NTNU – Trondheim
Norwegian University of
Science and Technology

Securing Cloud Storage Service

Vytautas Zapolskas

Master in Security and Mobile Computing
Submission date: June 2012
Supervisor: Danilo Gligoroski, ITEM
Co-supervisor: Gerald Q. Maguire Jr., KTH
Fredrik Solsvik, Telenor ASA

Norwegian University of Science and Technology
Department of Telematics

Master Thesis Problem Description

Vytautas Zapolskas

15th February, 2012

As more companies turn to cloud solutions, securing cloud based services becomes increasingly important, because for many organizations, the final barrier to adopting Cloud computing is whether it is sufficiently secure.

This research is limited to data protection risks in cases of storing and transferring sensitive data between clouds. The student will design a service which could provide Security as a Service for cloud brokers and carriers in a federated cloud allowing customers to securely migrate from one provider to another. Such service would utilize various encryption techniques and also include identity and key management mechanisms, such as "federated identity management".

To support the design of the service the study will also
- identify most important Cloud Storage specific risks and compare them with traditional solutions, such as server-based model.
- describe data protection requirements for cloud storage services.

Supervisor: Professor Danilo Gligoroski
External supervisor: Fredrik Solsvik, Telenor ASA

Securing Cloud Storage Service

Master of Science Thesis

Vytautas Zapolskas

10^{th} June, 2012

Academic Adviser and examiner:
Prof. Danilo Gligoroski
NTNU Norwegian University of Science and Technology, Norway

Academic Adviser and examiner:
Prof. Gerald Q. Maguire Jr
KTH Royal Institute of Technology, Sweden

Supervisor:
Fredrik Solsvik
Telenor ASA, Norway

Vilnius, Lithuania

Abstract

Cloud computing brought flexibility, scalability, and capital cost savings to the IT industry. As more companies turn to cloud solutions, securing cloud based services becomes increasingly important, because for many organizations, the final barrier to adopting cloud computing is whether it is sufficiently secure.

More users rely on cloud storage as it is mainly because cloud storage is available to be used by multiple devices (e.g. smart phones, tablets, notebooks, etc.) at the same time. These services often offer adequate protection to user's private data. However, there were cases where user's private data was accessible to other user's, since this data is stored in a multi-tenant environment. These incidents reduce the trust of cloud storage service providers, hence there is a need to securely migrate data from one cloud storage provider to another.

This thesis proposes a design of a service for providing Security as a Service for cloud brokers in a federated cloud. This scheme allows customers to securely migrate from one provider to another. To enable the design of this scheme, possible security and privacy risks of a cloud storage service were analysed and identified. Moreover, in order to successfully protect private data, data protection requirements (for data retention, sanitization, and processing) were analysed. The proposed service scheme utilizes various encryption techniques and also includes identity and key management mechanisms, such as "federated identity management".

While our proposed design meets most of the defined security and privacy requirements, it is still unknown how to properly handle data sanitization, to meet data protection requirements, and provide users data recovery capabilities (backups, versioning, etc.).

Some thoughts on information technology security:

Acknowledgements

I am most grateful to my academic thesis adviser and examiner Professor Gerald Q. Maguire Jr. (School of Information and Communication Technologies, Royal Institute of Technology (KTH), Stockholm, Sweden) for sharing his valuable ideas, comments, and suggestions. Special thanks for his comprehensive and fast answers.

My sincere thanks are due to my thesis supervisor Fredrik Solsvik (Telenor ASA, Trondheim, Norway) for his critical reviews on my work, support, and constant encouragement. I would also like to thank him for introducing me to the topic.

I would like to thank Professor Danilo Gligoroski (Department of Telematics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway) for his suggestions and especially for recommending me to narrow the thesis and to limit the scope of my study.

I would also take this opportunity to thank my study co-ordinators Ms. May-Britt Eklund-Larsson (NordSecMob Co-ordinator, KTH) and Ms. Mona Nordaune (NordSecMob Co-ordinator, NTNU) for extensive assistance during my stay in Scandinavia.

Thank you all for your help and making these studies such a wonderful experience.

Contents

List of Figures

List of Tables

Nomenclature

AES	Advanced Encryption Standard
API	Application programming interface
BBC	British Broadcasting Corporation
CDMI	Cloud Data Management Interface
CDN	Content Delivery Network
CSA	Cloud Security Alliance
CSP	Cloud Service Provider
DaaS	Data Storage as a Service
EAA	European Economic Area
ENISA	European Network and Information Security Agency
EU	European Union
ext4	fourth extended filesystem
GB	Gigabyte
IaaS	Infrastructure as a Service
IDM	Identity Management
ISP	Internet Service Provider
LHC	Large Hadron Collider
NIST	National Institute of Standards and Technology
NTFS	New Technology File System
PaaS	Platform as a Service
PB	Petabyte

PBKDF2	Password-Based Key Derivation Function 2
PC	Personal Computer
QoS	Quality of Service
RFC	Request For Comments
RIPA	Regulation of Investigatory Powers Act
RSA	Rivest Shamir and Aldeman
S3	Amazon's Simple Storage Service
SaaS	Software as a Service
SAML	Security Assertion Markup Language
SeaaS	Security as a Service
SLA	Service-level Agreement
SNIA	Storage Networking Industry Association
SSL	Secure Socket Layer
SSO	Single Sign-On
TLS	Transport Layer Security
U.S.	United States (of America)
UID	Unique IDentification
URL	Uniform Resource Locator
USB	Universal Serial Bus
XaaS	Anything as a Service
XRDS	eXtensible Resource Descriptor Sequence
ZFS	the Z File System

Chapter 1

Introduction

Cloud computing is not a completely new computing model. The concept has been adapted from the earlier grid computing paradigm, and other distributed systems such as utility computing and cluster computing. In September 2011, the definition and specifications of cloud computing were standardized by the U.S. National Institute of Standards and Technology (NIST). The definition of Cloud Computing introduced by the NIST is:

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models." [1]

Cloud computing is available in three different offerings: cloud computing, cloud storage, and Anything as a Service (XaaS). More details about each of these types of offers can be found in section 2.1. This thesis will explore security in the context of cloud storage, focusing specifically on the problem of transferring or migrating data stored in one cloud to another cloud.

1.1 Motivation

Areas such as eGovernment, eMedicine, media, and telecommunications are becoming more dependant on data storage services. Emerging high quality digital media formats require larger storage facilities. In [3] British Broadcasting Corporation (BBC) announced that it is shifting to fully tapeless content production and investing in developing their private cloud ecosystem. In addition, projects such as the Large Hadron Collider (LHC) and Hubble telescope are generating huge amounts of data. For instance the LHC computing grid generates 13 Petabytes (PB) of data each year [4]. This recent growth in generated data creates a problem

- where to store all that data. Moreover, the emergence of digital devices (smart phones, tablets, etc.) generates additional data such as photos, videos, etc. which preferably has to be backed up even though that data is not often used. Thus data storage services are gaining a significant role in our lives.

Moving data storage services to the cloud has its advantages and disadvantages. Businesses and individuals can achieve capital cost savings and simplify their operations with cloud storage, since they do not have to invest in their own storage servers nor do they have to maintain them. Most importantly these cloud storage services are scalable; with a few simple clicks your storage capacity will be expanded. Technical support, hardware renewal, and upgrades are no longer the customers' responsibility. Although these features seem very tempting, many customers are not rushing to transfer their data into the cloud.

The final barrier to adopting cloud storage is usually whether it is sufficiently secure. Privacy issues, data leakage, improper data sanitization; all of these and a lot more are reasons why a customer will choose a particular cloud storage provider. Additionally, changes in the customer's requirements or in the provider's offering may lead the customer to migrate their storage to another cloud storage provider.

Providing a secure service which handles sensitive data transfers between cloud storage providers in a federated cloud (see section 2.4.1) is an interesting and new problem domain. This problem has motivated this thesis project. Data transfers between the customer and the provider and between providers must be handled securely. Achieving this security is the obvious objective, but this has to be done in the context of maintaining compliance with the customer's security policies and meeting various regulatory and legislative requirements.

1.2 Related work

The Cloud Security Alliance (CSA) in [5] has done a great job defining Security as a Service. The European Network and Information Security Agency (ENISA) in [6] described the benefits and risks of cloud computing. Cloud storage services, such as SpiderOak (see section 2.2.2.3), introduced a zero knowledge approach which was used in the thesis. In [47], Educause clearly described a federated identity management concept.

In [53], Basescu et al. proposed a generic security management framework allowing providers of cloud data management systems to define and enforce complex security policies. In [54], Chow et al. addressed the problem of building a secure cloud storage system which supports dynamic users and data provenance. In [55], Yang and Zhang proposed a generic scheme to enable fine-grained data sharing over the cloud, which does not require key-redistribution and data re-encryption.

1.3 Scope and objectives

A recent security flaw in the Dropbox authentication mechanism [7] started a discussion about whether cloud storage services are sufficiently secure to store sensitive data. In addition, Dropbox announced that it will allow government agents to access customers' data. This means that there is a backdoor mechanism to access data which might be exploited. For both of these reasons, a lot of customers are considering migrating from Dropbox to a different cloud storage provider. However, there are solutions where experienced users simply use Dropbox as a drive and encrypt their files before doing operations on a virtual disk provided as a file via Dropbox [8].

The goal of this thesis project is to design and evaluate a service which could provide security functions for cloud brokers and carriers in a federated cloud allowing customers to securely migrate from one provider to another. Such a service would utilize various encryption techniques, a zero knowledge approach, and also include identity and key management mechanisms, such as federated identity management.

In addition, to support the design of the service this thesis will also identify the most important cloud storage specific risks and compare them with traditional solutions, such as storage offered by a server-based model. Moreover, the thesis will describe data protection requirements for cloud storage services.

1.4 Limitations

This research is limited to data protection risks in the cases of storing and transferring sensitive data between cloud providers. Since compliance with security policies and regulatory and legislative requirements differ between countries, the thesis will concentrate on those requirements relevant to the EU.

Due to time and resource constraints, no proof of concept will be attempted. However, similar solutions will be analysed and discussed. In addition, recommendations and design modifications will be provided to assist future work.

1.5 Methodology

First this thesis will analyse EU data protection requirements and security policies. Next, based on this analysis an abstract design of security as a service architecture will be proposed. The thesis will include an analysis and discussion of the proposed design, identifying its potential and limitations (if any). Finally, similar existing solutions will be compared with the proposed design, in order to identify modifications and make recommendations of changes that would be needed to comply with the EU's data protection requirements.

1.6 Outline

This thesis is organized into the following chapters:

Chapter 2 - Background provides an overview of cloud computing architectural framework and introduces cloud storage as a service model. In addition, cloud brokerage and cloud carrier services are described.

Chapter 3 - Cloud storage risks review the most important cloud storage specific risks and compares them with traditional solutions, such as those implemented by a server-based model. This discussion will be limited to data protection requirements in EU.

Chapter 4 - Security as a Service introduces a design of a security architecture which could enable cloud brokers to secure cloud storage services. The proposed scheme will secure consumer's data at rest and while migrating from from one cloud provider to another. The designed security architecture will take into account the data protection requirements from chapter 3.

Chapter 5 - Results - this chapter analyses the proposed design and identifies potential limitations.

Chapter 6 - Discussion - this chapter continues the discussion of our proposed design and our findings.

Chapter 7 - Future work and conclusions suggests possible enhancements and additional features that are applicable to the proposed design.

Chapter 2

Background

This thesis will use terms such as cloud computing, cloud broker, cloud carrier, cloud storage, federation of clouds, etc. In order to introduce the reader to these terms, a short background chapter with explanations is provided. This chapter covers cloud computing architectural framework, cloud storage, cloud service brokers and carriers, and finally describes a cloud federation.

2.1 Cloud Computing Architectural Framework

Cloud computing introduces a technical change in a way IT resources are delivered and consumed. This section describes five essential characteristics, four deployment models, and three service models that are common to cloud computing. In addition, this section explains multi-tenancy and covers five roles in the cloud ecosystem.

2.1.1 Essential Characteristics

The cloud model is composed of five essential characteristics:

- ***On-demand self-service.*** A customer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider. [1]

- ***Broad network access.*** Network access is available over the network and controlled through standard mechanisms that promote access by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations). [1]

- ***Resource pooling.*** The provider's computing resources are pooled to serve multiple customers using a multiple-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to customer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided

resources, but may be able to specify the location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth. [1]

- ***Rapid elasticity.*** Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the customer, the capabilities available for provisioning often appear to be unlimited and can be requested in any quantity and at any time. [1]
- ***Measured service.*** Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and customer of the utilized service. [1]

2.1.2 Deployment models

Following are the four types of cloud deployment models identified by NIST [1].

- ***Public cloud.*** The cloud infrastructure is open for use by the general public (i.e., any member of the general public can subscribe to use a public cloud's service).
- ***Private cloud.*** The cloud infrastructure is maintained by the organization itself and is used exclusively by a single organization.
- ***Community cloud.*** The cloud infrastructure is maintained by one organization for a set of organizations (the community) and used by all of them.
- ***Hybrid cloud.*** The cloud infrastructure is a composition of two or more distinct cloud infrastructures.

Private clouds are deployed on premises and accessible only to a single organization. In contrast Public cloud is deployed off premises and is accessible by any user. Hybrid and community clouds may be either internal or external. Figure 2.1 illustrates these cloud deployment types.

