Departament de Llenguatges i Sistemes Informatics

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Master in Computing

Master of Science Thesis

Privacy and Confidentiality issues in Cloud Computing architectures

David Jiménez Martínez

Advisor/s: Drs. Cristina Gómez Seoane
Dr. Xavier Franch Gutiérrez

09/09/2013

Acknowledgements

I would wish to be grateful to all people that supported me while working on this master thesis, whose input has proven invaluable to accomplish the defined objectives.

I'm thankful to my advisors, Drs. Cristina Gómez Seoane and Dr. Xavier Franch Gutiérrez, for encouraging me, guiding and helping me through the development of this Master Thesis by providing starting bibliography and tips, and for continuously reviewing my work. Our discussions have made this master thesis possible.

I would like to give a special mention to Christoph Fischer, a student of Business Information Technology from the Zurich University of Applied Sciences (Switzerland) who was working on a project about non-functional requirements in Cloud Computing, with who we shared our gathered bibliography about the topic, tips and with who I could prepare some questions for interviewing through him several people at the Swiss Cloud Conference 2012.

Also, I would like to thank to my dear friend, Noemí, for sharing working experiences with me and infuse me with motivation.

Finally, I am specially grateful to my girlfriend Judit and my mother for being patient with me, giving me wise tips, support and motivate me while working on this master thesis, making my days better on these tough economic times.

Index of Contents

Index of figures

Index of tables

CHAPTER 1

INTRODUCTION

Chapter 1: Introduction

This document represents the Master Thesis of the Master in Computing, at Barcelona School of Informatics (Facultat d'Informàtica de Barcelona) of the Technical University of Catalonia (Universitat Politècnica de Catalunya).

Cloud computing is a computing paradigm in which organizations can store their data remotely in the cloud (Internet) and access applications, services and infrastructure on-demand from a shared pool of computing resources. It is clear that cloud technologies have proven a major commercial success over recent years (since the appearance of products and cloud offerings like Amazon EC2 [75] and Microsoft Azure [65]). According to Gardner, Cloud Computing will play a large part in the ICT (Information and Communication Technologies) domain over the next 10 years or more, since it provides cost-savings to enterprises thanks to virtualization technologies, opening gates for new business opportunities as well.

However, Cloud Computing has to face several challenges and issues. Storing and processing data out of the boundaries of your company raises security and privacy concerns by itself. Nowadays information is the commodity of XXI century, and certain information can mean power and market advantage. As pointed out by Andreas Weiss, Director of the EuroCloud, in an interview we held with him (refer to Appendix C), data is one of the most important and valuable resource any company has. Therefore, security mechanisms to protect this data are necessary to make the right choices and decisions for the company without worrying about data safety. In the paradigm of Cloud Computing we will have to trust a Cloud Service Provider (CSP), creating an extra dependency to a third party which some customers, depending on the value of their data, will inevitably feel uncomfortable. Outsourcing business data in a place not owned by oneself can scare organizations from using the benefits of Cloud Computing in an optimal way.

1.1. Motivation

Privacy and confidentiality (refer to section 3.1 and 3.2 for definitions) are continuous hot topics. As of 2013, year of publication of this Master Thesis, media is flooded with a stream of news related to privacy and confidentiality issues and confrontations.

Just to enumerate some examples, one of the most relevant topics is the case of Edward Snowden, an ex-agent of the United States NSA (National Security Agency) who leaked information and details of several top-secret U.S. surveillance and espionage programs to the press. He defends that he did this to inform the public about what is done in their name and for the purpose of transparency. Since June 2013, he is wanted for the charges of "*theft of government property, unauthorized communication of national defense information, and willful communication of classified intelligence to an unauthorized person*". He is both called a hero and a traitor by sectors of population. Some related news are [61], [62].

Another relevant news can be read in [59], in which Google, as a defense against an action lawsuit for the lack of privacy from Gmail users, declared that whoever uses a service like Gmail should not expect privacy for an information that voluntarily is delivered to third-parties. Google explained that Gmail Terms of Service (ToS) and Privacy Policies (refer to section 2.5.2 for definitions) presents its automated mail process analysis and that the user, when he accepts the use of the service, he is also obliged to accept these terms. A related news, in the same line, about Facebook is [63].

As a last example, in [60] we can read on a popular Spanish computing magazine that Sweden forbids the use of Google Apps in the public sector, arguing that Google Apps fails to protect users' privacy and questions the equality of conditions in the use of Google Services.

As it can be noted, there are a hefty amount of issues and concerns around privacy and confidentiality of data, and they are a continuous subject of controversy in our society which raises questions difficult to answer due to multiple moral implications [2]. Google Apps, Facebook, and other software solutions delivered through Internet are considered Cloud solutions and are not exempt of these concerns. Some examples of these questions that may arise from Cloud customers: "How secure is my data? Can I trust my cloud provider? Which are the risks and mitigations for any existing issue on my data in the cloud?". Privacy is a high concern in the security requirement of Cloud Computing. As we will see throughout this document, even though a lot of effort has been put from governments, standard organizations and cloud industry into infusing trust and attract more customers to use Cloud services, Cloud technologies and models have not yet reached their full potential and have not yet acquired a degree to satisfy all potential circumstances of usage.

For these reasons, this Master Thesis aims at providing a state-of-the-art about Cloud Computing security, focusing on privacy and confidentiality matters which are the most significant ones. It will serve as a good starting point for readers interested in getting general knowledge about Cloud Computing and researchers interested in contributing on Cloud Computing privacy and confidentiality matters, allowing them to acknowledge the current status regarding privacy of this trendy paradigm. In this document we identify the issues and challenges privacy and confidentiality has to face and provide some basis about several relevant researches about the topic.

