox.com/
[5] https://drive.google.com
[6] https://www.icloud.com/

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

In dieser Arbeit sollen einige Sicherheitsaspekte des Auslagerns von Dateien in die Public Cloud betrachtet, sowie neue Lösungsansätze diskutiert werden. Hierfür wird die Annahme zu Grunde gelegt, dass ein deutsches Unternehmen seine möglicherweise vertraulichen oder personenbezogenen Daten in einer *Public Cloud* speichern möchte. Im Folgenden werden nun verschiedene Anbieter von *Cloud Storage* in der *Public Cloud* betrachtet. Für die Diskussion der Sicherheitsaspekte stehen in dieser Ausarbeitung die BSPs im Zentrum der Betrachtung. Im Einzelnen wurden die folgenden drei Dienste stellvertretend näher untersucht: Amazon S3, Google Storage, Rackspace Cloud Files.

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 [7]. 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.

Im nachfolgenden Kapitel II soll zunächst auf allgemeine rechtliche Fragen im Zusammenhang mit Cloud Storage eingegangen werden. Es schafft die Grundlage für die anschließende Diskussion verschiedener Sicherheitsziele im Kapitel III sowie Lösungsansätze in IV. Anschließend wird ein Verfahren vorgestellt, mit dem die meisten identifizierten Probleme gelöst werden können. Im darauf folgenden Kapitel V werden ausgewählte Ergebnisse einer umfassenden Evaluation des entwickelten Systems vorgestellt. Abschließend wird in Kapitel VI ein Fazit des Beitrages gegeben.

2. Rechtlicher Kontext

Bevor Daten in die Cloud ausgelagert werden können, muss mit einem Dienstleister ein Vertrag geschlossen werden. Hierbei gibt es ein paar grundsätzliche Fragen zu klären, die in diesem Kapitel näher erläutert werden sollen. Die Parteien, welche den Vertrag abschließen, unterliegen nicht zwingend dem gleichen Recht bzw. haben nicht den gleichen Standort. Hier gilt Vertragsfreiheit und es muss entschieden werden,

welches Landesrecht für die Auslegung des Vertrages anzuwenden ist. Außerdem muss geklärt werden, was im Falle von Vertragsverletzungen passiert und wie bei einem möglicherweise unvorhergesehen Vertragsende mit den gespeicherten Daten umgegangen wird. Daneben kann der CSP an zusätzliche Gesetze (wie z.B. den US-amerikanischen *Patriot Act)* gebunden sein. Außerdem macht das deutsche Bundesdatenschutzgesetz (BDSG) spezielle Vorgaben für den Umgang mit personenbezogenen Daten. Dessen Grundsätze kommen auch zur Anwendung, wenn ein deutsches Unternehmen solche Daten in die Cloud auslagern möchte.

Anzuwendendes Recht und Gerichtsstand

Bei Verträgen, die über das Internet abgeschlossen werden oder wenn die Vertragspartner ihren Sitz in unterschiedlichen Ländern haben, muss im Fall von Streitigkeiten bezüglich des Vertragsinhaltes geklärt werden, welches Gericht die Streitigkeit verhandelt und welches Recht anzuwenden ist. Hier kann es sein, dass verschiedene Rechte oder Gerichtsstände eine der Parteien unterschiedlich begünstigen. Daher ist es für den Kunden, also den Nutzer von *Cloud Storage* besonders wichtig, sich diesbezüglich zu informieren.

In Deutschland gilt — wie in vielen anderen Ländern der Welt auch — die Vertragsfreiheit nach Art. 2 I Grundgesetz (GG), d.h. die Vertragspartner können den Gerichtsstand und das Landesrecht, welchem sie den Vertrag unterwerfen, frei wählen[7]. Die untersuchten *Cloud Storage* Provider machen von dieser Freiheit Gebrauch und legen in ihren Vertragsbedingungen explizit den Gerichtsstand und das anzuwendende Recht fest. Sie alle entscheiden sich dafür, US-amerikanisches Recht zur Anwendung kommen zu lassen. Im Einzelnen gilt für Verträge mit Amazon das Recht des Staates Washington, für Verträge mit Google das Recht des Staates Kalifornien und für Verträge mit Rackspace das Recht des Staates Texas.

Falls es keine explizite Klausel für die Rechtswahl oder Wahl des Gerichtsstandes gibt, muss im Streitfall nach implizierten Faktoren entschieden werden. Hierzu zählen der Ort, an dem die für den Vertrag charakteristische Leistung erbracht wird, die Währung, in welcher die Vergütung vereinbart wird und die Sprache des Vertrages. Um unnötige Unsicherheit im Fall von Streitigkeiten zu vermeiden, sollte sich nicht auf die Interpretation implizierter Faktoren verlassen werden und explizite Klauseln formuliert werden. Ein deutsches Unternehmen, das im Streitfall deutsches Recht anwenden lassen will, sollte also Klauseln vereinbaren, welche die Anwendung von Bürgerlichem Gesetzbuch (BGB) und Handelsgesetzbuch (HGB) festhalten. Allerdings werden sich die drei in dieser Ausarbeitung betrachteten CSPs auf solche Regelungen nicht einlassen, da ihre Standardverträge andere Vorgaben machen (siehe oben). Nutzer dieser Dienste müssen sich also mit US-amerikanischem Recht auseinandersetzen.

[7] Ausnahmen gibt es von dieser Regelung nur bei Verträgen, die mit Verbrauchern geschlossen werden. Diese werden hier nicht weiter betrachtet.

Art des Vertrages

Die Rechte und Pflichten der Vertragspartner sind nicht nur abhängig von den in dem Vertrag festgelegten Klauseln. Je nach Art des Vertrages greifen verschiedene gesetzliche Regelungen, aus welchen sich Rechte, Pflichten und Ansprüche für beide Seiten ableiten lassen können. In diesem Absatz werden die für die Deutsche Unternehmen relevanten Aspekte näher beleuchtet. Daher sollten sich die Vertragspartner über die Art des Vertrages im Klaren sein. Es ist ratsam, die beabsichtigte Vertragsart im Vertrag festzuhalten. Rechtsgebend ist jedoch weniger die festgehaltene, beabsichtigte Vertragsart als die tatsächliche Natur des Vertrages, welche durch die vereinbarten Leistungen bestimmt wird.

Verträge können danach unterschieden werden, ob sie individuell zwischen den Parteien ausgehandelt werden oder ob es sich um einen wiederverwendeten Vertrag, sog. Allgemeine Geschäftsbedingungen (AGB) handelt. Dies ist wichtig, da bestimmte Klauseln in AGBs unzulässig sind, in einem Individualvertrag jedoch gelten. Für den betrachteten Fall, dass seitens eines Unternehmens eine Absicht besteht bestimmte Daten in die *Public Cloud* auszulagern, ist eher die AGB-Natur des Vertrages anzunehmen. Der Vertrag wird nicht zwischen zwei Parteien speziell für ihren Fall ausgehandelt. Stattdessen bietet der CSP seinen Dienst mit den gleichen Klauseln bzw. AGB für alle Nutzer an. Dies ist auch der Fall für Amazon S3, Google Storage und Rackspace Cloud Files.