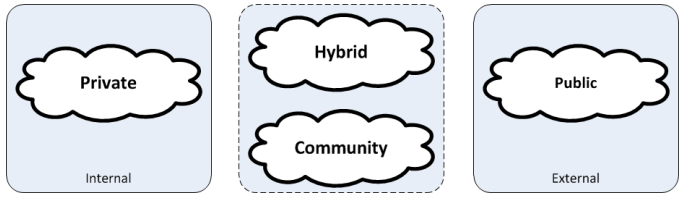

Figure 2.1: Cloud deployment types

2.1.3 Service models

Cloud computing has three fundamental models, these are:

- ***Infrastructure as a Service (IaaS)*** allows customers to use hardware through commonly available interfaces such as Secure Shell (SSH) or a web browser. Amazon Elastic Cloud EC2 offers such a service.
- **Platform as a Service (PaaS)** provides customers with a platform for executing and deploying services through a specific interface. PaaS enables collaboration, so multiple users can work on the same project, thus increasing productivity. An example of PaaS is Google's App Engine.
- ***Software as a Service (SaaS)*** enables users to access the provider's applications running on a cloud infrastructure through a simple client interface, such as a web browser. SaaS applications are installed on remote machines, so that clients do not have to install them on every machine. An example of SaaS is Google's gmail.

2.1.4 Multi-tenancy

In a multi-tenant environment consumers utilize CSP's infrastructure which is shared between other consumers. As stated in [5] "multi-tenancy suggests an architectural and design approach to enable economies of scale, availability, management, segmentation, isolation, and operational efficiency; leveraging shared infrastructure, data, metadata, services, and applications across many different consumers."

2.1.5 Roles

Five roles have been defined in the NIST's Cloud Computing Reference Architecture [61], these are:

Cloud consumer - Person or organization that uses services provided by cloud providers.

Cloud provider - Person, organization, or entity that provides services to cloud consumers.

Cloud auditor - An independent party that assesses various characteristics of a cloud service, such as security, privacy, Quality of Service, etc.

Cloud broker - A party that negotiates relationships between cloud consumers and cloud providers.

Cloud carrier - An organization that provides networking, computation resources, storage, etc. of cloud services.

2.2 Cloud Storage

Cloud storage is a new business model for delivering virtualized storage to customers on demand. The formal term proposed by the Storage Networking Industry Association (SNIA) for cloud storage is Data Storage as a Service (DaaS) - as

> "Delivery over a network of appropriately configured virtual storage and related data services, based on a request for a given service level." [2]

Allocation of costs is important for DaaS. Providing virtualized storage on demand does not require organizations to preorder a defined amount of storage capacity. This enables organizations to save a significant amount of capital because storage costs depend only on the actual amount of storage space used. This business model is extremely cost efficient for startups and small organizations. However, it is not cost effective for organizations that know (or can predict) the amount of storage that they actually need.

Capital cost savings for organizations are very tempting. However, this simply shifts the challenge to the cloud providers. Cloud storage services require deployment of accurate metering and billing mechanisms. Additionally, cloud providers have to meet the potential user's peak demands **without** expanding existing facilities and at a price that is less than or equal to the non-cloud alternative. [2]

A cloud storage service presents a container for data, and the user does not really care **how** the cloud provider implements, operates, or manages their resources within the cloud. A client, via the network, makes requests to the cloud storage to securely store and subsequently retrieve data at an agreed level of service (see Figure 2.2). Although seemingly abstract and complex, cloud storage is actually rather simple. Regardless of the data type, cloud storage represents a pool of resources that are provided in potentially small increments with the appearance of unbound capacity. [2]

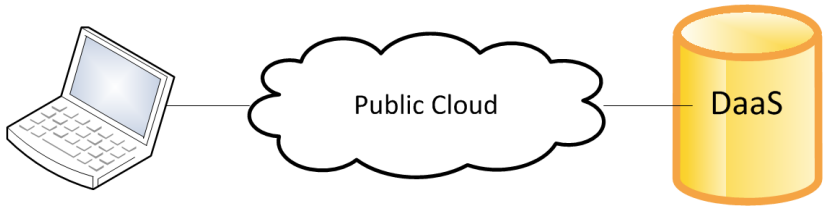

Figure 2.2: Data Storage as a Service. Adapted from [2]

Today's IT environment offers a wide variety of products to store, backup, archive, protect, and make the customer's data available for other business processes. Since DaaS infancy, cloud service providers began to make their own implementations available to users. As a result, a multitude of interfaces have been supplied that have been re-purposed for DaaS, such as block-based access via iSCSI; POSIX interfaces (NFS, CIFS, and WebDAV); object-based CRUD (Create, Read, Update, Delete) interfaces over HTTP; and a plethora of proprietary interfaces for database or table access (see Figure 2.3). Compared to the simplicity of the abstract cloud model, the existing cloud storage model is rather complex, because there are so many interfaces that may be required to meet the different demands of end users for accessing storage. [2]

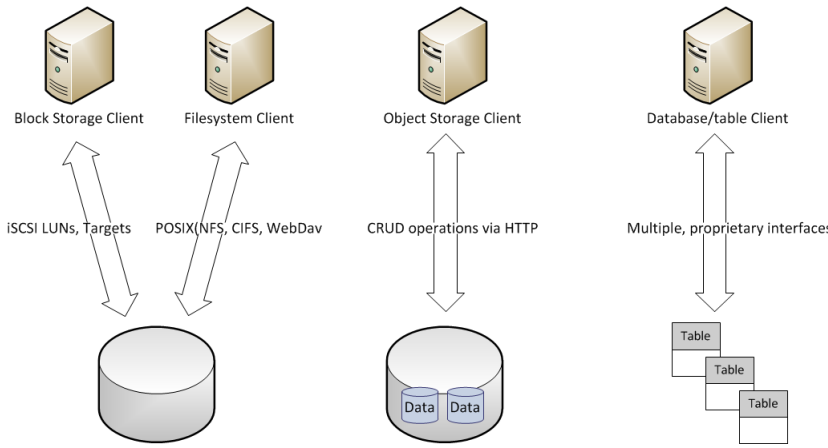

Figure 2.3: Existing interface standards for data storage. Adapted from [2]

2.2.1 CDMI

As cloud storage provides benefits to businesses, such as scalability and cost savings, interest in and adoption of cloud storage solutions is growing. However, each cloud storage provider offers its own cloud storage interface. As a result, multiple

standards exist, which locks clients into proprietary solutions. Thus, there is a requirement to simplify and allow interoperability across disparate cloud solutions. The SNIA's response has been to develop the Cloud Data Management Interface (CDMI), an extensible standard that accommodates vendors' requirements and ensures consistency and interoperability for users. In [2, 60], SNIA describes the details of the Application programming interface (API) .

2.2.2 Existing Cloud storage services

Security should be a top priority when it comes to choosing a cloud storage service. Cloud storage services should employ robust security measures to safeguard the customer's data during transmission and when stored in the cloud. The most basic protection is SSL encryption of the data during transit, password-protected accounts, and multi-level security in the cloud.

Cloud storage services come with different levels of security and privacy. These services can be divided into three categories:

1. Strong security and privacy rules [9, 10], i.e. SpiderOak [11];
2. Modest security and privacy mechanisms [10], i.e. DropBox [12];
3. Weak or no security and privacy mechanisms [10], i.e. Amazon Cloud Storage; This should not be mistaken for Amazon's Simple Storage Service (S3) [13].

As in [10], three cloud storage services, representing each level of security, were compared in terms of their security and privacy, in this case: SpiderOak, DropBox, and Amazon Cloud Storage.

2.2.2.1 DropBox

Dropbox encrypts data in transit with Secure Socket Layer (SSL), while stored data is protected with Advanced Encryption Standard (AES)-256 bit encryption. Data file names are in their original (plain text) form. Dropbox uses Amazon's Simple Storage Service (S3) for storage. S3 has a robust security policy of its own. The overall service design has security flaws and is a subject to known attacks [14].

Dropbox states that their employees are allowed to see only metadata, but not the data itself. However, when legally required to there are some employees who are allowed to view the customer's data.

Although the user's data is encrypted, [14] states that Dropbox employees are capable to decrypt any data. Those who are interested in protecting their data should consider adding an extra layer of encryption before synchronizing data with Dropbox, as described in [8].

2.2.2.2 Amazon Cloud Drive

Amazon's Cloud drive offers no encryption at all. Amazon's Cloud Drive "Terms of Use" [15] states that the provider can do whatever he likes with the user's data. Providing a free cloud audio player encourages users to upload music to their storage. However, all that music is periodically inspected for illegal (i.e. unlicensed content). As stated in [15], Amazon is able to access, retain, use, and disclose any account information and files. In other words if a user wants to use this cloud storage service he has to give up all privacy or protect the data himself or herself.

2.2.2.3 SpiderOak

SpiderOak uses strong encryption techniques while applying both symmetric and asymmetrical cryptography. A combination of 2048 bit Rivest Shamir and Aldeman (RSA) and 256 bit AES keys makes brute-force attacks either infeasible or impossible. In addition, keys are derived from a combination of a pass-phrase and a 32 bit salt value, thus preventing pre-computation or rainbow table attacks.

SpiderOak servers do not store any passwords. In fact, the password used to generate a symmetric key never leaves the user's personal computer (PC). All data on the servers, including encryption keys, are encrypted; thus there is no risk that a rogue employee will be able to decrypt any data to its plaintext.

Tools and libraries used to create SpiderOak are periodically provided as independent open source components. Although the SpiderOak client is still closed source, there are plans to make the entire source code open source.

Ultimately security comes with a price: you will not be able to recover your data if a key or pass-phrase used for encryption has been lost. Some cloud storage vendors are unaware of the user secret that was used to protect the user's data. While other vendors provide mechanisms to recover the user's data. The latter situation might happen if a user forgets their password or the hard drive (typically a Universal Serial Bus (USB) flash) where the user stored their passwords and secret keys has been corrupted.

The dilemma is: Are we (users) paranoid enough to risk their data being irreversibly lost in the cloud? If the providers have the keys to our secure storage vaults, where should we put our jewels (i.e., our valuable data)? Since we already entrust our money to third parties (banks) perhaps it is time for third party key escrow services. However, as Abdullah Azfar has shown in [16] that N of M escrow techniques can be used, thus we do not have to trust any individual escrow agent, while still having a high probability of recovering our own keys.

Table 2.1 shows a comparison between the three different cloud storage services.

Table 2.1: A comparison between the three different cloud storage service

CSP	Cloud Drive	Dropbox	SpiderOak
Security	None	SSL and AES-256 bit	SSL, 2048 bit RSA, AES-256, PBKDF2
Privacy	Access is granted to those who know the password. Amazon employees can freely read your data as it is stated in the "Terms of Service".	Some Dropbox employees are empowered to read user's data. Dropbox comply with the US-EU Safe Harbour Framework and the US-Swiss Safe Harbour Framework. Dropbox manages and stores Encryption keys.	"Zero knowledge" approach; SpiderOak encrypts the data on user's computer before uploading them to the server shared containers.
Overall	Perfect for storing non-private data, especially music and video.	On 19th June, 2011 an authentication bug was discovered [7]. Although it was patched in a timely fashion, it implies that the security by design is insecure. The "Terms of Service" implies that DropBox is cooperating with the government and making them capable to wiretap the communication.	Slowly approaching as an open source product. Security through obscurity is not a good idea, thus going open source is a very good decision. Sharing mechanism design has security flaws since a password is sent with an url. However, without that password it is impossible to decrypt the data.

2.3 Cloud service brokers and carriers

Although the terms cloud broker and cloud carrier are not new in areas such as real estate and telecommunications, in cloud computing these two roles are relatively new. Since so many cloud providers have entered the market, it is hard for a customer to choose a suitable cloud provider for their needs from the many cloud service providers (CSPs). It is even harder to integrate cloud solutions across different providers. Thus, cloud brokers and cloud carriers will arise in a near future, in order to provide customers simplified methods to adopt and utilize cloud services. These two terms are further defined in the following subsections.

2.3.1 Cloud service brokerage

In [61], the U.S. NIST divides cloud service brokerages into 3 main categories:

- **Intermediation**: In intermediation a cloud service broker enhances a given service by adding additional intermediation services. The broker's influence and capabilities depend on where the broker is placed. Such intermediation broker is capable of measuring service usage and supervising pricing and billing. Figure 2.4 illustrates cloud brokerage service.

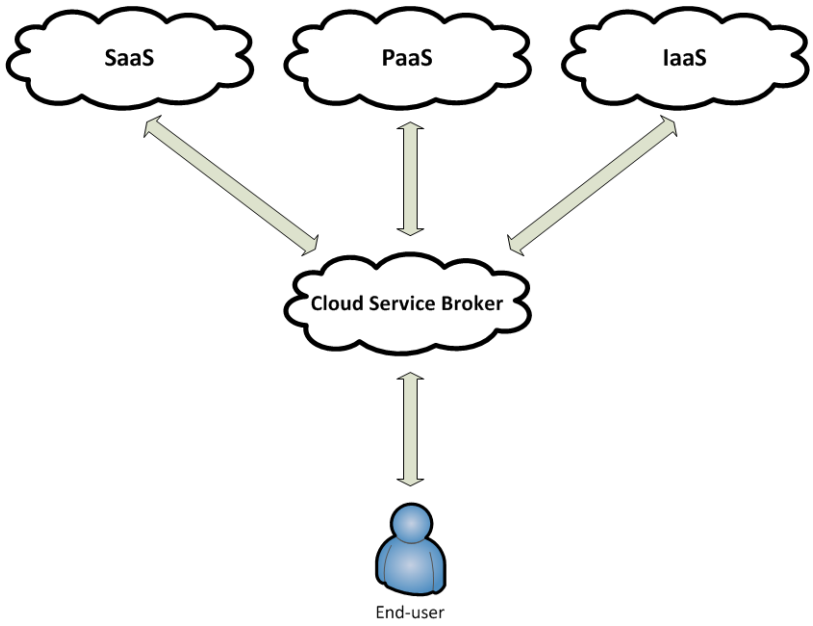

Figure 2.4: Illustration of a cloud brokerage service

A cloud brokerage service might be available at three locations. First, the brokerage service might add additional governance at the cloud service

provider's side by providing services such as access control or identification. Second, the same type of governance might be added at the customer's side. Finally, intermediation brokerages might appear as discrete services placed between the cloud service provider and the cloud customers.