More specifically, in this Master Thesis we make use of several computing research sources to look for definitions and research about Cloud Computing platforms and put all the pieces together in order to provide a deep understanding about privacy and confidentiality requirements, issues and challenges.

1.2. Thesis objectives and organization

The objective and main contribution of this master thesis consists in elaborating a state-of-the-art about the privacy and confidentiality requirements, issues and challenges applied to the Cloud Computing paradigm, identify the existing evidence on this topic and establish relationships among works to find gaps and conflicting areas.

In more detail, this document is structured as follows:

1) Chapter 2 describes background information and definitions about the Cloud Computing paradigm, with the aim of providing basic knowledge to fully understand the concepts described in the following chapters.

2) Chapter 3 explains which has been the research methodology used in this document to collect data for the elaboration of the Systematic Literature Review and the results of its application.

3) Chapter 4 explains the performed Systematic Literature Review. The chapter describes definitions about privacy and confidentiality applied to the paradigm of Cloud Computing, classifies the issues and challenges found in the literature and outlines the solutions found for these issues or research gaps in the different identified areas.

4) Chapter 5 exposes conclusions we obtained in the Systematic Literature Review through an analysis and outlines the research gaps, open areas and issues we identified.

5) Chapter 6 exposes final remarks about the work done for this Master Thesis and describes future work that could be done related to what we have explained so far.

6) Additionally, appendixes have been included with glossary, a short overview of different Cloud vendors and an interview with the director of EuroCloud, Andreas Weiss.

In this sense, the contributions made in this Master Thesis will benefit the Cloud Computing community, customers and Cloud industry. Through this document:

- Cloud customers will be able to grasp a better idea about privacy and data confidentiality protection technologies available in Cloud Computing paradigm as well as the existing issues, which will allow them to plan ahead, make better decisions and gain trustiness in the cloud.
- Researchers interested in the field of privacy and confidentiality in Cloud Computing will be able to acknowledge research areas and challenges in privacy and confidentiality which may need further work.
- Cloud industry will be able to identify existing issues and vulnerabilities regarding the protection of customer's data privacy and offered the chance to follow researchers' suggestions to provide better services to build up trust and increase the amount of Cloud adopters.

1.3. Starting point

Cloud Computing is a concept which has been around for some time. A hefty amount of work, reviews and research have been done in the topic of Cloud Computing, specially on concerns regarding security in the Cloud. So, in order to get a starting basis and begin the work of this Master Thesis, we have read papers which review concepts about Cloud Computing like those defined in the references, e.g. [1] and [10]. However, as Cloud Computing is a fairly recent paradigm of interest, research is disperse and there exists gaps which must still be covered in order to make cloud a full and complete reality. Bringing all the knowledge together in an organized way will help to better understand the issues that revolve around this paradigm.

In addition, some works used in this Master Thesis like [SLR16], [SLR19] or [SLR20] already provide reviews on security, privacy and confidentiality issues in Cloud Computing, which will help us providing a more accurate classification of works related to the topic for this Master Thesis.

CHAPTER 2

CLOUD COMPUTING: BACKGROUND INFORMATION AND DEFINITIONS

Chapter 2: Cloud Computing: Background information and definitions

In this chapter, in order to completely understand our systematic review about privacy and confidentiality in Cloud Computing, we will introduce background information to position the reader into the context. We will not provide a deep background on Cloud Computing as it is out of the scope of this Master Thesis, but the general concepts that will enable readers to firmly grasp the idea of what is cloud computing and its benefits, and understand the rest of this document. If you are interested in getting more knowledge on Cloud Computing, we firmly recommend reading [SLR16].

Also, this chapter will mention several examples of Cloud solutions. Appendix B of this Master Thesis makes a short overview on some of the most popular Cloud vendors with the goal of providing a better understanding and insight on the topic.

This chapter is structured as follows:

1) Section 2.1 defines what is Cloud Computing and its main characteristics.

2) Section 2.2 enumerates the actors that come into play in the Cloud Computing paradigm and describe its functions and responsibilities.

3) Section 2.3 describes the three most popular cloud delivery models by which Cloud providers offer their services.

4) Section 2.4 describes the most popular cloud deployment models, the different channels Cloud providers have to offer their services.

5) Section 2.5 describes which are the main components that make Cloud Computing possible.

2.1. Definitions of Cloud Computing

Various definitions and interpretations of “clouds” and / or “cloud computing” exist. Depending on the usage scope, we will try to give a representative set of definitions.

The National Institute of Standards and Technology (NIST) and Mather et al. [SLR16] provide a working and official definition of cloud computing [35]:

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.

Other works like [1] and [38] define Cloud Computing as platform or infrastructure in which dynamically scalable (elastic) resources are provided as a service through Internet, enabling users to process the data outside the boundaries of the company, providing economical benefits through virtualized and shared infrastructure without the need of expertise nor knowledge over the underlying technology.

In either definition, both describe a paradigm in which users can demand services through Internet (servers, applications, infrastructure, development platforms) whenever they need it, like a commodity. Take the example of the recently published app Cloud Photoshop, a popular image designer and editor. Customers will use Photoshop and pay for what they use and need, no more, no less. This saves the necessity of buying expensive 1000€ licenses, which represents cost-savings.

Just to provide an easy to understand comparison, Cloud Computing is the water or gas of computing. At home you usually don't have a water pump, nor a gas generator, but your home is connected with a set of pipes where the water and gas arrive, previously demanding or contracting a service to a provider company, and usually you pay for what you consume. The more water and gas you consume, the more you pay. The idea of Cloud Computing is pretty much the same, replacing water and gas for services or infrastructure and replacing the pipes for an Internet connection.