Für die weitere Unterscheidung des Vertrages muss das anzuwendende Recht zu Grunde gelegt werden weil je nach Recht verschiedene Vertragsarten definiert sein können. So kennt das deutsche BGB z.B. u.a. den Mietvertrag (§§535 ff.), den Werkvertrag (§§631 ff.) und den Dienstvertrag (§§611 ff.). Ein Vertrag über die Nutzung von *Cloud Storage* hat nach herrschender Meinung in Deutschland in der Regel Mietcharakter und es kommen die allgemeinen Regelungen des BGBs zum Mietrecht zur Anwendung[8]. Teilweise wird ein solcher Vertrag aber auch als Werkvertrag nach §631 BGB ausgelegt. Juristen unterscheiden dabei, ob das primäre Ziel des Vertrages die Überlassung des Speichers (§535 BGB, Mietvertrag) oder die Erreichbarkeit der Daten über das Internet (§631 BGB, Werkvertrag) ist. Ein Gerichtsurteil, welches den Charakter von Verträgen über *Cloud Storage* eindeutig klärt, existiert noch nicht.

Die Rechte und Pflichten der Parteien in den verschiedenen Vertragsarten nach Gesetz sind sehr unterschiedlich. Aufgrund der ungeklärten Frage zum Charakter von Verträgen über *Cloud Storage* sollten sich die Vertragspartner jedoch nicht auf die gesetzlichen Regelungen verlassen, sondern alle Anforderungen, Rechte, Pflichten und Ansprüche explizit im Vertrag selbst festhalten. Dies kann z.B. über vertraglich vereinbarte *Service Level Agreements* (SLA) geschehen. Diese sollten die zugesicherten quantitativen und qualitativen Eigenschaften des Vetragsgegenstandes in einer nachprüfbaren Art und Weise festhalten (z.B. Aussagen über die

[8] http://www.heise.de/resale/artikel/Das-Recht-in-der-Cloud-1193203.html

Verfügbarkeit in einem bestimmten Zeitraum, Wartungsintervalle, Reaktionszeit, usw.).

Vertragsstrafen

In dem Vertrag bzw. den SLAs sollten unter anderem auch Art und Umfang der Strafen bei Verletzung des Vertrages festgehalten werden. In dem betrachteten Fall betrifft das vor allem die in den AGB festgelegten Strafen für den CSP, falls dieser die Zusicherungen aus den SLAs nicht erfüllen kann. Solche Strafen finden sich auch in den Bedingungen von Amazon S3[9], Google Storage[10] und Rackspace Cloud Files[11]. Alle drei schließen jedoch die Haftung für Folgeschäden, die aufgrund einer SLA-Verletzung entstehen, aus. Stattdessen erstatten die CSP bei SLA-Verletzung nur die Nutzungsgebühr für den Monat zurück, in dem die Verletzung aufgetreten ist. Tatsächlich kann jedoch der Schaden, der dem Nutzer durch die Verletzung einer zugesicherten Leistung entsteht, viel größer sein als die angebotene Entschädigung (z.B. wenn es durch die SLA-Verletzung zu einem unwiederbringlichen Datenverlust kommt, siehe auch Kapitel 3).

Der Nutzer muss hier entscheiden, ob die im Vertrag festgelegten Strafen ausreichend sind, um den CSP zur Einhaltung der vereinbarten SLAs zu motivieren. Hierzu ist anzumerken, dass zusätzlich zur Strafe, die dem jeweiligen Kunden ausgezahlt werden muss, mit jedem Vertragsbruch auch ein Imageverlust für den CSP einhergeht. Für den Nutzer stellt sich außerdem die Frage, inwieweit das Schadensmodell von der Alternative im eigenen Datenzentrum abweicht. Für den Fall, dass Daten nicht wiederherstellbar verloren gehen, be- kommt ein Nutzer auch beim eigenen Datenzentrum keinen Schadensersatz für Folgeschäden. Es stellt sich vielmehr die Frage, ob der CSP oder das eigene Personal im eigenen Daten- zentrum eher dafür geeignet ist, dem Fehlerfall vorzubeugen und für Fehlertoleranz zu sorgen (zur Verfügbarkeit siehe auch Kapitel 3).

Vertragsende

Bei einem Vertragsabschluss sollte auch berücksichtigt werden, was nach Vertragsende mit den Daten der Nutzer passiert. Cloud Services basieren normalerweise auf der monatlichen Abrechnung der genutzten Ressourcen und somit stellt das Ende der Ressourcennutzung das gewohnte Ende des Vertrages dar. In diesem Fall dürften sich beide Parteien über das Ende des Vertrages einig sein. Interessanter sind jedoch die Fälle, bei denen sich die Parteien uneinig über die Bewertung der Situation sind, z.B. weil der Nutzer seine Daten weiterhin bei dem CSP speichern möchte, dieser ihm aber wegen eines (aus Sicht des Anbieters begangenen) Verstoßes gegen die AGB kündigt. Alle BSP haben eine Reihe von Bedingungen, denen Nutzer des Dienstes zustimmen müssen. Hierzu gehören *Service Terms,*

[9] https://aws.amazon.com/de/s3- sla/

[10] https://developers.google.com/storage/docs/sla

[11] http://www.rackspace.com/cloud/legal/cloudfilessla/

Customer Agreement und eine *Acceptable Use Policy*. Letztere legt die Art der Daten fest, welche gespeichert bzw. nicht gespeichert werden dürfen. Die Einschränkungen sind nicht in jedem Fall klar formuliert, was die Einschätzung eines Verstoßes verkompliziert. Das grundlegende Problem dieser Art Bedingungen ist jedoch ein anderes: Die CSP behalten sich in der Regel das Recht vor, nach eigenem Ermessen über die Natur der Verstöße zu entscheiden. Das heißt, Dienstanbieter entscheiden in erster Instanz selbst über die Einhaltung ihrer AGB. Ist der CSP der Ansicht, dass ein Verstoß vorliegt, können sie den Vertrag aus ihrer Sicht beenden und beispielsweise direkt den Zugriff auf die API sperren oder den betroffenen Nutzeraccount löschen[12].

Gerade für den Fall des für den Nutzer unvorhergesehenen Vertragsendes ist es notwendig, für wichtige Daten Vereinbarungen zu treffen, was mit den Daten bei Vertragsende geschieht. Für den Fall, dass der Nutzer nicht auf das Ende des Vertrages vorbereitet ist, sollte eine Regelung vorhanden sein, welche dem Nutzer zumindest die Möglichkeit einräumt, wieder in den Besitz der Daten zu gelangen. In der Regel muss festgehalten werden, dass sich der CSP verpflichtet, die Daten korrekt und vollständig an den Kunden zu übermitteln und anschließend die Daten in dem *Cloud Storage* zu löschen.