- **Aggregation:** Multiple services provided by different cloud service providers might be aggregated into an entirely new service or set of services by an aggregation broker. For instance, one provider might offer a webmail service, while another provider adds security measures for incoming and outgoing data (e.g. spam filtering). An aggregation broker ensures secure data movement between multiple providers and integrates a number of service components.

  Although each individual could aggregate several cloud services and integrate them into their own systems, that would require a lot of effort. Aggregation brokerages would exist in the cloud as a separate cloud service providers, forming a new (meta-)application layer. This type of brokerage service is thought to be permanent, as once a broker has chosen a set of cloud service providers, it will not be that easy for this broker to switch them to alternative ones. Although, if the broker has chosen CSPs with standard APIs, then it would be easy to change CSPs.

- **Arbitrage**: Cloud service arbitrage is more flexible than an aggregation broker, since this approach allows a customer to migrate from one provider to another quite easily. For instance, if an arbitrage broker has integrated three different service providers with similar capabilities and is now ready to deliver a new service, e.g. data storage service; then it is unlikely that all three integrated services are in use, as only the the one which offers best price is likely to have customers. Once another provider offers a better price, the broker will initiate service migration between the two providers. This will ensure that no monopoly is present, but it could lead to an oligopoly market.

  Arbitrage brokers can also provide entirely new services by combining multiple services from different cloud service providers. This new service will be directly available to the end-user directly, however the resources being used will be not the arbitrage broker's services, but rather the different cloud service providers' services, hence these different cloud service providers became virtual partners due to the arbitrage. These providers may need to become true partners as a result of this arbitrage because the arbitrage broker wants to ensure that the end-user will get the new service that is offered.

Rapid innovations in cloud computing lead to new and compatible APIs, thus cloud service brokers will become an important part of the overall cloud ecosystem. Cloud service brokerage can abstract a number of interfaces, each provided by different CSPs, thus helping customers to take even greater advantage of the cloud model.

2.3.2 Cloud carrier services

Cloud service carriers provides a dedicated transport level infrastructure to the cloud. This infrastructure interconnects CSP and its customers. Although the telecommunication companies providing these services are assumed to operate in the lowest part of the stack, these days the carriers are providing transport solutions tailored to the needs of various cloud providers and consumers. Figure 2.5 illustrates this interconnection.

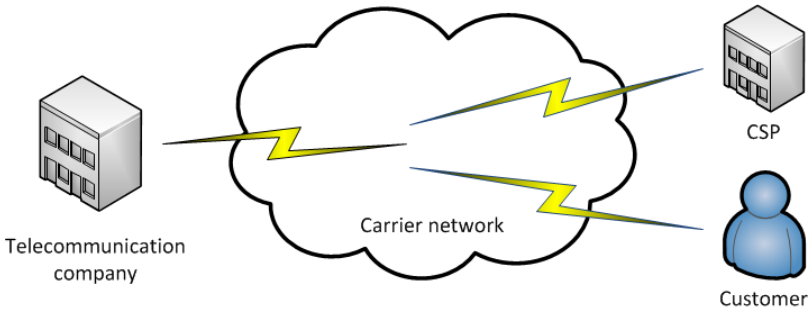

Figure 2.5: Illustration of a cloud carrier service

2.4 Federated Clouds

Currently cloud services are offered by many providers. The cloud service providers (CSPs) include: Amazon, Rackspace, Microsoft, and others. Most CSPs offer proprietary solutions which are not interoperable and as a result customers are locked-in to a single CSP. Furthermore, customers have different service requirements in terms of Quality of Service(QoS), security standards, data privacy, etc. Regulations and legislation also vary from country to country. This means that in order to satisfy the customer's needs the CSPs have to be more versatile. This can be achieved by joining a cloud federation, where different cloud service providers share common APIs and offer diverse Service-level Agreements (SLAs).

2.4.1 What is a federated cloud?

On October 25, 2011 NIST published a final version of their cloud computing definition. However, that document does not describe federated cloud ecosystems. Subsequently NIST released another document - defining a Cloud Computing Reference Architecture [61]. While this later document is still a draft and is a subject to being modified, some clues are given in their definitions of cloud carriers and brokerage services, but there is still nothing written about a federated cloud ecosystem. Krishnan Subramanian [17, 18] states that the essential characteristics of a federated cloud ecosystem are:

- **Multiple providers:** The most important feature of a federated cloud is to have at least 2 independent providers. Having even more providers in a cloud federation enables load balancing across providers. For instance provider A utilizes 90% of its resources and predicts that resource consumption will reach its limits within a short period of time. Since provider A has a SLA with provider B, further customer requests to provider A are redirected to provider B. This migration is transparent to the customer and provides benefits to both providers in this federation.

- **Diverse cloud platforms:** A idea behind a cloud federation is that end-users are important and no monopoly provider or platform is present. As a result end-users should not be tied to a specific platform. For instance if a platform is no longer adequate, then a customer simply migrates to another provider whom is more suitable for their needs. Although some open source activists might think that there would be no problems if all providers used open source platform, Krishnan Subramanian, a research analyst, in [17] states that:

  > "Monopoly at any layer is bad for the users ... I don't support monopoly even if the underlying platform is open source."

  Since homogeneous platforms may result in a customer lock-in on a platform level [17], the best way to ensure customer satisfaction and avoid a monopoly is to bring heterogeneous platforms together in a federated cloud.

- **Interoperability:** In [19] Bill Claybrook states that an ability to use the same management tools, server images, and other applications is considered to be interoperability. In a cloud storage context this would be use of a file system (the Z File System (ZFS), fourth extended filesystem (ext4), New Technology File System (NTFS), etc.), common encryption mechanisms (AES, blowfish, etc.), and so on. Interoperability is very important in a federated cloud ecosystem, because if CSPs can not communicate with each other via common APIs, then there would be no point in establishing a cloud federation in the first place. Ensuring interoperability is not an easy task, but it is achievable by using standardized APIs (e.g. CDMI) and open formats.

- **Migration:** The capability to move your applications and/or data from one provider to another allows customers to avoid vendor lock-in. Migration is a crucial feature when a CSP changes its policy and no longer provides satisfactory service to an end-user. Imagine a situation where an employee of a cloud storage service X was caught selling a customer's private data to a third party. A company might want to immediately migrate all of its data to another provider. Unfortunately, this will not address the issue of whether the company's data has already been "sold" or if there is another copy of the company's data which is not under its control and could be "sold" or used by someone else [48].

  Currently, if an end-user wants to migrate from one provider to another he or she would have to do so manually. This can become tricky and time consuming, especially when you have a lot of stored data. Additionally, when migrating from a compromised provider security will be a major concern. As a result migration should be handled in a secure manner. One of the most difficult aspects of migration is ensuring that the previous copy of the data is now unusable or ideally no longer in existence. Unfortunately, putting your data into the cloud is a bit like Java programming's "Write once", but now the attribute is "Read anywhere" rather than "Run anywhere" [48].

- **Geographical distribution:** Although cloud computing may use a Content Delivery Network (CDN) to allow faster access to the content, cloud providers in a federated cloud ecosystem have to locate specific data in different regions, while avoiding placing specific data in other regions. This should be done not only to ensure faster content delivery, but also meet various regional regulations and local legislation.

We can summarize our discussion of a federated cloud using Techtarget's definition of a Federated cloud:

> "A federated cloud (also called cloud federation) is the deployment and management of multiple external and internal cloud computing services

to match business needs. A federation is the union of several smaller parts that perform a common action. Service-level agreements will take a key role in a federated network." [20]

2.4.2 Service-level agreements

In a federated cloud ecosystem SLAs define many service aspects, such as uptime, customer support, data privacy and security, legal jurisdictions, pricing, etc. Steve Caughey from Arjuna Technologies Limited states that "...the glue which connect Clouds together - Service Agreements." [21] Table 2.2 shows common SLA elements as stated in [49, 50].

Table 2.2: Elements of common SLAs

Element	Description
Auditing	Independent and unbiased third party assesses CSPs.
Monitoring	Measures performance of the services and checks if load is managed as defined in the SLA.
Metering	Assurance of accurate billing.
Availability	How availability is measured is subject to some degree of interpretation. For accurate billing this must be clearly defined in a SLA. For instance most CSPs state that their services are available 99.9%
Performance	SLA may define a bottom and top performance thresh-olds (i.e. read/write speed, link speed, etc.).
Operational recovery	Includes details of how a service's operation failure should be handled. In addition, recovery period should be defined.
Disaster recovery	An enforceable and detailed SLA should define a disaster recovery procedure.
Security	Data must be encrypted before being sent out. In addition, encryption keys are typically kept by the user and should not be available to CSP (in this case key recovery procedure is not available). There must be a guarantee that the client's data is isolated and is not accessible by other clients in a multi-tenant environment.
Interoperability	Consistent, open standard interfaces (i.e. CDMI) for accessing and managing private and public cloud services (e.g. Amazon's S3) are required.
Portability	Open standards must ensure that data at rest can be migrated across different CSPs.

SLAs might also define interoperability, migration, and other features among providers within a cloud federation. Although SLAs have a major part in establishing a federated cloud, this thesis will not discuss them due to the fact that they are usually confidential.

2.5 Identity management

Identity Management (IDM) in a cloud manages control points, virtual devices or service identities, etc. An IDM for a cloud storage service requires dynamic governance of typical IDM issues like lifecycle management, provisioning and de-provisioning, entitlement, synchronization, etc.

2.5.1 Identity lifecycle management

Lifecycle management's functionality is divided into the provisioning and administrative components. Together these components manage user identities, their credentials, and entitlements. The administrative component is responsible for defining delegation rules. The provisioning component defines provisioning and de-provisioning procedures, manages policies and password maintenance tasks. In addition provisioning component is responsible for entitlements, deactivation, and proliferation of on-demand user ID.

2.5.1.1 Provision an de-provisioning

On-demand provisioning is based on a trust model and does not require service providers to exchange data about users in advance. De-provisioning should be real time and user account should be synchronized with all its service providers instantly.

2.5.1.2 Entitlement

Entitlement defines user's access rights and privileges. These entitlements are managed by giving a set of attributes to a user. CSPs should use same type of attributes, preferably defined by a standard, otherwise interoperability becomes challenging.

2.5.1.3 Proliferation of On-demand User ID

Users might have multiple accounts with different CSPs. Occurrence of multiple identities for the same user poses a challenge to interoperability as the access right and privileges should be synchronized. This issue can be resolved by using OpenID or SAML standards which are discussed in the next subsection.

2.5.2 Federated identity standards

Federated identity management allows customer identification and management through single sign-on. Identity provider is an entity responsible for creating, maintaining, and authenticating all user identities that enables users to securely operate among network members. Users need only sign on once with any member to access Web sites in the circle of trust.The following are common standards in the industry which enable the federated identity management.

2.5.2.1 Security Assertion Markup Language (SAML)

SAML is an XML standard for exchanging authentication and authorization data between security domains, that is, between an identity provider and a service provider. SAML enables web-based authentication and authorization scenarios including single sign-on (SSO) or identity federation regardless of the underlying architecture.

2.5.2.2 OAuth

OAuth is an open standard for authorization and it allows users to share their private resources to a third party site (service provider) without sharing their access permissions or the full extent of their data. The details of this standard are described in RFC 5849 [57].

OAuth uses tokens to allow requesting service to granularly access user's shared data. For instance if a service provider gets a token to access user's contacts, this will not allow the same provider to access user's photo albums. OAuth is vendor neutral, which is a major advantage for developers who would otherwise need to support each vendor's authorization mechanism.

2.5.2.3 OpenID

OpenID provides a simplified way to sign up or sign in to a service. With OpenID a user has only one globally unique identifier and it unifies information about the user, but only that information that is made public. Users identify themselves by providing a password to their OpenID identity provider which afterwards grants access to the site if the authentication was successful. No website other than user's provider ever sees user's password, so users do not have to worry about an insecure website compromising their identity.

OpenID is as secure as other means of authentication, however if user's identity gets compromised this will result on a large scale. Most OpenID identity providers offer password recovery by sending a recovery email. Then a compromised email account may result in compromising all other services which uses OpenID.

Chapter 3

Cloud Storage Security and Privacy Risks

This chapter examines the risks to privacy and security of storing data in the cloud (sections 3.1 and 3.2 respectively) and concludes with a section regarding data protection requirements (focusing on regulations in the EU).

3.1 Privacy risks in using Cloud Storage

Due to the separation between cloud users and their data, there are a number of serious privacy risks with storing information in a cloud. This section examines key privacy risks which can appear due to storage in the cloud.

3.1.1 Jurisdiction

Data in a cloud can potentially be stored, processed, and used in other ways within multiple jurisdictions. However, data protection laws differ in the various jurisdictions. As a result cloud based storage might be a serious threat to sensitive corporate or private data. Moreover, some of the different data protection legislations require that the data have a distinct ownership. However, in some cases it is in practice hard to identify the owner of the data.

3.1.2 Creation of new data

The cloud model has the potential to create and retain a huge amount of new data related to the activities of the cloud user. The creation of such data may raise concerns about the ownership of this data. This secondary data is generated by interactions with a cloud-based infrastructure. Although this data is not the actual data which is stored in a cloud by the cloud user, the ownership of this new data is a subject for debate. For instance, Facebook is storing information about what the users like, who their friends are, what music they listen to, what

movies they like, etc., and later related advertisements show up in their profiles. Some might say that data created by interacting with a cloud based infrastructure should be owned by the user who this data concerns and therefore be protected by data privacy legislation and hence not be resold to third parties without the user's explicit permission.

In the report "Reaching for the Cloud(s): Privacy Issues related to Cloud Computing" [28] the Office of the Privacy Commissioner of Canada states that "In the Pew Internet Study, users expressed great concern about the misuse of their data in the cloud 90% were concerned about their data being sold to another organization; 80% expressed concern about their photos or other data being used in marketing campaigns; and 68% said they would be concerned if their data were analysed and used to serve them with targeted advertising". This suggests that the users are becoming more concerned about their data privacy and in some countries there are those who believe that these users' rights should be protected by appropriate legislation. Finally, the secondary data created in the cloud may be personally identifiable information (according to the EU regulation 95/46/EC) and hence subject to restrictions. Additionally, individuals might be unaware of the existence of this data.