The figure 2-1 is a summary of Cloud different capabilities and features. Throughout this chapter we will overview them.

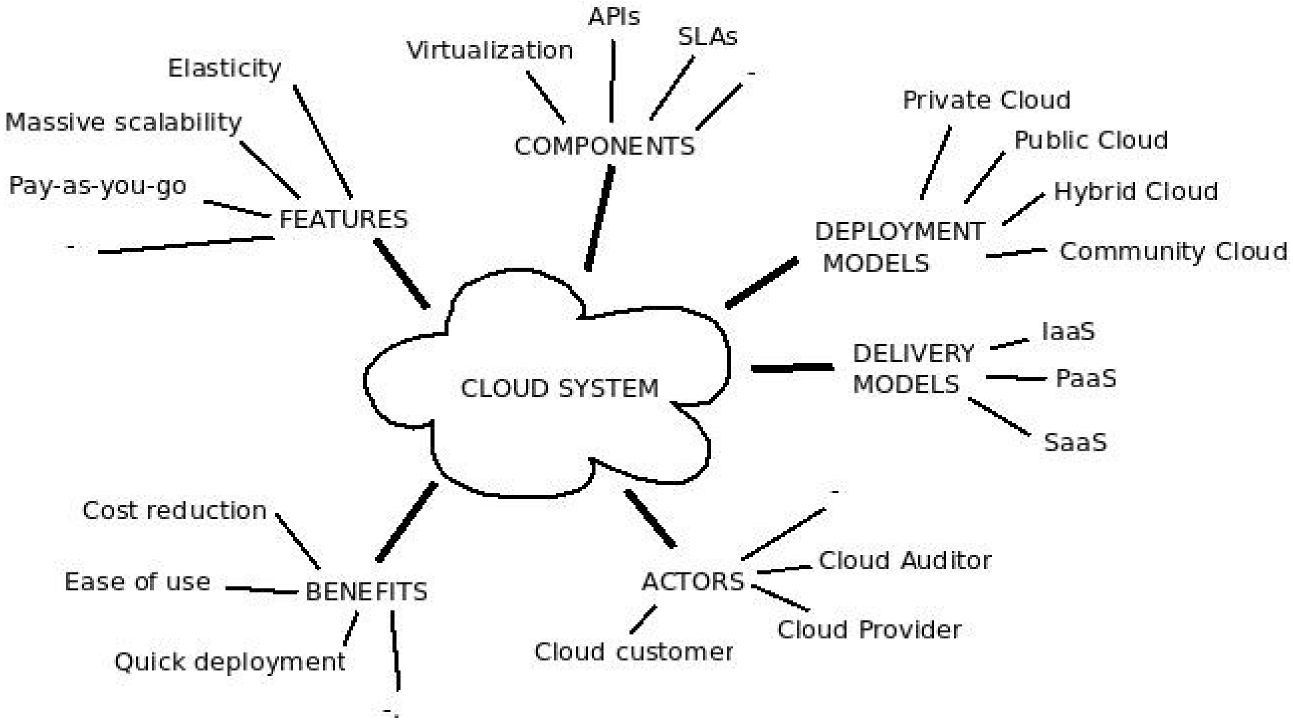

Figure 2-1: Features and capabilities of Cloud Systems

In their book, Mather et al. [SLR16] and NIST [35] define Cloud Computing with a set of essential characteristics that describe this paradigm, which fit into the definitions explained above:

- **Multitenancy** (shared resources): One of the benefits of Cloud Computing is that it is based on a business model where resources are shared among multiple users at the same time. This is usually reached through virtualization.

- **Massive scalability**: Cloud Computing provides the ability to scale to thousands of systems, and massively scale bandwidth and storage space.

- **Elasticity**: Users can rapidly increase and decrease their computing resources as needed, providing IT resources on demand and address spikes in usage, and release resources when they are not required any more. Elasticity can be achieved by using Load Balancers, which is a mechanism to self-regulating properly the workloads among servers, hard drives, network, and other IT resources.

- **Pay-as-you-go**: Users pay for the resources they use and only for the time they required them.

- **Self-provisioning of resources**: Users self-provision resource like processors, software, storage, network resources... without much intervention from the Cloud Provider.

- **Location-Independent Resource Pooling**: the resources may be located at multiple places, being this physical separation transparent to the consumer.

- **Ubiquitous Network Access**: Customers can access their demanded services wherever they need them, being either a web browser, several offices on a company, etc...