Auswirkendes Recht

Unternehmen, welche ihre Daten bei einem CSP speichern wollen, sind von verschiedenen Gesetzen betroffen und unterliegen verschiedenen Rechten. Zunächst das Recht am Sitz des Unternehmens: Ein deutsches Unternehmen ist an deutsche Gesetze gebunden

Als nächstes das für den Vertrag ausgehandelte, anzuwendende Recht: Dieses kann frei gewählt werden, wird in der Regel aber das Recht von dem Sitz des CSP oder des Nutzers sein. Abgesehen davon, ist der Nutzer immer indirekt von dem Recht betroffen, welches am Sitz des CSP und an den Serverstandorten gilt.

Das an den Serverstandorten geltende Recht nimmt spätestens dann Einfluss auf den Dienst, wenn gegen einen Nutzer des Dienstes ermittelt wird. Angenommen gegen einen Nutzer eines CSP läuft ein Ermittlungsverfahren und im Zuge dieses Ermittlungsverfahrens sollen die Daten, welche der Nutzer in der Cloud gespeichert hat, untersucht werden. In diesem Zusammenhang stellt sich die Frage, wie die ermittelnden Behörden an die jeweiligen Daten gelangen können. Ressourcen in der Cloud sind virtualisiert, d.h. bei einer physischen Beschlagnahmung des Servers sind nicht nur die Daten des verdächtigen Nutzers betroffen. Auch wenn nur die Server mit den Daten des tatverdächtigen Nutzers untersucht werden sollen, sind durch die logische, aber nicht zwingend physische Trennung auch Daten anderer Nutzer involviert.

Zusätzlich wirken für den Nutzer auch die Landesgesetze, denen der BSP unterliegt — selbst wenn diese gar nicht im Vertrag genannt werden. Da die betrachteten

[12] https://developers.google.com/storage/docs/policy

Dienstleister ihren Sitz in den USA haben, gilt für sie auch der *Patriot Act.* Der *Patriot Act* verpflichtet Unternehmen dazu, Behörden auf Anfrage Zugriff zu ihren Daten zu verschaffen. Zugriff verschaffen bedeutet in diesem Fall, nicht nur die Herausgabe von Daten, die in den USA gespeichert sind. Das Gesetz umfasst auch die Schaffung einer Möglichkeit zur Einsichtnahme in Daten, welche nicht physisch in den USA gespeichert sind (z.B. bei einem Tochterunternehmen in Europa). CSPs mit Sitz in den USA können nach eigenen Angaben ihren Nutzern nicht garantieren, dass amerikanische Behörden keine Einsicht in gespeicherten Daten nehmen werden [2]. Nutzer müssen dabei nicht zwingend über die Preisgabe der Daten unterrichtet werden.

3. Erfüllung von IT-Sicherheitszielen

In diesem Kapitel soll für einige ausgewählte IT-Sicherheitsziele dargelegt werden, inwieweit sie von CSPs gewährleistet werden und inwiefern der Nutzer selbst zum Schutz der Daten beitragen kann. Es soll für die betrachteten BSPs (Amazon S3, Google Storage, Rackspace Cloud Files) hervorgehoben werden, wo Defizite bei der Erfüllung der Sicherheitsziele vorliegen, besonders im Vergleich zum eigenen Datenzentrum.

Verfügbarkeit

Verfügbarkeit im Falle von *Cloud Storage* bedeutet die Abrufbarkeit der ausgelagerten Daten. Für Nutzer von *Cloud Storage* hängt die Verfügbarkeit der Daten, abgesehen von der Verfügbarkeit der Internetverbindung, von der Verfügbarkeit des CSP ab. Ein Ausfall beim CSP, z.B. aufgrund eines Stromausfalls, Hardwareschäden oder Naturkatastrophen, kann mit der Nicht-Verfügbarkeit beim Nutzer des Dienstes einhergehen. Es kann dabei zwischen temporären und dauerhaften Ausfällen unterschieden werden. Bei temporären Ausfällen kann die Verfügbarkeit der Daten z.B. durch Wiederherstellen der Stromverbindung wiedererlangt werden, während es bei dauerhaften Ausfällen (z.B. durch Naturkatstrophen) zum unwiederbringlichen Datenverlust und damit zum unwiederbringlichen Verfügbarkeitsverlust kommt. Für den Nutzer stellt sich zunächst die Frage, worin er seinen Anspruch auf Verfügbarkeit rechtlich begründen kann. In dem Fall, dass deutsches Recht für die Auslegung des Vertrages zur Anwendung kommt (siehe Kapitel 2), ist dieser Anspruch je nach Art des Vertrages entweder in §535 BGB (Mietvertrag) oder §631 BGB (Werkvertrag) grundlegend definiert. Die hier betrachteten BSPs schließen jedoch die Anwendung des Deutschen Rechts aus. Daher muss eine andere Anspruchsgrundlage gefunden werden. Diese findet sich z.B. in den schon erwähnten *Service Level Agreements* (SLAs), welche Bestandteil des Vertrages mit dem CSP sind.

Die hier betrachteten BSPs sichern in ihren SLAs dem Nutzer einen bestimmten Verfügbarkeitsgrad zu. Das heißt in den SLAs von Amazon S3, Google Storage und Rackspace Cloud Files finden sich quantitative Angaben über die von dem Dienst versprochene Verfügbarkeit der Daten. Alle drei BSP sichern eine Verfügbarkeit von

99,9% zu[13] [14]. Dies entspricht einer erlaubten bzw. vertragskonformen Ausfallzeit von ca. 44 Minuten je Monat. Dabei sollte beachtet werden, dass die Anbieter tatsächlich verschiedene Verfügbarkeiten zusichern, auch wenn sie sich alle in der gleichen Prozentangabe niederschlagen. Ursache hierfür ist die Berechnungsgrundlage, welche für die Errechnung des Verfügbarkeitswertes verwendet werden soll. Allgemein ist Verfügbarkeit durch die Gleichung aus Formel 1 definiert. Für Cloud Storage lässt sich dies durch Formel 2 definieren.

$$\frac{\text{uptime}}{\text{uptime} + \text{downtime}}$$

Formel 1: Allgemeine Verfügbarkeit.

$$\frac{\text{Anzahl erfolgreicher Anfragen}}{\text{Anzahl aller Anfragen}}$$

Formel 2: Verfügbarkeit für CSPs.

Amazon S3 benutzt genau diese Definition[15]. Bei Rackspace Cloud Files hingegen werden Anfragen nur dann als nicht erfolgreich angesehen, wenn es sich um fehlgeschlagene Anfragen innerhalb von zwei aufeinanderfolgenden 90-Sekunden Intervallen handelt. Bei Google Storage handelt es sich nur um einen Ausfall, wenn innerhalb eines Intervalls mindestens 5% aller Anfragen fehlschlagen. Hier benutzt Amazon S3 zum Einen die transparenteste Berechnungsgrundlage und zum Anderen sichert Amazon auch die tatsächlich höchste Verfügbarkeit zu. Die anderen beiden Anbieter erlauben sich aufgrund ihrer Berechnungsgrundlage kürzere Ausfälle, ohne dass diese als solche gezählt werden.