3.1.3 Securing the data

The internet is not a safe place for sensitive private data to travel. Additionally the cloud model does not define what security measures should be taken in order to secure the data while it is inside the cloud. All security related decisions depend upon the specific policies and actions of each CSP. This raises security risks both in the protection of data and in the safeguards applied to this data. According to [28], recent studies show that CSPs have tended to provide their services *without* strong security solutions. However, Christopher Soghoian recommends that CSPs should use the kind of encryption which is currently used by on-line banks. Moreover, data protection should be applied to data at rest, in transition, and while processing it.

3.1.4 Lawful access

Cloud computing raises additional concerns when the private data in the cloud has to be accessed by the government, its agencies, etc. For instance a lawful access request can target a certain individual or a company whose data is stored in the cloud. However, if there is data which belongs to multiple data subjects, this data may also be exposed. This actually raises four privacy risks. First, the court order or other lawful access request may result in access to information above and beyond what was intended. Second, the CSP client who is not the target of the lawful access request might be unaware of the possible data intrusion and might never be informed of this intrusion. A third risk is that the target of the lawful access request might also never be aware of the intrusion. A fourth risk is that the government agency which receives this information might not securely handle the data or they may retain the data for longer than it should be retained.

3.1.5 Misuse of processing data

The CSP should be bound to the privacy requirements equal to those used within the organization whose data is going to stored or processed in the cloud. A CSP must ensure that access and modification procedures are possible and that deletion procedures are adequate and appropriate. These procedures and privacy requirements are important because there is a possibility that a CSP might access, manipulate, or mine data in an inappropriate way [28]. In that case, regulators may have to distinguish whether the data were processed for a specified purpose or purposes in order to know which regulations or laws are relevant.

3.1.6 Permanence of data

In the contract between an organization or a person and a CSP there should be a statement of what measures will be taken to ensure that the data is protected while it is held in the cloud by the CSP. However, there is a security and privacy risk to the data when the contract expires. Methods should be introduced to securely remove the customer's data from the cloud infrastructure. A client should be acquainted with what will happen to his data after the end of the contract and within what time period these operations are guaranteed to be carried out. Moreover, in Megaupload's case [56] customers' data is no longer accessible to these customers since some of them violated copyright law. All 25 PB of data residing in the data center is seized by the law enforcement authorities and is not available even to those customers whom did not violate copyright law. From the perspective of the data center this case brought a huge financial loss since the government is not willing to pay for operational costs of data retention and is does not allow to delete that data.

3.2 Cloud Storage data protection risks

Customers who store their data in a cloud should be familiar with the risks of data being collocated in a shared environment. NIST in [34] defines the main data protection risks for stored data in a CSP and the risks when migrating data between providers.

3.2.1 Data concentration

Currently, information has a huge value and that data is consolidated into a huge cloud-based data storing facility. Because this data has such high value and it is all in one place it is a clear target for an attack. The basic reasons for such data being desirable target is due to the economy of scale - as a successful attack has a greater yield for the effort of carrying out the attack. As a result an attacker is more interested in exploiting a system which has a lot of data, even though a successful attack may require more effort than an attack against a target that has little data and requires slightly less effort.

Such information storage vaults require sophisticated security measures including proper password reset operations. As stated in [41, 42] a famous social networking site Twitter [43] was exploited because the site's administrator's account password was reset by someone who successfully answered the security questions. The correct answers were gathered by social engineering. A similar weakness was found in Amazon's Grid Computing Services [40]. An attacker who controls a mail service can access a tremendous number of user accounts as frequently lost passwords for cloud services can be reset by using a Uniform Resource Locator (URL) or code word sent via electronic mail [34]. If an attacker eavesdrops the communication link through which a password reset mail is sent, he or she may effectively take over that account.

3.2.2 Data isolation

In cloud storage data can take many forms, for instance it can be a container of data or simply a set of files and associated metadata. In addition, part of a customer's data might be stored within a database (for example, private data such as name, address, payment card number, etc.). To successfully secure data from unauthorized access, a suitable access control mechanism should be used. Identity management is one of the biggest issues in cloud storage as physical authentication is not possible, hence it is easier in the internet to impersonate another person than in a reality. Currently data centres offer high-level physical security [44]. However, there is always a possibility that a rogue employee steels or alters data. Encryption should be used to protect the data, with the encryption keys stored *outside* of the data center, preferably held by an key escrow service. However, Bruce Schneier has stated in [45] that 'A variety of "key recovery," "key escrow," and "trusted third-party" encryption requirements have been suggested in recent years by government agencies seeking to conduct covert surveillance within the changing environments brought about by new technologies' so that these government agencies may continue to conduct covert surveillance.

In [34] Jansen and Grance say "Data must be secured while at rest, in transit, and in use, and access to the data must be controlled". Data transfers have been secured by introducing standardized security protocols such as SSL [36] and Transport Layer Security (TLS) [46]. However, protection for data at rest has not been standardized yet [34].

3.2.3 Data sanitization

Data stored in a cloud should be deleted with great care, as forensic tools can be used by both criminals and law enforcement authorities in order to restore deleted data - even in a multi-user environment. Since customers of a cloud storage service share the same storage media, there is a possibility that a cloud storage user can restore other customers' data from a given container. Moreover, it is easy for a rogue employee to recover insecurely deleted customer data. There are cases [37, 38]

where it was possible to recover data from hard drives that had been disposed of by selling them on the Ebay Online Store [39].

Kissel, et al. [35] provide guidelines on how data storage should be properly sanitized. Sanitization involves the expunging of data from storage media by overwriting, degaussing, or other means, or the destruction of the media itself, to prevent unauthorized disclosure of information [35]. Data sanitization applies to repurposed equipment (usually after hardware upgrade), backup copies, and also to any data which is in storage after the end of the contract.

3.3 Data Protection Requirements

This section describes the terms used various data protection acts and legislations. Next, specific legislations such as the EU Data Protection Act, United Kingdom's Regulation of Investigatory Powers Act (RIPA), and the UK Data Protection Act, are introduced.

3.3.1 Data protection basics

We begin our discussion of data protection requirements by starting some key definitions:

- **Data** means information which - is being processed or is recorded with the intention of being processed or stored on a computer or similar equipment.
- **Personal Data** is information which can identify an individual. If this data relates to other information which can identify an individual, then that data is also considered to be personal data. For instance, if an USB drive contains a spreadsheet with Unique Identification (UID) numbers , then although, these numbers do not directly identify an individual, it may still possible to relate an individual to these UIDs if there is a match in another system containing personal data (such as the tuple: UID, Name, Surname).
- **Sensitive Personal Data** - is information about an individual's health or criminal record.
- **Data controller** - an organization that determines the purposes and manners in which way any personal data is processed.
- **Data subject** - is the individual who can be identified from his private data.
- **Processing** - means obtaining, recording or holding the information or operating with that data.
- **Permission Based Marketing** Opt-in means that a user has to give an explicit permission for a specific purpose in order for it to operate. Opt-out means that a user has to give an explicit denial that disables an operation.

3.3.2 EU Data Protection Act

The EU Data Protection directive (95/46/EC) [22] regulates the processing of personal information within the EU. Personal information is defined as any direct or indirect information which can link or identify a natural person (i.e. the data subject). The directive provides personal information recommendations such as that all data subjects should be given notice when their data is being collected; that data should only be used for its intended purpose and it should not be disclosed without the data subject's consent. Additionally, the collected data should be secured and available for modification by the data subject in order to correct any inaccuracies.

3.3.3 Safe Harbour Agreement

The EU has for many years had a formalized system of privacy legislation, which is regarded as more rigorous than that found in many other areas of the world [27]. The Safe Harbour agreement is a part of the EU Data Protection Directive (EU directive 95/46/EC), it sets strict privacy protection requirements for EU citizens. Basically this agreement prohibits EU organizations from transferring personal data outside the European Economic Area, unless there is a guarantee that the EU mandated data privacy requirements will be met. For example, US companies can verify that they comply with these principles, or hire a third-party to perform the assessment if they agree to meet EU standards under the directive's Safe Harbour Principles.

3.3.4 RIPA

Regulation of Investigatory Powers Act 2000 (RIPA) [27] is an act of the Parliament of the United Kingdom regulating the powers of government agencies to carry out surveillance, investigation, and interception of communications. RIPA can be invoked by government officials specified in the Act on the grounds of national security and for the purposes of detecting crime, preventing disorder, ensuring public safety, protecting public health, or in the interests of the economic well-being of the United Kingdom.

The act enables government agencies to demand access to an Internet Service Provider's (ISP's) customer's communications and allows mass surveillance of transit communications. Additionally, the act allows authorities to demand UK citizens to hand over keys to protected and encrypted information [52] in order to enable government to monitor people's internet activities.

3.3.5 UK Data Protection Act of 1998

The UK Data Protection Act of 1998 [25] aims to implement the European Data Protection Directive [22], hence introducing new provisions for the regulation of the processing of information relating to individuals, including the obtaining, holding, use, or disclosure of such information.

3.3.5.1 Eight Data Protection Principles

The act defines eight data protection principles. By following these eight principles, organisations will comply with the Data Protection Act. These eight principles are fundamental to understand the Data Protection Requirements in the United Kingdom and are a good example of how the EU Data Protection Directive should be implemented within the European Union. These eight data protection principles are:

1. **Processing personal data fairly and lawfully**

   Individuals should know what data will be collected and for what it will be used. Some organisations (charities, etc.) might share personal data with other similar organisations. However, organisations may sell, trade, or even rent this information. In any case, the information processing must be done fairly.

   An organization should notify an individual about all opt-out services which will be activated after an individual enters into a relationship with the organisation. Ideally, an organisation should provide an opt-in choice for an individual in order to comply with the first principle. However, not all organisations that offer opt-out should be considered as fair. Additionally, many fair organisations simplify the process of data processing based on their previous relationship with an individual (for example by transferring privacy settings, etc. to a new service).

   In any case, if an organisation intends to disclose an individual's information to other parties, it must get the individual's consent. Usually, this is stated in a "Terms of Service" agreement. Unfortunately, most individuals do not read this document and they expect a simpler form of privacy notification.

   Fairness requires an organisation to:

   - Clearly and honestly state their true identity;
   - Define for what intend purpose any personal data will be used;
   - Handle personal data in a way that an individual would reasonably expect; and
   - Not to use any information that could have a negative effect, unless this intention is well-grounded.

   Moreover, the first principle requires the information processing to be lawful. However, the Data Protection Act itself does not specifically define what is

lawful processing and what is not. In order to comply with this principle an organisation has to follow common sense and local laws.

Committing a crime with processed personal data is directly unlawful (this is in addition to any industry-specific legislation). In addition, processing may be unlawful if it results in a breach of a duty of confidence where confidentiality is expected, such as in client - attorney communication, medical records, etc. Copyright infringement and intrusion into private and family life (the Human Rights Act 1998) are also considered to be unlawful.

2. **Processing personal data for specified purposes**

The Data Protection Act says:

> *"Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes."*

Principles 1 explains that personal data should be processed in a fair and lawful way. Principle 2 adds an additional requirement for processing of a private data, i.e. that the entity that obtains personal data must specify a single or multiple purposes for this data that will be used when an organisation collects the data. Anything, that the organisation does with this information, should be compatible with the stated purpose of collecting it.

An organisation wishing to use or disclose personal data of an individual has to take into consideration that an individual reasonably expects such use of his data and that this disclosure would not have any unfavourable outcome concerning the individual. In addition, the secondary purpose has to be lawful and fair as stated in Principle 1. If an individual does not expect that his personal data is being used and/or that this use has an unfavourable effect, then the purpose of collecting data and secondary purpose are incompatible and therefore an organisation does not comply with the Data Protection Act of 1998.

For instance, a matchmaking organisation collects an individual's private information such as name, surname, age, email address, and specifies in its privacy notice that this data will be used to find a potential wife or husband. However, if this organisation is subsidiary to a parent company which specializes in providing leisure activities for couples and these two organisations share a unified client database, then if the parent organisation sends offers of leisure activities to the clients of the matchmaking organisation the action of the parent organization is not lawful unless the policy notice

of the matchmaking organization also specifies that the data will be used to offer leisure activities, therefore making the use of personal information for this purpose is compatible with the purpose for which it was collected.

Unfortunately, in practice, organisations ask upfront for an individual's permission to use personal data for undisclosed secondary purposes.

The third, fourth, and fifth principle are tightly related as they define standards that must be applied to personal information. As defined in [23] these standards are:

- Adequate, relevant, and not excessive;
- Accurate and where necessary, kept up to date; and
- Data can not be kept for no longer than necessary.

3. **Information standards - the amount of personal data that an organisation may hold**

The Act says:

> *"Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed."*

This means that an organisation should not store any data unrelated with the purpose or purposes of the original use of the personal information. This process is also known as "data minimisation" which is defined in [26] as "collect only what is essential, store it only for as long as is necessary and resist a tendency to collect more personal information".

In practice this means that an organisation must have just the right amount of personal data for the purpose specified prior to collecting that data. An organisation must be clear why it is holding and using that data. It can not store personal information just because that information might have some value in the future. However, it is permissible to retain information for an event which might or might not happen. It is uncertain, for how long an organisation can hold data collected for an anticipated event. Since, this is not stated in the Act, organisations may take advantage of this flaw and create privacy threats to individuals' personal data by creating events which might occur at some indefinite point in the future.

4. **Information standards - keeping personal data accurate and up to date**

   The Act says:

   > *"Personal data shall be accurate and, where necessary, kept up to date."*

   It is quite obvious that keeping all personal data accurate is a formidable task. The Act says that personal information is incorrect or misleading, then it is inaccurate. For instance if an individual stated that his email is name@domain1, and his current email address is name@domain2, then this personal information is inaccurate. In some cases organisations wish to keep a record of the inaccuracy (mistak)e for accounting or other reasons. The Act allows this data to be retained but only if those records are accurate.