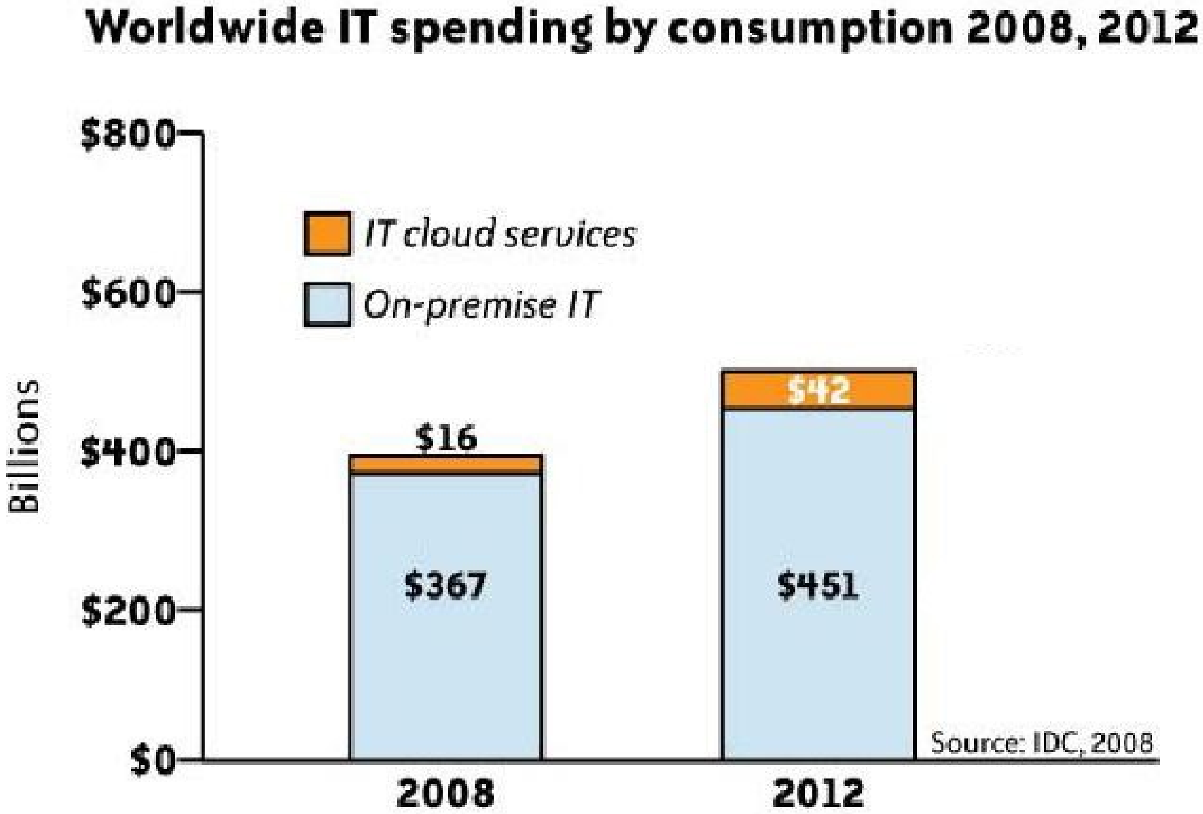

Figure 2-2: Spendings increase in cloud computing services compared with on-site infrastructure

Thanks to Cloud Computing, you can access and make use of powerful computing infrastructure at a very reduced cost compared to buying the whole infrastructure yourself, and acquire it on demand, avoiding the costs of infrastructure obsolescence (when you buy a computer, that piece of hardware has less value for every day it passes). As the time passes, the interest of Cloud Computing raises among companies and it is foresighted that cloud technologies will increase its presence. Figure 2-2 illustrates the increase the spendings in Cloud technologies compared with the on-site IT solutions.

2.2. Actors

According to the NIST Cloud Computing Reference Architecture [10], five major actors have been identified who carry out unique and specific cloud computing activities. This section will summarize the responsibilities of these actors. Figure 2-3 illustrates the actors and their responsibilities.

The main two actors are:

- **Cloud Service Costumer:** They purchase and use services from a Cloud Service Provider or a Cloud Broker. The Cloud Service Costumer could be a company that is interested in moving to a cloud-based solution, e.g. an SaaS solution, such as e-mail.

- **Cloud Service Provider (CSP):** provides the cloud services (service layer), manages the resource allocation and control (Resource abstraction and control layer) and the physical infrastructure (physical resource layer) that conforms the cloud, and they are responsible for the Cloud Service Management (resource provisioning, monitoring, business related services and migration of services between clouds).