Im Hinterkopf sollte dabei behalten werden, dass die CSPs auch bei einer Verletzung der garantierten Verfügbarkeit keine Verantwortung für die durch Ausfälle entstandene Folgeschäden übernehmen. Stattdessen wird im Allgemeinfall nur die Nutzungsgebühr anteilig für den Zeitraum der SLA-Verletzung zurück erstattet (siehe Kapitel 2). Für Datenkorruption bzw. Datenverlust wird ebenfalls nur für einen Monat die Nutzungsgebühr erstattet. Ab dem Folgemonat müsste für die Speicherung der korrupten Daten wieder voll bezahlt werden, sofern diese nicht gelöscht werden. Die niedrigen Strafen können durchaus abschreckend auf potentielle Nutzer wirken, da die tatsächliche Schaden wesentlich höher sein kann als die monatliche Nutzungsgebühr.

Vertraulichkeit

Vertraulichkeit als Sicherheitsziel bedeutet die Gewährleistung des Schutzes vor unbefugter Preisgabe von Informationen. Im Fall von *Cloud Storage* beinhaltet Vertraulichkeit erstens die Vertraulichkeit der Daten während der Übertragung zwischen Nutzer und CSP und zweitens die Vertraulichkeit der beim CSP gespeicherten Daten. Der letzte Punkt lässt sich aufgliedern in die Vertraulichkeit gegenüber unbefugten Dritten und die Vertraulichkeit gegenüber dem CSP selbst. Vertraulichkeit baut auf anderen Sicherheitszielen, namentlich Authentifizierung und

[13] https://developers.google.com/storage/docs/sla

[14] http://www.rackspace.com/cloud/legal/cloudfilessla/

[15] https://aws.amazon.com/de/s3- sla/

Autorisierung auf. Darüberhinaus ist Verschlüsselung das wichtigste Mittel zum Erreichen von Vertraulichkeit. SSL/TLS ist ein etabliertes Protokoll, um Vertraulichkeit während der Übertragung von Daten zu erreichen. Die Möglichkeit SSL/TLS-verschlüsselter Verbindungen zu nutzen, ist bei allen hier betrachteten BSP gegeben. Für den Authentifizierungsprozess wird beim Übertragen sensibler Nutzerdaten diese verschlüsselte Art der Verbindung sogar erzwungen. Die Verwendung von SSL/TLS ist für CSP leider nicht selbstverständlich. Einige ASP verwenden statt des etablierten, offenen Protokolls eigene Protokolle, was als unsicher gilt [2].

Die Daten werden bei den BSP üblicherweise unverschlüsselt gespeichert. Vertraulichkeit gegenüber Dritten wird bei den BSP für den Online-Zugriff allgemein über Authentifizierung und Autorisierung erreicht, indem nur autorisierte und authentifizierte Personen Zugriff auf die Daten bekommen. Eine weitere Möglichkeit, Vertraulichkeit gegenüber Dritten zu erreichen, ist serverseitige Verschlüsselung. Hier werden die Daten bei der Ankunft von dem CSP verschlüsselt, werden verschlüsselt gespeichert und bei Anfrage des Nutzers vor dem Übertragen wieder entschlüsselt. Die Verschlüsselung kann dabei entweder mit einem Schlüssel geschehen, der nur dem CSP bekannt ist, oder mit einem Schlüssel, der von dem Nutzer zur Verfügung gestellt wurde und von dem CSP verwaltet wird. Die erste Variante wird von einigen ASP verwendet, die zweite wird von Amazon S3 angeboten.

Auch wenn die Vertraulichkeit gegenüber Dritten für den Online-Zugriff auf die Daten über Authentifizierung und Autorisierung ausreichend gegeben ist, müssen auch Offline-Zugriffe direkt auf die Hardware im Rechenzentrum des CSPs betrachtet werden. Die BSP sichern alle zu, nur einer begrenzten Anzahl an Personen Zutritt zu den Serverräumen zu gewähren. Die hierbei genutzten Schutzmechanismen können jedoch von dem Nutzer nicht überprüft werden. Dabei sind nicht nur die technischen Angaben relevant, sondern auch die beim BSP vorgesehenen Prozesse sowie deren Umsetzung (z.B. physischer Zugang zu den Serverräumen).

Ein weiteres Problem stellt in diesem Zusammenhang die Datenremanenz dar, denn die Befugnis, Informationen zu speichern, kann von dem Besitzer der Daten auch wieder entzogen werden und es stellt sich die Frage, wie dies im Falle von *Cloud Storage* umgesetzt ist. Zum Beispiel könnte ein Nutzer versehentlich eine sehr sensible, vertrauliche Datei an einen *Cloud Storage* übertragen. In einem solchen Fall kann der Nutzer die Löschung der Datei beim CSP beantragen (d.h. eine „Lösch"-Anfrage ausführen), doch es ist unklar, ob, wann und wie sicher die Datei tatsächlich gelöscht wird. Die hier betrachteten BSP machen alle Aussagen darüber, was bei dem „Löschen" einer bei ihnen gespeicherten Datei nach außen hin passiert: Je nach implementiertem Datenkonsistenzmodell existiert nach wenigen Sekunden (abschließende Konsistenz) oder sofort (Read-after-Write-Konsistenz) kein Fernzugriff mehr auf die Datei. Es existiert jedoch keine Aussage darüber, wann die Daten tatsächlich physikalisch auf dem Speichermedium des CSPs gelöscht werden. Auch wenn angenommen wird, dass die Daten sofort von den eingesetzten Server-Festplatten beim BSP gelöscht werden, ist fraglich, ob der BSP nicht immer noch über

eine Kopie der Daten verfügt (z.B. in Form von internen Backups, welche durchaus geografisch verteilt sein können). Letztendlich wären die Aussagen, welche von den BSP getroffen werden, für den Nutzer auch nicht nachprüfbar. Somit muss der Nutzer dem CSP entweder vertrauen oder er verschlüsselt seine Daten in der Form, dass sie für den Cloud-Storage-Anbieter nicht auswertbar sind (Vertraulichkeit gegenüber dem CSP).