   Not all information has to be updated. If the data was collected for statistical purposes, it is obvious that this data should not be updated, unless an error was found. In practice, it is not always obvious whether information should be or should not be updated. In addition, it is sufficient for an organisation to check a sample of personal information to see whether it is accurate or not, i.e., they are not required to check that all of the information that they are storing is actually correct.

5. **Information standards - retaining personal data**

   The Act says:

   > *"Personal data processed for any purpose or purposes shall not be kept for longer than necessary for that purpose or those purposes"*

   It is unclear for how long an organisation may retain personal data. If an organisation deletes private information too soon, this will disadvantage their business and may require to collect private data more often than is necessary. In contrast, if an organisation holds data for too long, it will be outdated and is more likely to be inaccurate, thus this data is likely to be prone to errors.

   As stated in [23], "personal data held for longer than necessary will, by definition, be excessive and may also be irrelevant. In any event, it is inefficient to hold more information than necessary". Purging unnecessary data should be done on regular basis.

6. **The right of individuals**

The Act states:

> "Personal data shall be processed in accordance with the rights of data subjects under this Act"

As stated in [23] individuals have rights to:

- access to a copy of their personal data;
- object to personal data processing if it is in any way harmful to an individual;
- prevent their personal data from being utilized for direct marketing;
- correct or delete inaccurate personal data; and
- claim compensation caused by a misuse of their personal data.

7. **Information security**

The Act states:

> "Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data."

As in [23], an organisation should take these practical measures:

- design and organise its security (both physical and technological) to fit the nature of the personal data it holds and the harm that may result from a security breach; and
- assign responsible personnel who ensure information security and to train all staff to respond to a security breach in a timely fashion and according to well defined procedures.

Individual's personal information security is of top priority as its misuse can lead to identity fraud or identity exposure of service personnel (this is especially true for secret service personnel, witnesses, etc.). The consequences of breached security may result in individual's embarrassment, distress, or even a risk to life.

As stated in [23], the Act does not require an organisation to have state-of-the-art security technology to protect the personal information it holds, but the organisation should regularly review its security measures as technology evolves.

> Before deciding what security measures to take, an organisation should assess the value of the personal information, and what consequences could be caused if this data would be exposed or lost. The security measures should ensure that personal data is only available to authorised personnel within their scope of authority. In addition, all personal data should be backed up and available for recovery if an accident occurs.

8. **Sending personal data outside the European Economic Area (EEA)**

The Act says:

> "Personal data shall not be transferred to a country or territory outside the EEA unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data."

Figure 3.1 illustrates the member countries of the EAA which are: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden.

According to the Act, there are no restrictions on the transfer of personal data within the countries shown in Figure 3.1. However, if the data is intended to be accessed (over internet, or by any other means) from countries outside the EAA, then the recipient country must provide adequate protection. [23, 30] states that Argentina, Canada, Guernsey, Isle of Man, Jersey, Switzerland, and the Faroe Islands have an adequate level of protection for personal information. In addition, [30] states that the US has an adequate protection if the personal information is sent under the Safe Harbour [26] scheme.

Before an organisation transfers an individual's personal information to countries outside the EEA which are not listed in [30] as having an adequate protection, then an organisation has to make an assessment of adequacy. This includes:

1. analysing how the data protection standards were adopted in its law;
2. ensuring that these standards are achieved in practice; and
3. making sure that there is a procedure where individuals could get compensations if their personal data were misused.

If an organisation is transferring personal information to a third party which acts as a data processor, an organisation must ensure that this third party has adequate security and acts only on the instructions given by the organisations. It is best to have a contract with the data processor, defining the procedures and measures specifying how an individuals' personal information should be handled.

Figure 3.1: Member countries of the EEA (source: [29])

In any case it is not recommended to transfer personal data to unstable countries such as Egypt, Syria, Libya, etc.

Other options of how to transfer personal information outside the EEA is to:

1. authorise transfers by the Information Commissioner;
2. adopt binding corporate rules [31]; and
3. use the three sets of standard contractual clauses in a contract [32].

There are exceptions to the above procedures. An organisation may transfer individuals' personal information with their consent, however an individual has to have a choice to make that consent or not, without any harmful consequences (employee dismissal, denial of service, etc.). In addition an individual should be aware of the reasons for the transfer and possible threats. Finally transfers related to fulfilling contracts are excluded.

3.4 Cloud security comparison to traditional IT

After analysing cloud storage security and privacy risks, we came up to a conclusion that *cloud is neither more or less secure than the traditional IT*. However, due to the fact that data stored in the cloud is concentrated more than in traditional IT, the consequences of the security breach in a cloud are far more severe. In addition, as data stored in the cloud are more valuable (see subsection 3.2.1), cloud storage services have potentially higher risk to be attacked and exploited.

3.4.1 Security in reality

When businesses move from traditional IT to the cloud model, they give up most of the control to the provider. This, in any case does not necessary mean that security is or will be compromised. In contrast, most CSPs make a great effort to secure their services as much as possible. Moreover, it is more likely that a startup deployed in the cloud ecosystem would have a better security implemented than the startup which has deployed its services on their own in a traditional IT environment. However, this might not apply to enterprises having their own IT security departments.

Studies [51] show that a perception that a cloud model poses greater security risks (compared to traditional IT deployments) is not supported by the empirical data. In contrast, the study [51] showed that:

- CSPs had lower occurrence rates for every class of incident examined;
- CSP's customers experienced lower threat diversity; and
- Cloud environments were twelve times less likely than traditional IT environments to have common configuration issues.

The study [51] explains these findings as in a cloud environment each customer has fewer applications with tightly controlled network access. This results in a relatively small area for an attack.

3.4.2 Recommendations

Cloud model increases agility, flexibility, scalability, and boosts business efficiency by handing over their IT management to a CSP. To successfully move into a cloud, an organization should understand the cloud environment and consider how it impacts the organization from a security standpoint.

Transitioning to the cloud might be less painful by first moving services in low risk areas that does not put the organization at increased security risk. As trust and working relationship with the CSP builds up, the organization should identify potential security risks and customize the SLA to meet their security expectations.

Chapter 4

SeaaS for Cloud Storage

This chapter introduces a scenario where a customer requests a broker to securely migrate his personal data from one CSP to another within a cloud federation. Next, the criteria for a secure cloud storage service shall be identified and compared to existing solutions. Finally, weaknesses of existing solutions shall be identified and solutions shall be proposed which will lead to the design of the secure cloud storage service.

4.1 Problem scenario

The scenario consists of three parts. First a customer signs a contract with a broker. Next his data is moved to one the CSPs. And finally his data is moved from one CSP to another.

4.1.1 Contract with the broker

First of all the customer contacts the cloud storage broker in order to initiate cloud storage service. This broker might not necessarily provide only cloud storage services, in addition the broker might include XaaS, PaaS, and IaaS in its portfolio. In this scenario we assume that the broker provides only cloud storage service.

Since, several new CSPs are entering the market every year its becoming hard for cloud service customers to choose the best CSP for their needs. A cloud brokerage service comes into play at this stage, communicating with the CSPs on the behalf of the customer. In this scenario each CSP has different properties and offers diverse QoS, thus the pricing of the cloud storage may differ. The broker examines the requirements of the service which are given by the customer and moves its data to the most suitable CSP.

The broker is capable of measuring service usage and supervising pricing and billing. In this scenario the broker has characteristics of an intermediation

brokerage service (see section 2.3.1). It may appear as a discrete service placed between the CSPs and the customer. Figure 4.1 illustrates this.

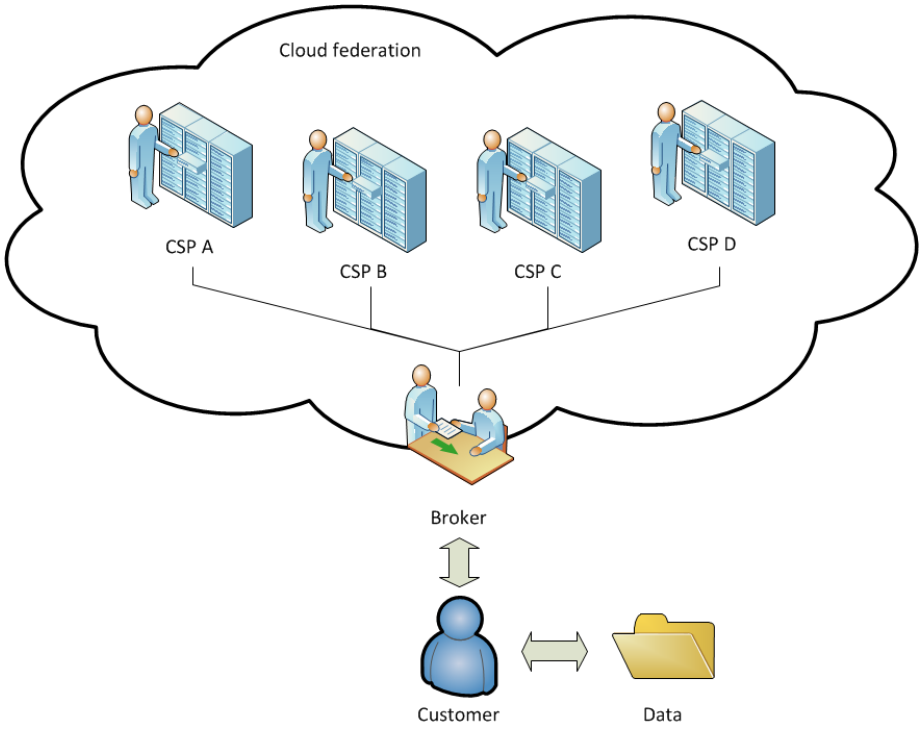

Figure 4.1: Customer contacts a broker

The CSPs in this scenario are part of a cloud federation and each CSP provides a cloud storage service to the broker. These services have different SLAs, hence the broker can offer the customer the best service satisfying certain requirements. In this scenario the broker chose to store the customer's data in the service provided by the CSP A.

4.1.2 Synchronization with CSP A

CSP A provides a basic data storage service which is accessible through a proprietary API. This cloud storage service does not offer any encryption for data at rest and relies only on a transport layer security (i.e. TLS or SSL). The broker initiates data synchronization between the CSP A and the customer. During this transfer the data is encrypted and decrypted when it reaches the CSP. Figure 4.2 illustrates this process.

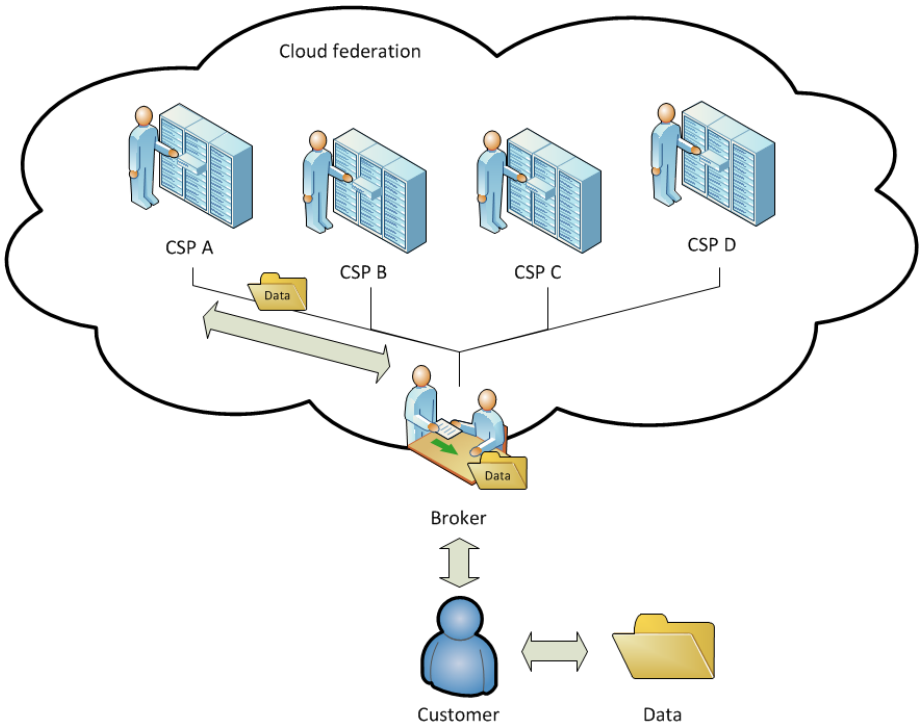

Figure 4.2: Broker initiates the synchronization of the customer's data

The customers are unaware that their data is sent to a CSP as they think that the service is provided directly by the broker. However, the broker does not have a sufficient storage capacity to store customer's data and therefore intermediates with the CSPs in the cloud federation and with the customer.

4.1.3 Data migration

Further in this scenario the customer gets acquainted with the cloud model and likes the benefits which this model brings. However, the customer's requirements for the cloud storage service have changed as he intends to store some sensitive data to the cloud. Since the customer is not familiar with encryption, the broker suggests that it provide all the security that is needed by the customer after he upgrades his service.

The broker decides to migrate the customer's data from the CSP A to the CSP D (see Figure 4.3). In this case the encryption mechanism offered by the CSP D

uses the broker as a secret key escrow entity.