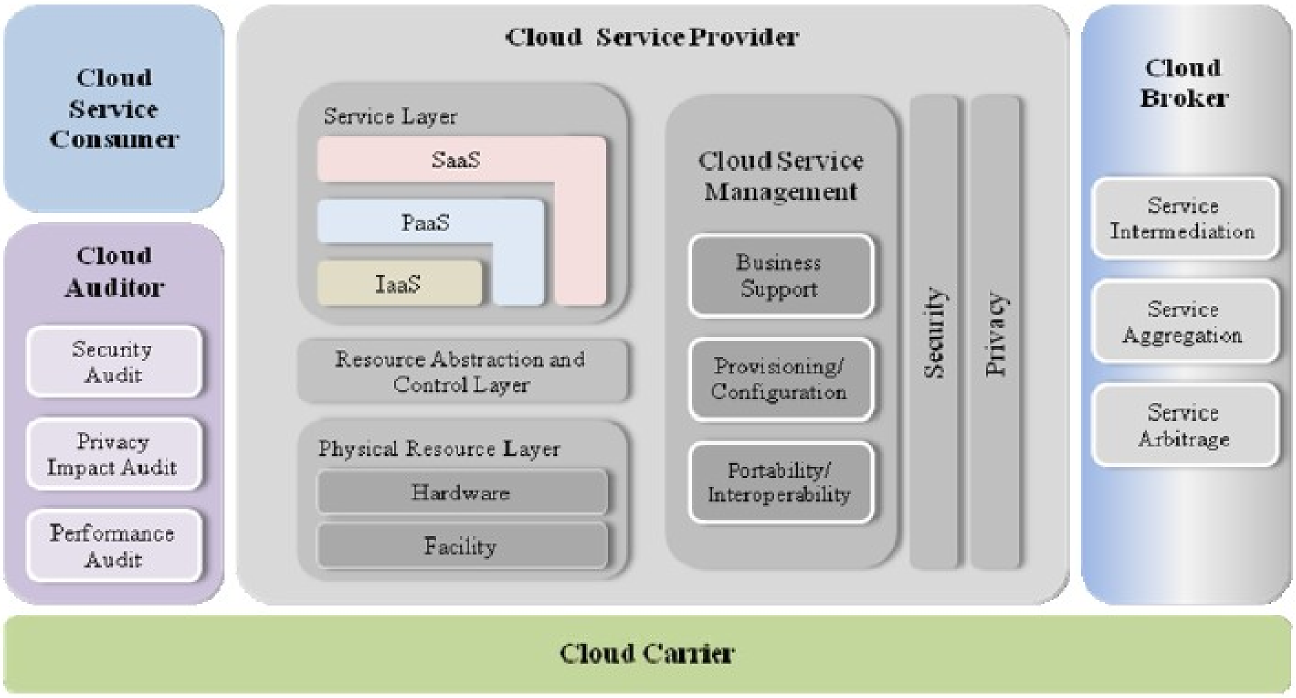

Figure 2-3: NIST Cloud Computing Actors Model

Other actors that take part into the Cloud model are:

- **Cloud Broker:** responsible of integrating and combining cloud services to create enhanced services and providing value-added services to consumers. It behaves as a provider when interacting with a consumer or as a consumer when interacting with a cloud provider.

- **Cloud Auditor:** evaluates and audits the services provided by a cloud provider in terms of security, privacy, performance, etc., to ensure that the vendor operates as expected and that security requirements are met. According to NIST, auditing is especially important for federal agencies, and they should include a contractual clause allowing third parties to assess security controls of cloud providers.

- **Cloud Carrier:** provides connectivity and transport of cloud services from Cloud Providers to Cloud Consumers e.g. an Internet Service Provider (ISP). Cloud carriers provide access to consumers through network, telecommunication and other access devices. A cloud provider usually sets up Service Level Agreements (SLAs) with a cloud carrier to provide services consistent with the level of SLAs offered to cloud consumers. Consumers may require the cloud carrier to provide dedicated and encrypted connections.

2.3. Delivery models

CSPs offer their services according to three fundamental delivery models, depending on the abstraction level provided and the service model of providers [43]: *Infrastructure as a Service*, *Platform as a Service*, and *Software as a Service*. Figure 2-4 illustrates the layered organization of the cloud stack from physical infrastructure to applications, alongside definitions (which are explained in more detail below) and examples of existing market solutions. This model represents the level of abstraction, and can be viewed as a layered architecture where services of a higher layer can be composed from services of the underlying layer.

The descriptions of these delivery models are:

Infrastructure as a Service (IaaS)

In the IaaS delivery model, the CSP provides superior IT infrastructure (storage, processing power, memory...) to run applications and Operating Systems (OS) commonly through virtualization technologies like Vmware [77].

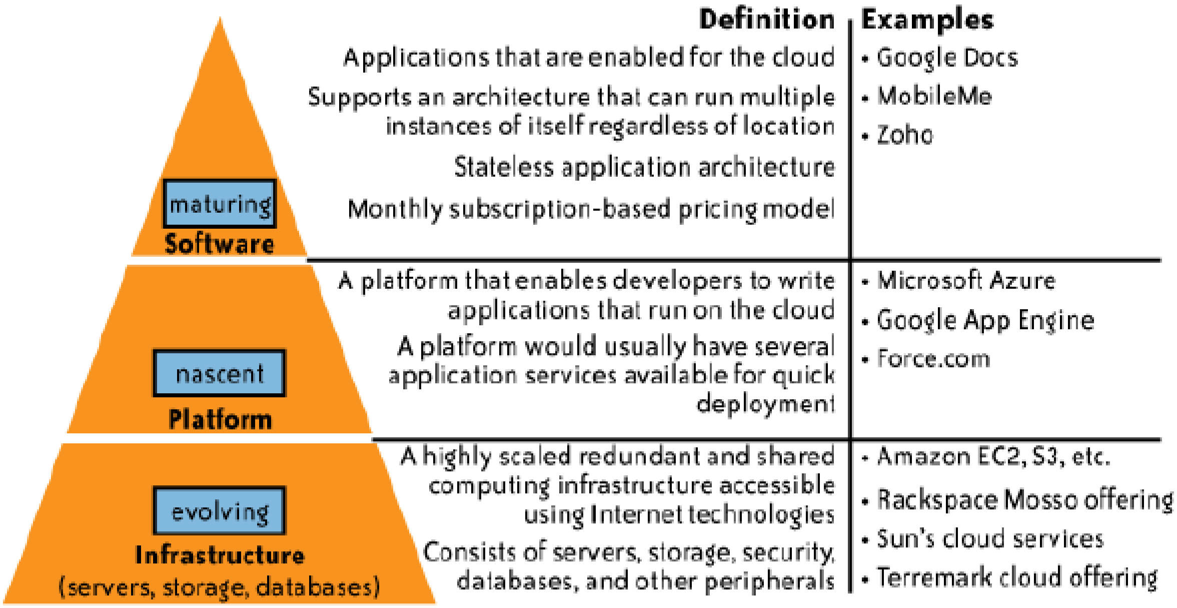
Definition
Examples
Applications that are enabled for the cloud
Supports an architecture that can run multiple instances of itself regardless of location
Stateless application architecture
Monthly subscription-based pricing model
• Google Docs
• MobileMe
• Zoho
maturing
Software
A platform that enables developers to write applications that run on the cloud
A platform would usually have several application services available for quick deployment
• Microsoft Azure
• Google App Engine
• Force.com
nascent
Platform
A highly scaled redundant and shared computing infrastructure accessible using Internet technologies
Consists of servers, storage, security, databases, and other peripherals
• Amazon EC2, S3, etc.
• Rackspace Mosso offering
• Sun's cloud services
• Terremark cloud offering
evolving
Infrastructure
(servers, storage, databases)

Figure 2-4: The cloud computing delivery models pyramid

Some examples of IaaS are: Amazon S3 [64], SQL Azure [65], Terremark [66], Rackspace [67].