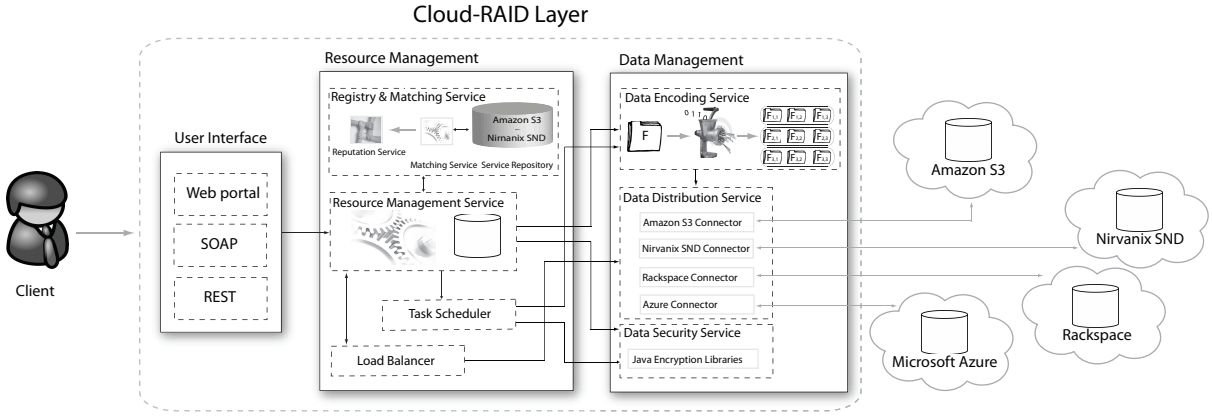

Abbildung 1: Grundlegende Architektur und Arbeitsweise der Cloud-RAID Lösung.

4. Lösungsansatz

Bisher wurde ein Überblick über den allgemeinen rechtlichen Kontext bei der Nutzung von *Cloud Storage* sowie die Erfüllung der IT-Sicherheitsziele Verfügbarkeit und Vertraulichkeit sowie Datenschutz gegeben. Zusammenfassend kann gesagt werden, dass die Auswahl eines externen Anbieters sehr sorgfältig vorgenommen werden sollte. Bei bestimmen Daten müssen dabei gesetzliche Bestimmungen beachtet werden (z.B. in Bezug auf den Schutz oder den geographischen Standort der Aufbewahrung). Darüberhinaus reichen die Zusicherungen öffentlicher Cloud-Speicher-Anbieter in Bezug auf die Verfügbarkeit und die Vertraulichkeit zur Aufbewahrung geschäftskritischer Daten nicht aus. Für einen solchen Einsatz sollte die Kontrolle über die Sicherheit sowie Verfügbarkeit der auszulagernden Inhalte stets beim Besitzer der Daten bleiben. In diesem Kapitel wird eine Anwendung vorgestellt, welche im Rahmen eines laufenden Forschungsprojektes die diskutierten Herausforderungen zu lösen versucht. Details zur Implementierung können unseren früheren Publikationen entnommen werden [8, 9, 10, 11]. In diesem Abschnitt wird lediglich die Funktionsweise der wichtigsten Komponenten vorgestellt.

Unser Ansatz verfolgt die Idee, bei Nutzung von Cloud-Speicherressourcen die Anwenderdaten nicht einem einzigen Anbieter anzuvertrauen sondern gleichmäßig auf mehrere - unter Einhaltung nutzerspezifischer Anforderungen - 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 (siehe Abbildung 1). 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 gleiche Prinzip für Cloud-Speicher-Ressourcen ein. Zur Aufteilung sowie Rekonstruktion der Daten werden Erasure Codierung Techniken eingesetzt. Der große Vorteil bei diesem Vorgehen liegt in dem günstigen Verhältnis des Speicherbedarfs relativ zur dazu gewonnenen Datenverfügbarkeit. Der Ansatz ermöglicht unserer Plattform, den Ausfall von einem oder sogar mehreren Online-Speicherdiensten zu tolerieren, ohne Daten der Anwender zu verlieren. Die Anzahl der zu tolerierenden Ausfälle hängt dabei von den Anforderungen der Nutzer an die Verfügbarkeit der Daten ab.

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 die verfügbaren technischen Features. Damit können Nutzer ihre Daten verwalten sowie Anforderungen 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 Anforderungen 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.

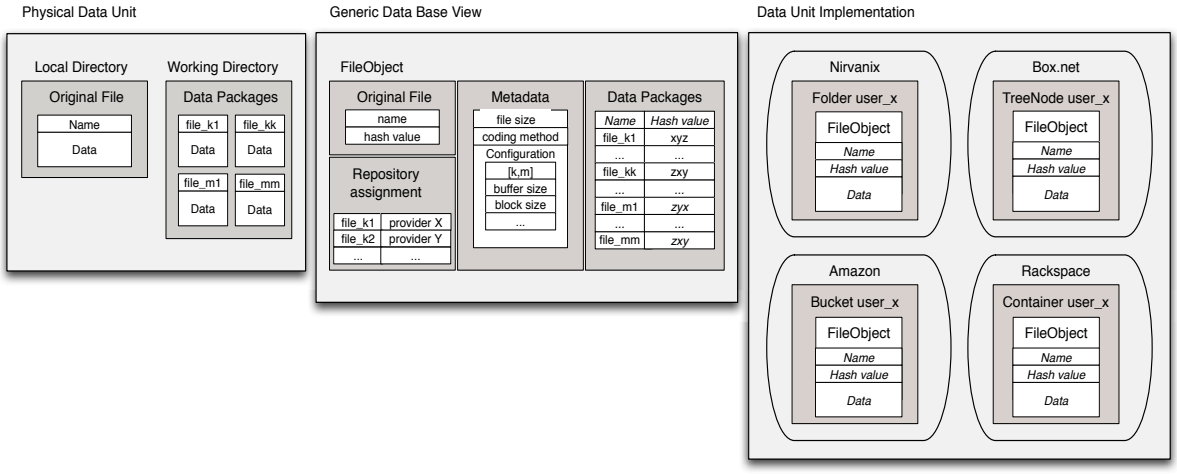

Abbildung 2: Das Datenmodell des Cloud-RAID-Systems.

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:

- *Encoding Service*: ist für die Kodierung und Dekodierung einzelner Datenobjekte mittels Erasure-Algorithmen verantwortlich. Hierbei wird ein Datenobjekt in $n=k+m$ Blöcke kodiert (Datenfragmente). Die Zahl n bezeichnet dabei Anzahl der Datenblöcke, welche nach dem Encoding entstehen und anschließend auf verschiedene, disjunkte Speicherquellen verteilt werden. Der Vorteil bei dieser Form der Kodierung liegt darin, dass jeder Datenblock mit beliebigen m der n Fragmente zu rekonstruieren ist [12]. Die Parameter m und k sind dabei frei wählbar. Der zusätzliche Speicherbedarf ergibt sich aus dem Verhältnis m/k. Bei Verteilung der Daten auf 11 Anbieter (mit einer exemplarischen Erasure-Konfiguration [10,1]) wäre das System in der Lage den Verlust von einer Quelle zu tolerieren. Der zusätzliche Speicherbedarf, der zusätzliche Kosten nach sich zieht, würde dabei um lediglich 10 Prozent ansteigen. Bei naiver Replikation steigen dagegen die Kosten proportional zur Anzahl verteilter Kopien. Aktuell unterstützt das System nur 12 Dienstleister, wobei die Ausdehnung der Funktionalität auf weitere Anbieter geplant ist. An dieser Stelle kann angemerkt werden, dass die eigentliche Kodierung der Datenobjekte sehr performant erfolgt, so dass die dazu benötigte Zeit bei dem gesamten Übertragungsprozess vernachlässigt werden kann (siehe Abbildung 3 und Abbildung 4). Der eigentliche Kodierungsschritt nimmt in dem Beispiel, der in den Abbildungen gezeigt wird, weniger als 0,5 Prozent der gesamten Datenübertragung in Anspruch. Die einzelnen Datenpakete wurden dabei an folgende Anbieter übertragen: Google US, Amazon EU, Amazon (US-west-1), Nirvanix, Azure und Google EU.