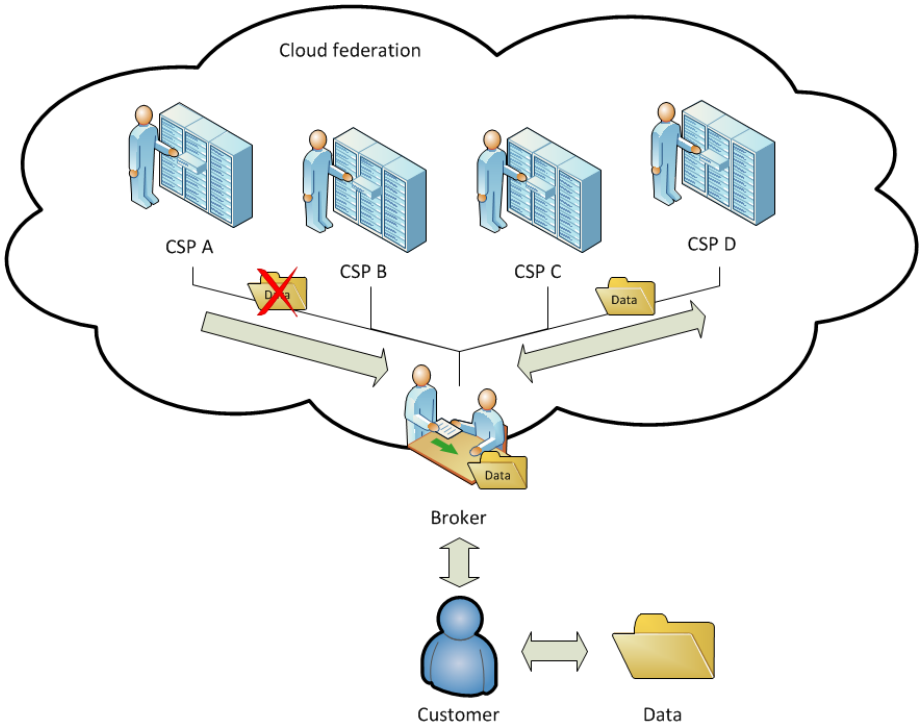

Figure 4.3: Data migration

At this stage the customer's data is encrypted, however a secret key used to encrypt and decrypt data is exposed to three entities (customer, broker, and CSP). Of course this could be limited to two entities (customer and broker) if the broker or the customer did the encryption, however in this scenario we assume that encryption capabilities are located in the CSP.

4.1.4 Changing a broker

This last stage of the scenario illustrates that due to the lack of standardized APIs and overall interoperability changing a cloud storage provider might become a troublesome objective. If trust in a CSP is lost, then the broker initiates data migration within its federated cloud as described in section 4.1.3. A problem appears when a broker is no longer trusted.

There might be many reasons why a customer wants to change their broker. Of course there is also a possibility that a customer will change a broker without

cause. In any case there should be procedures to securely migrate data from one broker to another. In this scenario the customer's data is not stored directly within the broker, in fact it is stored in one of the CSPs in a federated cloud.

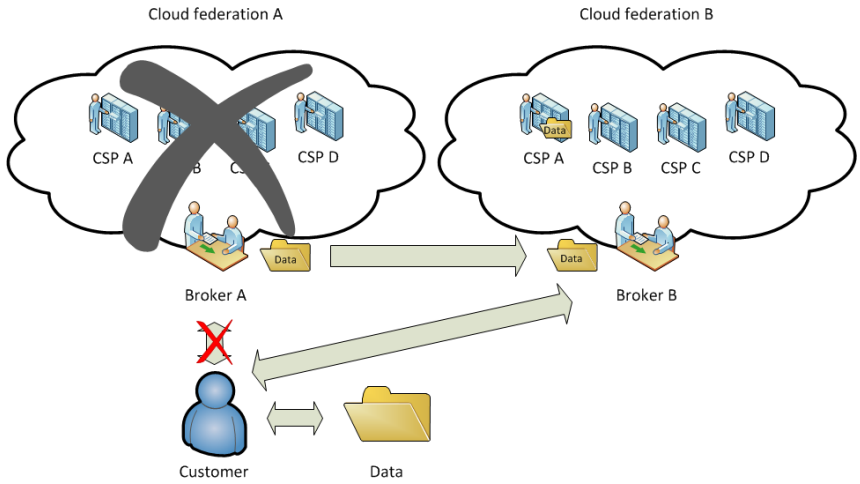

Figure 4.4: Migrating data between federated clouds

First the customer initiates the end of a contract as he no longer trusts the broker. Ideally all customer's data should be moved (copied and deleted afterwards) from the broker A to the broker B (see Figure 4.4). However in reality if the broker B is using incompatible proprietary APIs it is most likely that broker A will not bother to migrate data and will leave this task to the customer. This problem must be addressed along with the problem of proper data sanitization after the user's contract terminates.

4.2 Design criteria

This section identifies criteria for the data protection requirements which will lead to a scheme of a secure cloud storage service. There are three main categories of the criteria - functional, security, and privacy requirements.

4.2.1 Functional requirements

Functional requirements for a secure cloud storage service are straightforward:

1. the service should be able to store the user's data;
2. the data should be accessible through any devices connected to the internet;
3. the service should be capable to synchronize the user's data between multiple devices (notebooks, smart phones, etc.);
4. the service should preserve all historical changes (versioning);
5. data should be shareable with other users;
6. the service should support SSO; and
7. the service should be interoperable with other cloud storage services, enabling data migration from one CSP to another.

4.2.2 Security requirements

The key security risk in a cloud storage service is unauthorized access to the data. Successfully authenticating users and managing their permissions is a crucial requirement towards realizing a secure cloud storage service. In addition, data encryption and a proper secret key management are essential to meet the data protection requirements.

Our proposed secure cloud storage service design will be based on the following security requirements:

1. the service should support zero knowledge encryption;
2. the service should support user centric access control;
3. the service should have the ability to quickly provision and de-provision access to cloud assets;
4. the service should support dynamic trust propagation and dynamic authorization;
5. the service should be transparent to a user (thus security measures applied in the service should be available to cloud users and not rely on security through obscurity);

6. data should be properly sanitized after the end of a user's contract; and
7. the service should provide data backup and redundancy (disaster recovery).

4.2.3 Privacy requirements

The criteria for privacy requirements are based on the eight data protection principles which are defined in the UK Data Protection Act of 1998 (see section 3.3.5). These principles are in essence a code of good practice for processing personal data. Table 4.1 summarizes these eight data protection principles and assess the importance of each principle to a cloud storage service.

Table 4.1: Summary of eight data protection principles

Principle	Description	Relevant to cloud storage?
1	Personal data shall be processed fairly and lawfully	Yes
2	Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.	Yes
3	Personal data shall be adequate, relevant, and not excessive in relation to the purpose or purposes for which they are processed.	No
4	Personal data shall be accurate and, where necessary, kept up to date.	No
5	Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.	Yes
6	Personal data shall be processed in accordance with the rights of data subjects under this Act.	No
7	Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.	**Yes**
8	Personal data shall not be transferred to a country or territory outside the European Economic Area, unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.	**Yes**

In order to make privacy requirements for our service, we have to distinguish whether these principles may be applied to a cloud storage. Principles 3,4, and 6 are less related or non-related to cloud storage services as compared to other cloud based services, such as Facebook where sensitive personal data is of a bigger importance. We can summarize the criteria for privacy requirements in a cloud storage service:

1. the cloud storage service should process entrusted data fairly and lawfully;
2. not further process the data for purposes other than storing the data and backing it up;
3. retained data should not be altered;
4. backups of data should not be kept for longer then necessary;
5. access to the retained data should be monitored and controlled;
6. data should be protected against accidental loss or destruction (backed up); and
7. retained data should have an adequate protection (encryption).

4.3 Comparison of the alternatives

Three cloud storage services (Amazon's Cloud Drive, Dropbox, and SpiderOak) were assessed to evaluate whether they meet the functional, security, and privacy criteria described above. These services were chosen in order to represent 3 levels of security and privacy (see section 2.2.2). Tables 4.2, 4.3, and 4.4 summarize our findings.

4.3.1 Amazon's Cloud Drive assessment

As shown in Table 4.2, Amazon's Cloud Drive offers only the most basic functionality - it is capable of storing data, synchronize it within multiple devices, and the service is accessible on devices which are connected to the internet. In addition, the service provides data recovery and is partly transparent to the user, since it does not offer any kind of encryption or other security measures, which for instance are available in Dropbox and SpiderOak. We were not able to assess most of the criteria for privacy requirements as that information was not publicly available.

Table 4.2: Assessment of Amazon's Cloud Drive

#	Functional	Security	Privacy
1	Yes	No	?
2	Yes	No	?
3	Yes	No	?
4	No	No	?
5	No	Partly	?
6	No	?	Yes
7	No	Yes	No

4.3.2 Dropbox assessment

Table 4.3 summarizes Dropbox assessment according to our criteria. In addition to Amazon's Cloud Drive's functionalities, Dropbox offers data sharing and access control of this shared data. In addition, Dropbox preserves all historical changes and offers data encryption. Although data are encrypted and should not be altered, the encryption keys are kept by Dropbox, which enables them to read and modify encrypted data.

Table 4.3: Assessment of Dropbox

#	Functional	Security	Privacy
1	Yes	No	?
2	Yes	No	?
3	Yes	Yes	Partly
4	Yes	No	?
5	Yes	No	Yes
6	No	?	Yes
7	No	Yes	Yes

4.3.3 SpiderOak assessment

Table 4.4 shows SpiderOak assessment according to our criteria. SpiderOak, in addition to Dropbox's features provides a zero-knowledge approach, which allows SpiderOak to meet most criteria of our privacy requirements. Since SpiderOak does not keep the encryption key, it is not possible for it to alter the data without the user's consent. The zero-knowledge approach also guarantees that the stored data is processed fairly, lawfully, and only for the specified purpose (data storage), because SpiderOak is unaware of the data's content, and it is not able to read or alter the data. Moreover, SpiderOak reveals the source of most of its components to the public, which makes SpiderOak and its security transparent.

Table 4.4: Assessment of SpiderOak

#	Functional	Security	Privacy
1	Yes	Yes	Yes
2	Yes	No	Yes
3	Yes	Yes	Yes
4	Yes	No	?
5	Yes	Yes	No
6	No	?	Yes
7	No	Yes	Yes

Despite all the benefits of the zero-knowledge approach, sharing data becomes an issue. Since sharing requires revealing the password that is used for encryption, it becomes impossible to properly monitor and manage access control in such an approach.

4.3.4 Summary

Although SpiderOak meets most of the requirements for a secure cloud storage service, some criteria were not met. None of the compared services support SSO, interoperability, user centric access control, dynamic trust propagation, and dynamic authorization. In addition, an issue of proper data sanitization was not resolved by any of the compared services. There is a possibility that all services delete the user's data as recommended in the data sanitization guideline [35], however these services do not provide any proof of doing as mentioned in [35]. Another unresolved privacy issue is whether these services keep data backups for no longer then necessary. Unless CSPs start publishing such details, there will always be doubts about privacy and security.

4.4 Proposed scheme

In this thesis we chose a combination of OpenID and OAuth rather than using SAML, because OAuth has become a mainstream internet-scale SSO mechanism. While OpenID focuses on authentication, OAuth focuses on authorization. Moreover, there are OAuth API libraries for most languages commonly used in web development including Python, PHP, Ruby, Perl, Java, C#, and Objective-C. In addition, OAuth access token may be passed in the HTTP header which makes it more compatible with CDMI.

Our proposed design of a Security as a Service for cloud storage services provides both authentication and authorization. By using our service brokers may transparently allow their customers to choose a preferable OpenID identity provider (Google, Facebook, LinkedIn, etc.) leaving API implementations to the service.

We strongly believe that our proposed scheme addresses the issues stated in Section 4.1 and should meet most of the criteria listed in Section 4.2. Figure 4.5 illustrates our proposed design.

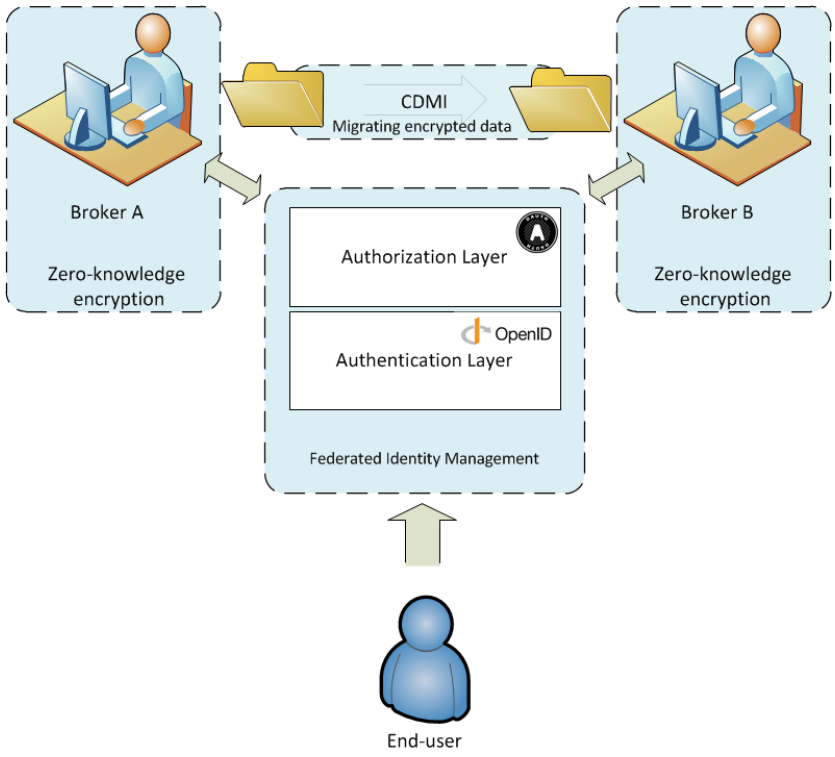

Figure 4.5: Proposed design

In order to successfully migrate data from one broker (CSP) to another, both brokers must identify the user through an abstract authentication layer. OpenID provides SSO features, hence broker B can identify the user by redirecting the authentication process to a trusted OpenID identity provider, such as Google. Of course, each broker could act as an OpenID identity provider, however trust issues may arise. Moreover, OAuth must be used in addition to OpenID, as the user would like to implement access control for the cloud storage service providers and their brokers.

4.4.1 Zero-knowledge encryption

In our design we are using a zero-knowledge approach to secure the user's data. In order to access or alter retained data users must authenticate by simply proving that they know their password without ever sending it over the network. First a user has to generate the encryption components which are shown in Table 4.5.