Do not confuse hiring a server for a specific amount of time for specific and fixed resources with IaaS. IaaS has several advantages over traditional hosting:

- It scales and adapts the infrastructure like memory, storage, processor and other computing resources on-demand to fulfil capability requirements almost at a real-time speed (elasticity).
- You purchase and pay for the amount of infrastructure required at any time.

Platform as a Service (PaaS)

In the PaaS delivery model, cloud vendors offer a browser-based development studio solution to application developers, upon which applications and services can be developed, hosted and offered to customers through the provider's platform. PaaS typically makes use of dedicated APIs (Application Programming Interfaces), development toolkits and standards for usually building web-based applications. These tools are aimed at providing monitoring of application and user activities, integration with external web services and databases, scalability, reliability and security, billing mechanisms and multi-tenancy without too much development.

Some examples of PaaS are Microsoft Azure [66], Google App Engine [68] or Force.com [69]

PaaS delivery model has several advantages:

- Allows the rapid propagation of software applications due to the low cost.
- Allows single developers and start-up companies to deploy web-based applications without the cost and complexity of buying and setting up servers.
- Allows software vendors to control, limit the use, prohibit copying and distribution and facilitates the control of versions of their software.

Software as a Service (SaaS)

Traditionally, customers purchase software licenses and install them in their own hardware. In the SaaS delivery model, users rent the software under a subscription or a pay-per-use model (sometimes free for a limited time or under several conditions like allowing advertisement) and access the software through Internet (for example with a web browser) under some authorization mechanism [SLR16]. More specifically, Cloud Providers offer implementations of specific business functions and business processes in the form of applications or services under an established cloud infrastructure or platform.

Some examples of SaaS are Google Docs [70], MobileMe, redirecting to iCloud [71], Zoho [72] and Salesforce CRM [73].

SaaS delivery model has several advantages:

- Allows organizations to outsource the hosting and management of applications to the CSP, reducing licensing, personnel and infrastructure costs.
- Allows software vendors to control, limit the use, prohibit copying and distribution and facilitates the control of versions of their software.
- Management of a SaaS application is supported by the vendor. However they cannot be completely customized.
- Unlike single-tenant architecture applications, SaaS physical back-end infrastructure is shared among different customers (but logically unique for the customers), maximizing the sharing of resources among them.

These are the most common types of Clouds provided by most providers, yet there exists other types of clouds which derive from the former three [4, 39], but we will not describe them in detail in this Master Thesis.

Just to illustrate a bit more the differences among the three delivery models, Figure 2-4 illustrates the level of control of several parts of the infrastructure. When the infrastructure is locate right in organization's offices, organization has full control of everything (except network, which is obviously managed by the ISP). When hiring hosted infrastructure, user controls the operating system which can host Virtual Machines and applications, but the control of the physical server relies more on the Hosting provider. When moving to a IaaS, customers can control the applications that run inside a Virtual Machine, but not the Virtual Machine manager itself nor the physical machine. On a PaaS, customers control applications and services under cloud vendors platform (which limits the control). Finally, on SaaS users don't have control of applications and they can just use it.

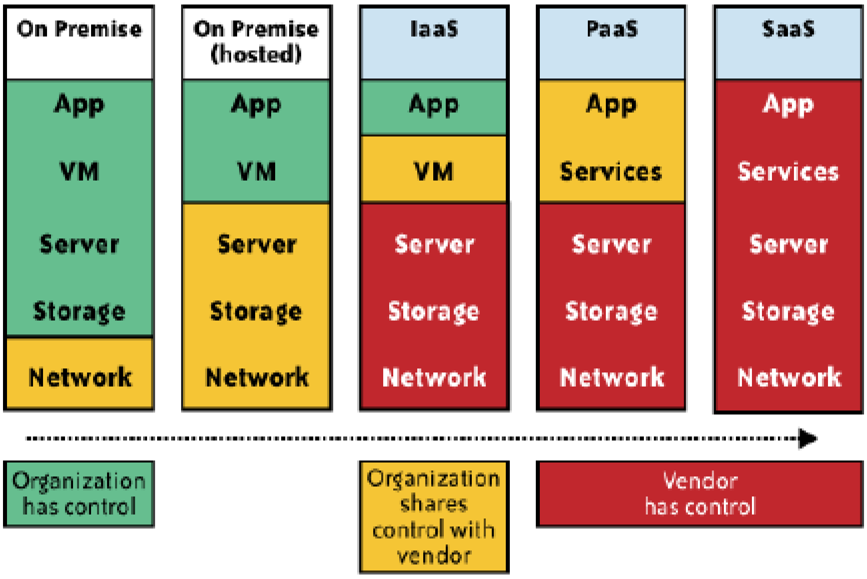

Figure 2-5: Areas of control between customer and vendor depending on the delivery model

2.4. Deployment Models

This section will describe the ways previously explained delivery models can be deployed in the cloud. There have been identified four deployments models in the literature:

Private Clouds

Also known as internal clouds, private clouds are designed for exclusive use (storage, computing...) by a single organization on a private network. A private cloud offers the highest degree of control over performance, reliability and security. However, they are purchased and completely managed by the organisation, so private clouds don't benefit from lower costs due to shared environments unlike other models and requires internal IT expertise or delegate the management to third parties.

Example: eBay [74].