- *Datenverteilungsservice*: verteilt separate Fragmente auf verschiedene Speicheranbieter. Im Wesentlichen ähnelt das Datenmodell unseres Systems (siehe Abbildung 2) dem des Amazon S3. Alle Datenobjekte werden in so genannten "Buckets" (Behälter) abgelegt. Jeder dieser Behälter 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 insgesamt *n* Speicherbehälter eingerichtet. Diese repräsentieren Buckets, die später auf verschiedene Cloud-Storage Anbieter verteilt werden. 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. Der Adapter unterstützt somit die grundlegende Funktionalität zur Datenmodifikation (*put, get, delete, list:* 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.

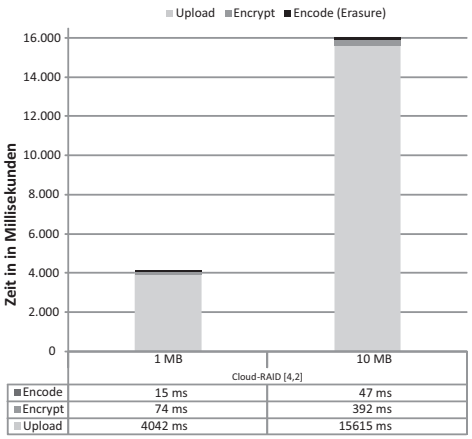

Abbildung 3: Vergleich Rechenaufwand für die Kodierung und Verschlüsslung und Übertragung von Datenobjekten mit Cloud-RAID.

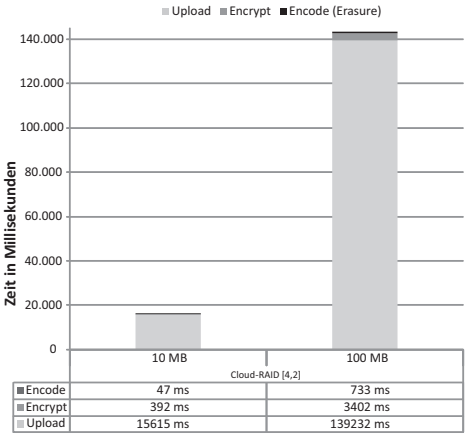

Abbildung 4: Vergleich Rechenaufwand für die Kodierung und Verschlüsslung und Übertragung von Datenobjekten mit Cloud-RAID.

- *Sicherheitsservice*: ist für die Durchsetzung festgelegter Sicherheitsrichtlinien verantwortlich. Das initiale Sicherheitsniveau wird durch die physische Trennung der Inhalte gewährleistet, indem die ursprünglichen Datenobjekte auf unterschiedliche Cloud-Provider verteilt werden. Der Einsatz der Erasure-Algorithmen schließt das Durchsickern wichtiger Informationen nicht aus. Besonders wichtige Inhalte sollten daher verschlüsselt aufbewahrt werden. Um dies zu ermöglichen, werden von dem System die Inhalte vor der Übertragung an Cloud-Anbieter verschlüsselt. Abbildungen Abbildung 3 und Abbildung 4

zeigen deutlich, dass auch die Verschlüsselung der Daten in unserer Anwendung im Vergleich zu der eigentlichen Datenübertragung kaum ins Gewicht fällt. Daher sieht die Standardeinstellung vor, dass Daten vor der Übertragung stets verschlüsselt werden.

5. Auswertung

Zur Einschätzung der Leistungsfähigkeit des Forschungsansatzes wurde am Hasso Plattner Institut (HPI) im Juli 2012 ein umfassender Test der entwickelten Software durchgeführt. Aufgrund des begrenzten Platzes, der der aktuellen Publikation zur Verfügung steht, werden in diesem Kapitel nur die ausgewählten Ergebnisse des durchgeführten Experiments vorgestellt. Der Test wurde insgesamt in einem Zeitraum von mehr als 336 Stunden (zwei Wochen) ununterbrochen ausgeführt (7x24). Das Testbed umfasste dabei sieben Cloud-Speicher-Anbieter mit verteilten Infrastrukturen, so dass insgesamt 12 voneinander getrennte „Container“ für den Cloud-Speicher bereitgestellt werden konnten. Da das HPI über eine Breitbandanbindung an das Internet verfügt (1Gb), konnte sichergestellt werden, dass die Testsysteme selbst zu keinem Zeitpunkt einen Engpass während des Versuchs erfahren müssen.

Beobachtungen und Ergebnisse

Wie in dem Kapitel 4 bereits beschrieben wurde, sieht die Implementierung des Cloud-RAID Systems eine Fragmentierung (als Ergebnis der Kodierung) und anschließende Verteilung einzelner Datenpakete auf mehrere Cloud-Speicher-Anbieter vor. Die Gesamtdauer der Datenübertragung hängt von der Durchsatzfähigkeit einzelner Anbieter ab, die in den jeweiligen Datenübertragungsprozess miteinbezogen werden. In unserem Test wurden alle möglichen Konstellationen der Anbieter berücksichtigt, die aktuell von unserem System unterstützt werden.

Grundsätzlich kann gesagt werden, dass der Einsatz von Erasure-Algorithmen die Leistungsfähigkeit der Datenübertragung gegenüber einer nativen Datenübertragung der gleichen Datei an einzelne Dienste deutlich erhöhen kann. Bei einer genaueren Betrachtung der Ergebnisse fällt auf, dass einige Anbieterkonstellationen besser abschneiden als andere. Eine Auswahl getesteter Konstellationen ist in einer Übersichtstabelle (Tabelle 1) auf Seite 18 aufgeführt. Dieses Verhalten lässt sich damit erklären, dass die im Rahmen des Versuchs betrachteten Anbieter die Schnittstellen der Dienste für bestimmte Anwendungsfälle optimiert haben. Im Folgenden soll dieses Verhalten verdeutlicht werden.

In erster Linie fällt ein großer Unterschied in der Performance zwischen dem Up- und Download bei einzelnen Anbietern auf. Nahezu alle getesteten Dienste zeigen im Download deutlich bessere Durchsatzwerte gegenüber dem Upload. Dies gilt sowohl für die Übermittlung kleinerer als auch größerer Dateien. So kann beispielsweise der Schreibvorgang einer 100kB Datei bei dem Dienst Google US oder Google EU ca. 12 bis 19 mal länger dauern als beim lesenden Zugriff (vergl. Abbildung 5 und Abbildung 6). Das Verhalten lässt sich auch bei Übertragung größerer Dateien beobachten, wenn auch nicht mehr mit dermaßen drastischen Schwankungen. Hier unterscheidet sich die

Performance um das vier- bis fünffache mit Ausnahme von dem Anbieter Rackspace, bei dem ein Schreibzugriff bis zu 19 bzw. 49 mal langsamer ausfallen kann als ein Lesezugriff (vergl. **Abbildung 10**, **Abbildung 11** und **Abbildung 12**).