Table 4.5: Data generated at the user's end

Component	Description	Size	Format
salt1	A random value	32 B	plaintext
salt2	A random value	32 B	plaintext
publicKey	Serialized RSA public key	2048 B	plaintext
challengeKey	A key used in the construction of authentication challenges	32 B	plaintext
keypair	Serialized RSA keypair	2048 B	encrypted

Next, the user generates a 32 byte challenge key ("challengeKey") which is used in the zero-knowledge authentication together with a 2048 byte RSA keypair which is used to protect the secret key that encrypts the user's data. Figure 4.6 illustrates how the challenge key and RSA keypair are generated. Password-Based Key Derivation Function 2 (PBKDF2) is used to derive a key based upon a salt and a password (see [59]).

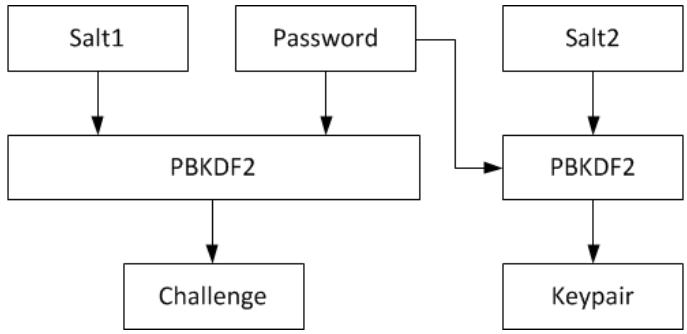

Figure 4.6: RSA keypair and challenge key creation

The cloud storage service authenticates the users without their password ever leaving their premises by first sending a challenge request containing a user's generated salt value, timestamp, and random data which is encrypted with the challenge key. The user must decrypt the random data and use it as a key to its own encrypted reply containing the timestamp. The details of this process are shown in Figure 4.7.

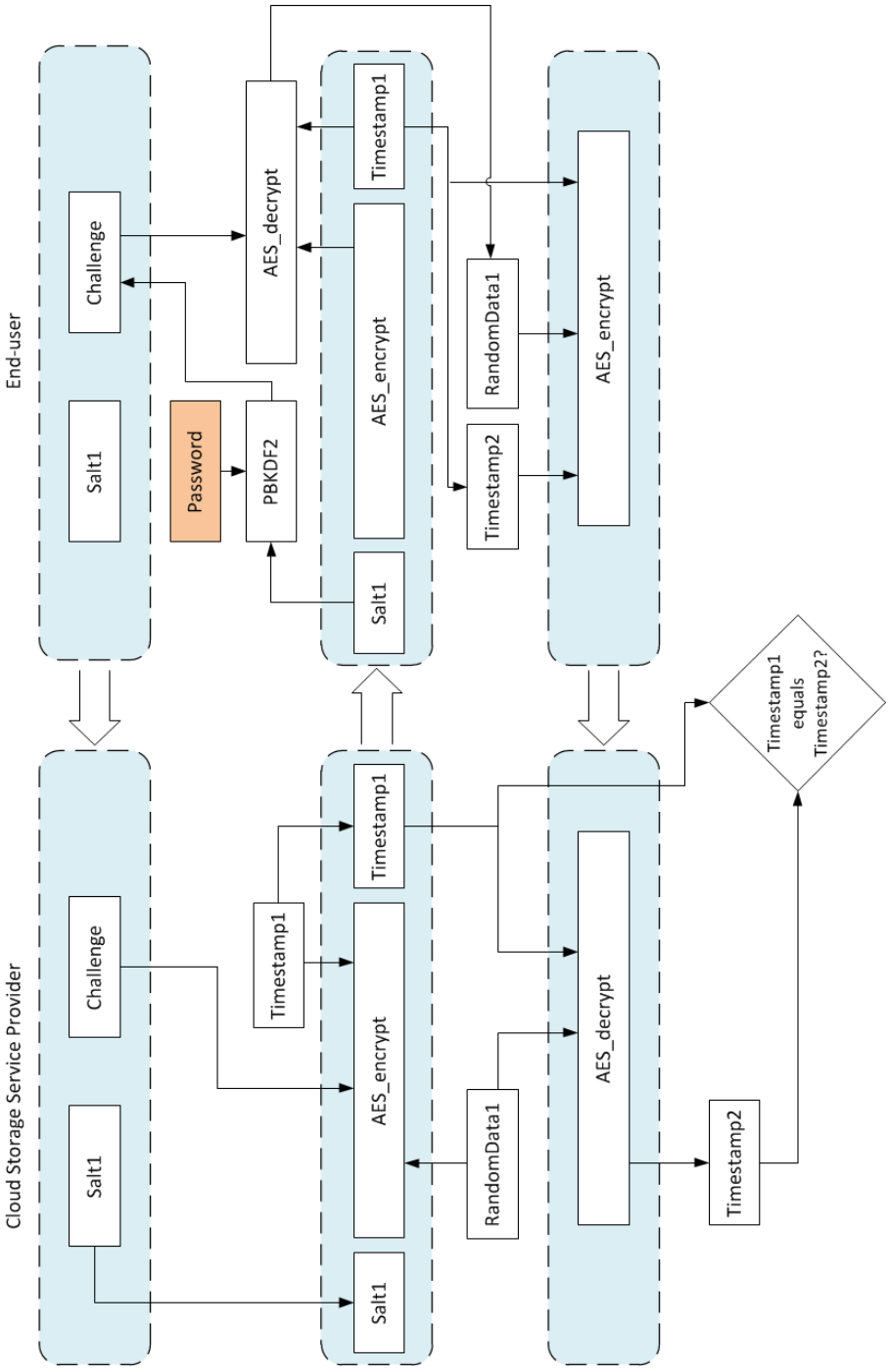

Figure 4.7: Zero-knowledge proof of knowledge

4.4.2 OpenId + OAuth

OpenID login authentication involves a sequence of interactions between a cloud storage service, the OpenID identity provider's login authentication service, and the end user. Figure 4.8 and sequence below describe the process.

1. The cloud storage service asks the end user to log in by offering a set of federated identity options (SSO).
2. The user selects the "Sign in with OpenID identity provider" option.
3. The cloud storage service sends a "discovery" request to the OpenID identity provider to get information on its login authentication endpoint.
4. The OpenID identity provider returns an eXtensible Resource Descriptor Sequence (XRDS) document, which contains the endpoint address.
5. The cloud storage service sends a login authentication request to the OpenID identity provider's endpoint address.
6. The user is redirected to the OpenID identity provider's login page and is asked to sign in.
7. The user logs in and approves the OpenID identity provider's authentication and access request.
8. The OpenID identity provider returns the user's identity and various OpenID and OAuth paramaters such as openid.return_to, openid.claimed_id, etc.
9. The cloud storage service uses the user's identifier which was provided by the OpenID identity provider and uses the request token to continue the OAuth sequence and to gain access to the user's services provided by the OpenID identity provider.

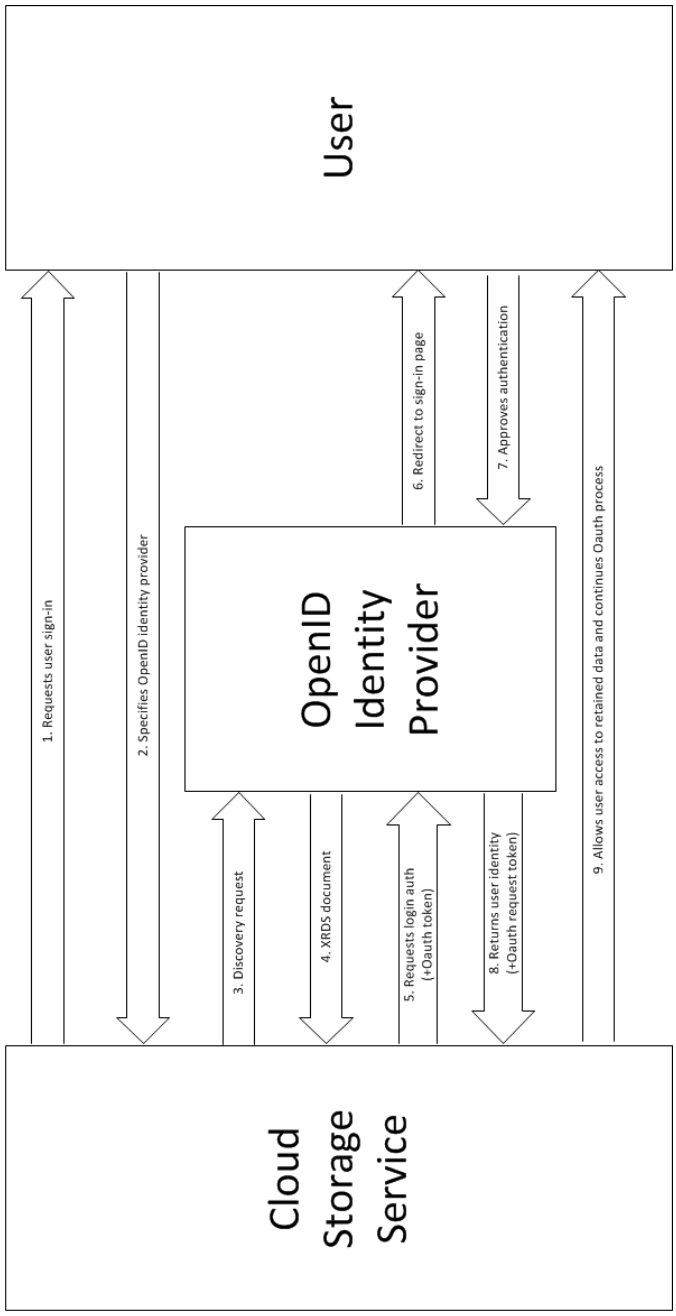

Figure 4.8: OpenID login authentication sequence (adapted from [58])

Chapter 5

Results

We have assumed that our proposed design is an enhancement to a currently active cloud storage service such as SpiderOak, rather than a completely new service. Our proposed design has focused on the areas of security, privacy, and interoperability. Table 5.1 shows which criteria (see Section 4.2) are met by our proposed design.

Table 5.1: Assessment of the proposed design

#	Functional	Security	Privacy
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	Yes	Yes
4	Yes	Yes	?
5	Yes	Yes	Yes
6	Yes	?	Yes
7	Yes	Yes	Yes

Compared to SpiderOak our proposed design additionally supports SSO and also is interoperable with other cloud storage service providers. In addition, our scheme supports user centric access control, dynamic trust propagation and dynamic authorization, and access to retained data is monitored and controlled.

However, some criteria were not met. Our proposed design does not affect what happens to the user's data after the end of a contract and whether backups are not kept for longer than necessary. Although it might be possible to properly delete the user's data by overwriting the user's storage space with random data after a

successful migration to another CSP, it is still unclear how to enforce proper data sanitization across different sites. Most CSPs backup users' data in different data centers, not to mention various backup media such as tape backups. However, the backups of user's data that was encrypted using a zero-knowledge approach, cause little to no risk that this data might cause harm to an individuals and their privacy.

Chapter 6

Discussion

OAuth enables service providers to share data in a secure, revocable, limited in scope, more private way. In addition OAuth encourages good user behaviour, as if a user violates the TOS of a service provider the consequences will be delegated to other service providers as well. However, OAuth does not attempt to solve other problems that can arise such as privacy policy management or data duplication and skew. Moreover, OAuth faces difficulties in propagation and development of trust relationships among service providers.

OAuth provides interoperable representation of the user's access rights and privileges. This simplifies the process of the cloud storage service development as there is no need to create expensive and customized syntactic translation components.

Our proposed design uses a hybrid protocol (OpenID + OAuth) in order to identify users and authorize them to access their data. In contrast, if only OAuth was used, the service would not really care who the user is. However, our proposed service securely redirects users to their preferable OpenID identity providers (Google, Facebook, LinkedIn, etc.), guarantees their privacy, and meets EU data protection requirements.

With OpenID our service makes it possible to maintain discreet profiles for logging in to other cloud storage providers across a cloud federation without needing a different password for each one of them. In addition, brokers and CSPs no longer have to store their users' passwords so the responsibility for a security breach (if any) would be taken by our proposed service. Moreover, our service focuses on data privacy and security rather than on the basic functionalities of a cloud storage service (store, delete data, etc.), thus security and privacy protection should be state of the art. In addition our proposed service maintains interoperability between CSPs in federated clouds.

Most countries have laws and legislation that requires organizations to protect

their customers' data. Zero-knowledge encryption assures customers that the confidentiality of their data is preserved irrespective of the actions of the CSP or the broker. As data is only stored in encrypted form, any law that concerns stored data has little to no effect on the customer. This reduces legal exposure for the customer and allows the CSP to make optimal use of its storage infrastructure, thereby reducing costs.

If an organization becomes the suspect of an investigation, legal authorities may send a subpoena requesting a CSP to give access to the organization's data. Customers of this organization may not be informed of such data intrusion. In our proposed scheme customers' data is encrypted and the CSP can not provide access for government agencies to that data since only the customer has the secret key. Thus, any request for the data must be made directly to the customer, as in [52]. Note that the CSP may (in some countries) be compelled to provide a copy of the encrypted data in order to preserve evidence or enable a government agencies to attempt to decrypt the data themselves.

Chapter 7

Conclusions and future work

7.1 Conclusions

The aim of this thesis was to identify cloud storage security and privacy risks and propose a Security as a Service design which could securely migrate data from one CSP to another. The motivation behind this research lies in the fact that for many organizations the final barrier to adopting Cloud computing is whether it is sufficiently secure.

This thesis project examined EU data protection requirements, federated identity management, and security and privacy risks and then compared them to traditional IT solutions. An important conclusion of my work is that the cloud is inherently neither secure nor insecure. The most important factor that a CSP can provide is the quality of management applied - just as this is the most important factor in any IT environment.