Public Clouds

Public clouds provide on-demand services to the general public over a common infrastructure hosted, operated and managed by a third-party vendor. Security management and day-to-day operations are relegated to the vendor. Public clouds offer several key benefits to customers, including no initial capital investment on infrastructure, lower costs and shifting of risks to providers' infrastructure.

However, customers have a low degree of control in these kind of control compared with private clouds, which raises a huge amounts of security and privacy concerns that are the basis of this document, as public clouds are the main traditional way of deploying Cloud Computing architectures.

Examples: Amazon EC2 [75], Windows Azure [65].

Hybrid Clouds

A hybrid cloud is a combination of public and private cloud models that tries to address the limitations of each approach. In a hybrid cloud, part of the service infrastructure runs in private clouds (e.g. core applications, sensitive data) while the remaining part runs in public clouds (e.g. non-core applications). Hybrid clouds provide more control and security over data compared to public clouds while still facilitating on-demand service and elasticity. However, the design of hybrid clouds require to carefully determine what should be split into public and private cloud components.

Examples: Juniper [76].

Community clouds

Community clouds are a less common type of cloud compared to the other deployment models (and somewhat experimental). In these kind of clouds, an organization shares infrastructure among several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), and can be either managed internally or by a third-party, and hosted internally or externally. The costs are spread over fewer users than a public cloud (but more than a private cloud), so there are cost-savings, but in a lower degree than public clouds, but more control compared to public clouds.

Examples: Zimory [68], RightScale [69]

2.5. Cloud components

[1][SLR16] Cloud Computing does not refer to a specific technology, but a combination of pre-existing technologies and protocols that made this paradigm possible. In this section we will describe the most relevant technological and non-technological components which support Cloud Computing architectures.

Just as a small remark, non-technological aspects described in section 2.5.2 were defined before the Cloud Era, but they are applicable as well under this paradigm (with several remarks and appointments which will be reviewed through Chapter 4).

2.5.1. Technological components

Data centres and server farms

Cloud services require large amounts of computing capacity in order to provide high flexibility and power, and are hosted in data centres and server farms (which can be distributed in multiple locations).

For example, Google links a large number of servers to provide their services, or Amazon EC2 provides virtualization in a data centre to create a huge amount of virtual instances.

Virtualization technologies

The most important technology that makes cloud computing possible. Virtualization technologies abstracts physical computing resources (CPU, storage, network, memory, databases...) from applications and end users, providing the capability of creating pooled resources accessible to anything authorized to use them.

This capability provides the multi-tenancy properties of cloud-computing, providing scalable and shared resource platform capabilities for all tenants.

Virtualizations technologies are supported by the hypervisor, a layer of software between an operating system and hardware platform that is used to operate Virtual Machines like launch or terminating, and implements and manages virtual CPU (vCPU), virtual memory (vMemory), event channels and memory shared by the hosted Virtual Machines (VMs), controls data Input/Output and memory access to devices.

When talking about virtualization applications, common users think on applications to create guest OS Virtual Machines like VMWare or VirtualBox, creating instances of Operating Systems that share resources with the host physical machine and other Virtual Machines. However virtualization does not limit to operating systems, but other components that can be virtualized, like storage (Storage Area Networks), databases, and software (Apache Tomcat, WebSphere, Jboss...).

Depending on the delivery model, virtualization appears at various layers. In IaaS commonly storage and OS is virtualized, while on SaaS and PaaS databases and software are the typically virtualized component.

Application Programming Interfaces (APIs)

APIs provide features to customers, like self-provisioning and control of services and resources, and allow communication between foreign applications and the cloud service. The type of APIs depend on the delivery model, and can go from simple URL manipulations to SOA (Service Oriented Architecture) programming models.

However each CSP usually has its unique API and no standards exist on this matter , which makes cloud applications not portable among clouds, harming interoperability and locking-in the customer into the vendor's cloud. Also, IT staff is required to become familiar with those platform-specific features (although some organisations like the Universal Cloud Interface, which attempts to unite APIs from different cloud providers to create and open and standardized cloud interface).

Encryption

Encryption is the process of encoding messages or information in a way that that only authorized parties can read that information, preventing unwanted parties and hackers read it. This is usually done with the use of an encryption key, which specifies how the message will be encoded. Then, the authorized party uses a secret decryption key to decipher the message and read it. Encryption is the common technique used to protect the confidentiality of messages.

2.5.2. Non-technological components

Service-level Agreements (SLA)

The mutual contract between providers and users is usually called SLAs which offers guarantees on the quality of the service by defining what is to be expected from the offered service, and defines compensatory schemas in case the provider violates these contract pre-defined agreements. Some examples of these agreements could be 99% of availability, resolution of incidents under a specified period of time, data security, etc.

Privacy Policies

Wikipedia [45] provides an excellent definition of the term policy, which will be very recurrent throughout this document. A policy is a statement of intentions, implemented as protocols or procedures, to guide decisions and achieve a desired goal. When applied under the privacy topic [46], these policies represent legal documents that explains and discloses the way a party gathers, uses, discloses and manages customer's data (e.g. name, address, medical history, financial records, commercial transactions, etc.). It also informs the client about what information is collected and whether it is kept confidential, shared with partners or sold to other firms of enterprises. These aspects are usually collected under a written document which outlines rules, provides principles that guide actions, sets roles and responsibilities, reflect values and beliefs and state a protocol of actions [47].

In this sense, privacy policies forge the individual's physical and moral autonomy, so a set of laws in different countries have been developed over the years to protect and enforce these policies (refer to section 4.3.2.2.2), although there exists differences in the way these privacy laws are implemented and enforced in different countries. While the European Union enforces that data protection laws must be met by any business operating or transferring personal information about any EU citizen or business, in United States privacy laws only apply to the public sector, not the private sector.