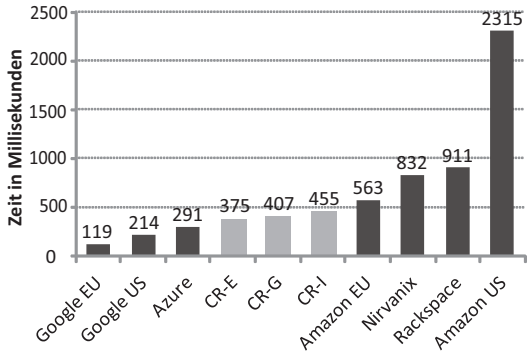

Abbildung 5: Download einer 100kB Datei.

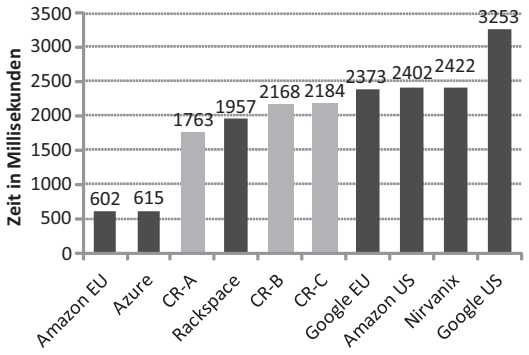

Abbildung 6: Upload einer 100kB Datei.

Andererseits konnte auch ein Zusammenhang zwischen der Übertragungsleistung und der Datengröße beobachtet werden. So erzielen bei der Übertragung kleinerer Dateien (bis 1 MB) die Anbieter Azure und Amazon EU (beim Upload) die besten Ergebnisse. Bei 10 MB fällt der Anbieter Amazon EU bereits zurück (vergl. Abbildung 6 und **Abbildung 12**). Ähnlich verhält sich auch der Dienst Google US. Allerdings verbessern sich hier die Performancewerte (deutlich) mit wachsender Dateigröße. Das beobachtete Verhalten kann damit zusammenhängen, dass die relativ große Reaktionszeit des Dienstes (welche in erster Linie auf die lange Entfernung zwischen unserem Testsystem und dem Zielknoten zurückzuführen ist) bei größeren Dateien immer weniger die Messung der Übertragungsdauer beeinflusst.

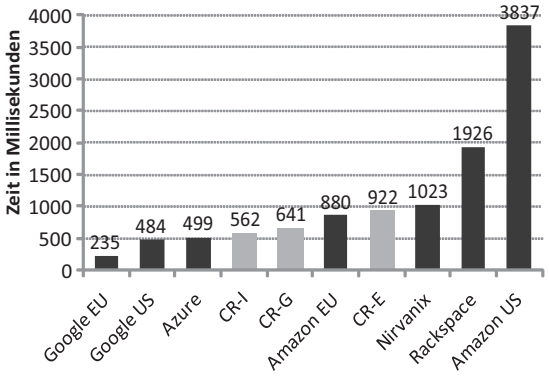

Abbildung 7: Download einer 500kB Datei.

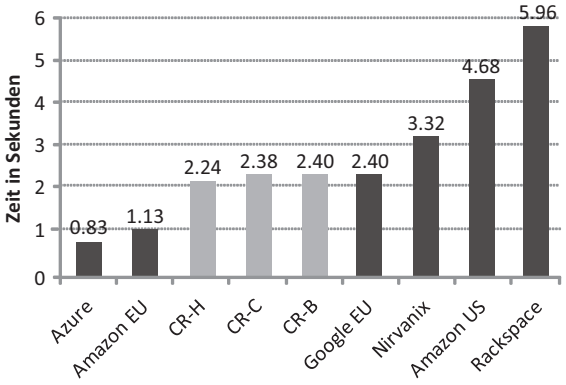

Abbildung 8: Upload einer 500kB Datei.

Noch deutlicher lässt sich jedoch dieses Verhalten bei dem Dienst Google EU beobachten. Bei der Übertragung kleinerer Dateien zeigt der Dienst eine relativ schlechte Performance im direkten Vergleich zu anderen Anbietern. Mit wachsender Dateigröße rückt der Dienst jedoch immer weiter nach Vorne, bis er schließlich bei der Übertragung von 100 MB Dateien die Spitzenposition erreicht und damit den Dienst Azure ablöst (vergl. Abbildungen Abbildung 6, **Abbildung 8**, **Abbildung 10** und **Abbildung 12**).

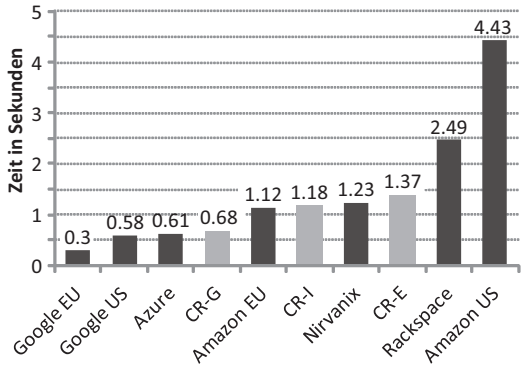

Abbildung 9: Download einer 1 MB Datei

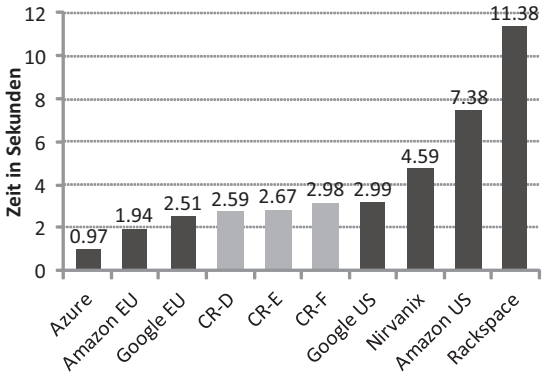

Abbildung 10: Upload einer 1MB Datei.

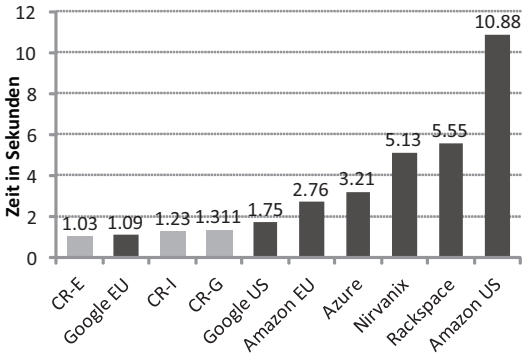

Abbildung 11: Download einer 10 MB Datei

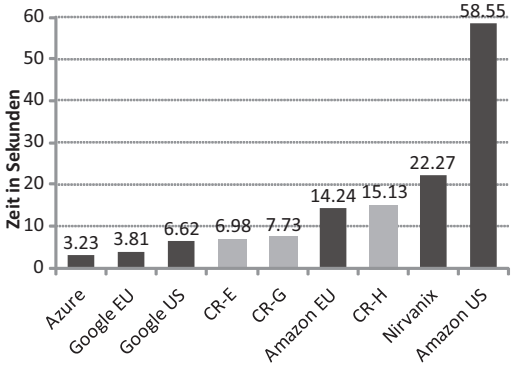

Abbildung 12: Upload einer 10 MB Datei.