After analysing cloud storage security and privacy risks, EU data protection requirements, and security applied to current cloud storage services (Amazon's Cloud Drive, Dropbox, and SpiderOak) we came up with our proposed design of the Security as a Service for cloud storage services. The following are the key aspects of our proposed design:

- zero-knowledge encryption;
- SSO;
- interoperability with other CSPs;
- user centric access control; and
- dynamic trust propagation and authorization;

Our proposed design of the service meets most of the defined security and privacy requirements (see section 4.2).

7.2 Future work

The service uses OpenID for authentication, OAuth for authorization, and CDMI to manage the user's data. However, it is unknown how these three technologies work together, therefore a proof of concept is required to prove that our design could realize the properties that we claim. It is already known that OAuth and OpenID work well together. The only question is whether CDMI will operate well when used together with OAuth and OpenID. Since OAuth may send access tokens to an API in the HTTP Authorization header, we strongly believe that OAuth should be able to pass these tokens to the functional CDMI interface and therefore authorize data access.

While our proposed design meets most of the defined security and privacy requirements of section 4.2, it is still unknown how to properly handle data sanitization, to meet data protection requirements, and provide users data recovery capabilities (backups, versioning, etc.). Although there are guidelines for proper data sanitization, in practise it becomes a troublesome objective to follow these guidelines, especially in the context of a CSP. Further research should be conducted and solutions should be proposed to address this matter.

Bibliography

[1] Peter Mell and Timothy Grance. The NIST Definition of Cloud Computing. Special Publication 800-145, National Institute of Standards and Technology, Information Technology Laboratory, September 2011. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[2] SNIA, Advanced Storage a Information Technology. Implementing, Serving,and Using Cloud Storage. Cloud Storage Initiative. October 2010, http://www.snia.org/sites/default/files/2010-10-WP-ImplementingServingandUsingTheCloud.pdf

[3] Spencer Piggott. BBC Internet Blog. BBC Technology Strategy. , January 2010, http://www.bbc.co.uk/blogs/bbcinternet/2010/01/bbc_technology_strategy_princi.html

[4] Geoff Brumfiel. High-energy physics: Down the petabyte highway. Nature 469: pp. 282283, doi:10.1038/469282a. January 19th, 2011, http://www.nature.com/news/2011/110119/pdf/469282a.pdf

[5] Cloud Security Alliance. Security guidance for critical areas of focus in cloud computing. Version 3. December 2011, https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf

[6] European Network and Information Security Agency (ENISA). Benefits, risks and recommendations for information security. November 2009, http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment/at_download/fullReport

[7] Derek Newton. Dropbox authentication: insecure by design. Dereknewton.com Information security insights. Accessed May 20th, 2012, http://dereknewton.com/2011/04/dropbox-authentication-static-host-ids/

[8] Melanie Pinola. How To Add a Second Layer of Encryption to Dropbox. Lifehacker. June 20th, 2011, http://lifehacker.com/5794486/how-to-add-a-second-layer-of-encryption-to-dropbox.

[9] Mike Masnick. SOPA Can Impact Companies Who Think They're Immune. TechDirt blog. December 2011, http://www.techdirt.com/blog/?company=spideroak.

[10] Kimpl. Website dedicated to quality and useful information on privacy, security, software, and internet. Most secure online backup and file sync service. December 2011. http://www.kimpl.com/847/secure-online-cloud-storage-service/

[11] SpiderOak website. Accessed May 20th, 2012, https://spideroak.com/

[12] Dropbox website. Accessed May 20th, 2012, https://www.dropbox.com/

[13] Amazon S3 website. Accessed May 20th, 2012, http://aws.amazon.com/s3/

[14] Bruce Schneier. Dropbox Security. Schneier on Security. May 23, 2011, http://www.schneier.com/blog/archives/2011/05/dropbox_securit.html

[15] Steven J. Vaughan-Nichols. No Privacy on Amazons Cloud Drive. ZDNet networking Blog. March 2011, http://www.zdnet.com/blog/networking/no-privacy-on-amazons-cloud-drive/882

[16] Abdullah Azfar. Multiple Escrow Agents in VoIP. Master thesis. KTH Royal Institute of Technology, School of Information and Communication Technology, Stockholm, Sweden, TRITA-ICT-EX-2010:109, June 2010, http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-91102

[17] Krishnan Subramanian. Defining Federated Cloud Ecosystems. October 2011, http://krishworld.com/?p=1285

[18] Krishnan Subramanian. The Cloud Is Rockin' & Rollin' In. March 2011, http://www.slideshare.net/krishnan/the-cloud-is-rockin-and-rollin-in

[19] Bill Claybrook. Cloud interoperability: Problems and best practices. Computerworld. June 2011, http://www.computerworld.com/s/article/9217158/Cloud_interoperability_Problems_and_best_practices

[20] Techtarget. Federated cloud (cloud federation) definition. Accessed February 26th, 2012, http://whatis.techtarget.com/definition/federated-cloud--cloud-federation-.html

[21] Arjuna. Service Agreements in the Cloud. October 2011, http://blog.arjuna.com/2011/10/11/service-agreements-in-the-cloud/

[23] Information Commissioner's Office. Guide to data protection definitions, principles, and practical examples. Accessed April 14th, 2012, http://www.ico.gov.uk/for_organisations/data_protection/the_guide.aspx

[22] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. October 24th, 1995, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

[24] UK Data Protection Act 1998. July 16th, 1998, http://www.legislation.gov.uk/ukpga/1998/29/enacted

[25] Eduardo Ustaran. Internation Association of Privacy Professionals. Global Privacy Dispatches - UK Data Minimisation. July 2008, https://www.privacyassociation.org/publications/2008_07_global_privacy_dispatches_uk_data_minimisation

[26] Safe Harbour Agreement. US Department of Commerce website. Accessed May 12th, 2012, http://www.export.gov/safeharbor/eu/eg_main_018476.asp

[27] UK Regulation of Investigatory Powers Act 2000. Accessed May 12th, 2012, http://www.legislation.gov.uk/ukpga/2000/23/contents/enacted

[28] Privacy Issues related to Cloud Computing. Office of the Privacy Commissioner of Canada. March 2010, http://www.priv.gc.ca/information/pub/cc_201003_e.asp

[29] Member countries of the EEA. Leonardo da Vinci - A European Vocational Training Initiative. Accessed May 13th, 2012, http://www.leonardo.ar.tum.de/english/countries.html

[30] Commission decisions on the adequacy of the protection of personal data in third countries. European Commission's data protection website. Accessed May 13th, 2012, http://ec.europa.eu/justice/policies/privacy/thridcountries/index_en.htm

[31] Binding corporate rules. European Commission's data protection website. Accessed May 13th, 2012, http://ec.europa.eu/justice/policies/privacy/binding_rules/index_en.htm

[32] EC model clauses. European Commission's data protection website. Accessed May 13th, 2012, http://ec.europa.eu/justice/policies/privacy/modelcontracts/index_en.htm

[33] Thomas Ristenpart, et al. Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds. ACM CCS'09,Proceedings of the 16th ACM conference on Computer and communications security, November 2009, http://cseweb.ucsd.edu/~hovav/dist/cloudsec.pdf

[34] Wayne Jansen and Timothy Grance. Guidelines on Security and Privacy in Public Cloud Computing. NIST Special Publication 800-144. December 2011, http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf

[35] Richard Kissel, et al. Guidelines for Media Sanitization. NIST Special Publication 800-88. September 2006, http://csrc.nist.gov/publications/nistpubs/800-88/NISTSP800-88_rev1.pdf

[36] A. Freier, et al. The Secure Sockets Layer (SSL) Protocol Version 3.0. Internet Engineering Task Force (IETF). Request for Comments: 6101. August 2011, http://tools.ietf.org/html/rfc6101

[37] Daily Mail Reporter. Computer hard drive sold on eBay had details of top secret U.S. missile defence system. May 2009, http://www.dailymail.co.uk/news/article-1178239/Computer-hard-drive-sold-eBay-details-secret-U-S-missile-defence-system.html

[38] Lucas Mearian. Computerworld. Survey: 40% of hard drives bought on eBay hold personal, corporate data. February 10, 2009, http://www.computerworld.com/s/article/9127717/Survey_40_of_hard_drives_bought_on_eBay_hold_personal_corporate_data

[39] Ebay Online Shopping. Accessed May 20th, 2012, http://www.ebay.com/

[40] Simson Garfinkel, An Evaluation of Amazons Grid Computing Services: EC2, S3 and SQS. Technical Report TR-08-07. Center for Research on Computation and Society. School for Engineering and Applied Sciences, Harvard University. July 2007, http://simson.net/clips/academic/2007.Harvard.S3.pdf.

[41] Twitter Email Account Hack Highlights Cloud Dangers. Infosecurity Magazine, July 23, 2009, http://www.infosecurity-magazine.com/view/2668/twitter-emailaccount-hack-highlights-cloud-dangers-/

[42] John D. Sutter. Twitter Hack Raises Questions about Cloud Computing. CNN. July 16, 2009, http://edition.cnn.com/2009/TECH/07/16/twitter.hack/

[43] Twitter. Social networking platform. Accessed May 20th, 2012. http://www.twitter.com

[44] Physical Security. Metro Data Center LLC. Accessed May 16th, 2012. http://metrodatacenter.com/data-center/physical-security/

[45] Bruce Schneier. The Risks of Key Recovery, Key Escrow, and Trusted Third-Party Encryption. May 27th, 1997, http://www.schneier.com/paper-key-escrow.pdf

[46] T. Dierks and E. Rescorla. The Transport Layer Security (TLS) Protocol Version 1.2. Network Working Group. RFC Editor, RFC 5246. issn 2070-1721. August 2008, http://tools.ietf.org/html/rfc5246

[47] Educause. 7 things you should know about federated identity management. September 2009, http://net.educause.edu/ir/library/pdf/EST0903.pdf

[48] G. Q. Maguire Jr., Personal communications, May 18th, 2012.

[49] Melvin Greer and Lockheed Martin. Cloud Standards Customer Council. The Practical Guide to Cloud Service Level Agreements. April 10, 2012, http://www.cloudstandardscustomercouncil.org/PGCloudSLA040512MGreer.pdf

[50] Cloud Standards Customer Council. Cloud Computing Use Cases Version 1.0. October, 2011, http://www.cloudstandardscustomercouncil.org/use-cases/CloudComputingUseCases.pdf

[51] Maureen Rogers. Alert Logic. Removing the cloud of insecurity. State of cloud security report. Spring 2012, http://www.alertlogic.com/wp-content/uploads/alertlogic%20state%20of%20cloud%20security%20spring2012.pdf

[52] John Leyden. Animal rights activist hit with RIPA key decrypt demand. The register. November 14th, 2007, http://www.theregister.co.uk/2007/11/14/ripa_encryption_key_notice/

[53] C. Basescu, et al. Managing Data Access on Clouds: A Generic Framework for Enforcing Security Policies. Advanced Information Networking and Applications (AINA), 2011 IEEE International Conference on. Pages 459-466. March 2011, http://hal.inria.fr/docs/00/53/66/03/PDF/GenericFrameworkForEnforcingSecurityPolicies.pdf

[54] Sherman Chow, et al. Dynamic Secure Cloud Storage with Provenance. Cryptography and Security: From Theory to Applications. Lecture Notes in Computer Science. Pages 442-464, volume 6805/2012. Accessed May 20th, 2012, http://dx.doi.org/10.1007/978-3-642-28368-0_28

[55] Yanjiang Yang and Youcheng Zhang. A Generic Scheme for Secure Data Sharing in Cloud. Parallel Processing Workshops (ICPPW), 2011 40th International Conference. Pages 145-153. September 2011,

[56] David Kravets. Judge Wont Purge Megaupload User Data, At Least Not Yet. Wired.com, April 13th, 2012, http://www.wired.com/threatlevel/2012/04/megaupload-data-flap/

[57] E. Hammer, Ed., et al. The OAuth 2.0 Authorization Framework draft-ietf-oauth-v2-26. Network Working Group. RFC 5849. May 1, 2012. http://tools.ietf.org/html/draft-ietf-oauth-v2-26

[58] Google developers. Federated Login for Google Account Users. Accessed June 3rd, 2012.https://developers.google.com/accounts/docs/OpenID

[59] S. Josefsson. PKCS #5: Password-Based Key Derivation Function 2 (PBKDF2) Test Vectors. Internet Request for Comments. Issn = 2070-1721, volume = RFC 6070 (Informational). RFC Editor. Jan 2011. http://www.rfc-editor.org/rfc/rfc6070.txt

[60] SNIA, Advanced Storage a Information Technology. Cloud Data Management Interface (CDMI). Technical Position Version 1.0.1. September 15, 2011, http://snia.org/sites/default/files/CDMI_SNIA_Architecture_v1.0.1.pdf

[61] Fang Liu, et al. NIST Cloud Computing Reference Architecture. Recommendations of the National Institute of Standards and Technology. Special Publication 500-292. September 2011, http://collaborate.nist.gov/twiki-cloud-computing/pub/CloudComputing/ReferenceArchitectureTaxonomy/NIST_SP_500-292_-_090611.pdf

Appendix A

Thesis Problem Description

Master Thesis Problem Description

Vytautas Zapolskas
15th February, 2012

As more companies turn to cloud solutions, securing cloud based services becomes increasingly important, because for many organizations, the final barrier to adopting Cloud computing is whether it is sufficiently secure.

This research is limited to data protection risks in cases of storing and transferring sensitive data between clouds. The student will design a service which could provide Security as a Service for cloud brokers and carriers in a federated cloud allowing customers to securely migrate from one provider to another. Such service would utilize various encryption techniques and also include identity and key management mechanisms, such as "federated identity management".

To support the design of the service the study will also
- identify most important Cloud Storage specific risks and compare them with traditional solutions, such as server-based model.
- describe data protection requirements for cloud storage services.

Supervisor: Professor Danilo Gligoroski
External supervisor: Fredrik Solsvik, Telenor ASA