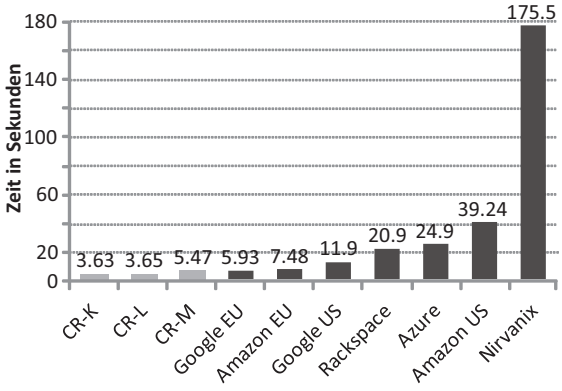

Abbildung 13: Download einer 100 MB Datei.

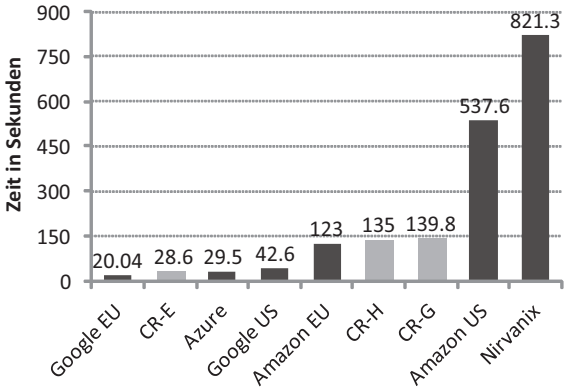

Abbildung 14: Download einer 100 MB Datei.

Ähnliche Zusammenhänge lassen sich auch beim Download beobachten. Bei der Übertragung kleinerer Dateien, gehört der Anbieter Azure zu den führenden Diensten. Mit wachsender Dateigröße fällt Azure jedoch immer weiter zurück (vergl. Abbildung 5 und Abbildung 13). Auch der Dienst Rackspace zeigt bei kleineren Dateien wesentlich bessere Performance. Bei der Übertragung von Dateien mit einer Größe

von 100kB gehört der Dienst noch zu den führenden drei Anbietern. Bereits bei einer Dateigröße von 500kB fällt der Dienst auf den sechsten Platz zurück.

Cloud-RAID Konfiguration [4,1]	Konstellation der beteiligten Anbieter
CR-A	Amazon EU, Amazon US, Azure, Nirvanix, Rackspace
CR-B	Amazon EU, Amazon US, Azure, Google EU, Rackspace
CR-C	Amazon US, Azure, Google EU, Nirvanix, Rackspace
CR-D	Amazon EU, Amazon US, Azure, Google EU, Nirvanix
CR-E	Amazon EU, Azure, Google EU, Google US, Nirvanix
CR-F	Amazon EU, Google EU, Google US, Nirvanix, Rackspace
CR-G	Amazon EU, Amazon US, Azure, Google EU, Google US
CR-H	Amazon EU, Amazon US, Google EU, Google US, Nirvanix
CR-I	Amazon EU, Azure, Google EU, Google US, Rackspace
CR-K	Amazon EU, BoxNet, Google EU, Google US, Nirvanix
CR-L	Amazon EU, Amazon US, BoxNet, Google EU, Google US
CR-M	Amazon EU, Amazon US, Azure, BoxNet, Google EU

Tabelle 1: Verwendete Anbieterkonstellationen.

Diskussion

In dieser Arbeit haben wir ein System vorgestellt, welches eine Metaebene zwischen Anwendern und Anbietern von Cloud-Speicherressourcen bereitstellt. Bei der Übertragung der Daten werden die Datensätze der Anwender mittels Erasure Codes fragmentiert und auf verschiedene, von einander unabhängige 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 (aller Datenfragmente) ist. Zur Wiederherstellung der Originaldaten ist dabei nur ein Teil der Datenblöcke notwendig. Das Vorgehen erhöht die Zuverlässigkeit bei externer Datenlagerung, reduziert das Lock-in Risiko sowie auch die Gefahr eines möglichen Datenmissbrauchs seitens der Dienstleister. Die Ergebnisse des durchgeführten Experiments zeigen deutlich, dass der Einsatz von Erasure Codes ein gutes Verhältnis zwischen Wirtschaftlichkeit [13] und Effizienz für externe Datenlagerung bereitstellt. Darüber hinaus abstrahiert der Ansatz die technische Komplexität im Umgang mit den proprietären Schnittstellen verschiedener Dienstanbieter und ermöglicht seinen Nutzern einen einfachen Zugriff auf Cloud-Speicherressourcen.

Literaturhinweise

[1] P. Mell and T. Grance, “The NIST definition of cloud computing,” in*National Institute of Standards and Technology, Information Technology Laboratory,* 2009.

[2] M. Borgmann, T. Hahn, M. Herfert, T. Kunz, M. Richter, U. Viebeg, and S. Vowé, “On the security of cloud storage services,” Tech. Rep., 2012.

[3] M. Chung and J. Hermans, “From hype to future. kpmg’s 2010 cloud computing survey,” 2010.

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

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

[6] 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.

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

[8] Maxim Schnjakin, Rehab Alnemr and Christoph Meinel. A security and high- availability layer for cloud storage. In Web Information Systems Engineering – WISE 2010 Workshops, volume 6724 of Lecture Notes in Computer Science, pages 449–462. Springer Berlin / Heidelberg, 2011.

[9] Maxim Schnjakin and Christoph Meinel. Implementation of Cloud-RAID: A Secure and Reliable Storage Above the Clouds. Proceedings of 8th International Conference on Grid and Pervasive Computing – GPC 2013, to appear in May 2013.

[10] Maxim Schnjakin, Dimitri Korsch, Martin Schoenberg, Christoph Meinel. Implementation of a Secure and Reliable Storage Above the Untrusted Clouds. Proceedings of 8th International Conference on Computer Science and Education – ICCSE 2013, to appear in April 2013.

[11] Maxim Schnjakin and Christoph Meinel. Platform for a secure storage- infrastructure in the cloud. Proceedings of the 12th Deutscher IT- Sicherheitskongress (Sicherheit 2011), 2011.

[12] J. S. Plank, S. Simmerman, and C. D. Schuman. Jerasure: A library in C/C++ facilitating erasure coding for storage applications - Version 1.2. Technical Report CS-08-627, University of Tennessee, August 2008.

[13] H. Weatherspoon and J. Kubiatowicz. Erasure coding vs. replication: A quantitative comparison. IPTPS, March 2002.