Terms and conditions

Also known as Terms of Service (ToS), they are rules which an entity must agree and accept in order to use a service. The terms-of-service agreement is mainly used for legal purposes by websites and Internet service providers that store a user's personal data (e.g. e-commerce sites and social networking services). A legitimate terms-of-service agreement is legally binding, and may be subject to change.

A terms-of-service agreement typically contains sections explaining the user rights and responsibilities of expected usage and potential misuse of the service; accountability for actions, behavior, and conduct; a privacy policy outlining the use of personal data; possible payment details such as membership or subscription fees, etc.; possibly a policy describing procedure for account termination; disclaimer/limitation of Liability clarifying the service's legal liability for damages incurred by users; and user notification when the terms are modified.

CHAPTER 3

THE REVIEW METHOD: SYSTEMATIC LITERATURE REVIEW

Chapter 3: The review method: Systematic Literature Review

This chapter describes the methodology, and the search and work selection strategies we used to elaborate the state-of-the-art over privacy and confidentiality issues in Cloud Computing architectures presented in the Chapter 4 of this document. We used the **Systematic Literature Review** methodology (SLR). The exposed results in this chapter were gathered in 11/10/2012.

The reason to perform a systematic literature review over privacy and confidentiality on cloud computing is to provide an state-of-the-art about this area, identify which are the privacy and confidentiality requirements that practitioners take into account to accordingly plan their architectures and make their decisions and summarize the existing evidence about the topic. SLR methodology is described on B. Kitchenham's guidelines [21], which proposes systematic review guidelines specific for software engineer researchers.

Kitchenham defines Systematic Literature Review as follows:

A systematic literature review (often referred to as a systematic review) is a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest. Individual studies contributing to a systematic review are called primary studies; a systematic review is a form of secondary study.

There are several features that differentiate a systematic review from a conventional literature review:

- *Definition of the review protocol:* Review protocol specifies the research question being addressed and the methods that will be used to perform the review.
- *Definition of the search strategy:* Search strategy aims to detect as much of the relevant literature as possible.
- *Documented searches*, so that readers can assess its rigour and completeness.
- *Explicit inclusion and exclusion criteria* to assess each potential primary study (a systematic review is a secondary study based on primary studies).
- Systematic reviews specify the information to be obtained from each primary study including quality criteria by which to evaluate each primary study.

- A systematic review is a prerequisite for quantitative meta-analysis.

Sticking with fidelity to the guidelines allows to perform reviews which can be replicated by other people which allows for an easy expansion of recent work (through performing the SLR at another point of time) and assess the veracity of results.

The systematic review takes three main phases, each one of them associated with several activities:

1) *Planning the review:* identify the need for the review, specify research question and, in general, develop a review protocol. This chapter is focused on describing this phase.
2) *Conducting the review:* this phase consists on the selection of primary studies, quality extraction and data synthesis. Chapter 3 explains the selection of primary studies and quality extraction, chapter 4 on data synthesis and chapter 5 on the analysis of review results.
3) *Reporting the review:* format the main report. Chapters 3, 4 and 5 are formatted following advisors' guidelines based on Kitchenham's guidelines.

3.1. Review protocol

As stated in [21], “A review protocol specifies the methods that will be used to undertake a specific systematic review. A pre-defined protocol is necessary to reduce the possibility researcher bias.”

So, the following points are indicated in the guidelines for the development of the protocol:

3.1.1. The review question

It is obvious that before finding a solution, we need to define which is the question we want to answer. Initially, we wanted to do a SLR on overall security aspects on Cloud Computing, so the questions were:

- How does security affect the architecture of Cloud Computing?
- Which are the currently identified issues regarding security aspects which should be addressed?
- Which are some of the solutions proposed to solve these issues?

However, security matters are broad, encapsulating a lot of fields. For practical matters, we decided later to switch the questions and focus on the security sub-area Privacy and

Confidentiality. The main reasons for this changes were the amount of references found on the first generic searches, as explained on section 3.1.2.3. The final questions for the SLR of this master thesis are:

- **What is the impact of privacy and confidentiality requirements in Cloud Computing architectures?**
- **Which are the currently identified issues and challenges regarding privacy and confidentiality in Cloud Computing platforms. What are some of the solutions proposed to solve these issues?**

The answer to these questions will provide a good background about the current status of privacy and confidentiality requirements on the Cloud Computing paradigm and identify which is the current direction Cloud Computing is heading towards providing trustiness on organizations' information and data. This answer will be interesting for cloud adopters, researchers and engineers as explained in section 3.1.2.

In order to be able to answer the review question, first we had acquired in Chapter 2 an idea about what is Cloud Computing, what it is its architecture and the necessary requirements to define its architecture. Then, with all this knowledge and further literature, we will identify possible uncovered research gaps on privacy and confidentiality.

This will allow to summarize the existing evidence and knowledge concerning law issues, technological aspects and provide a background which could lead and position new research activities and identify possible unknown gaps.

3.1.2. The search and data extraction strategy

In order to find proper information to answer our question, we need to define a search strategy. The strategy outline we followed was:

1) Selection of **proper information sources** (see section 3.1.2.1).
2) **Find keywords** in order to formulate a search expression which will retrieve the papers we need about the topic as most specific as possible (see section 3.1.2.2).
3) **Elaborate a search expression** to retrieve relevant literature about the topic (see section 3.1.2.3)
4) **Select articles based on their title**: in order to reduce the number of articles (see section

3.1.2.4, EC1).

5) **Read the abstract, the conclusions and check keywords**: sometimes the title is too generic or leads to confusion (see section 3.1.2.4, EC2).

6) **Read the article**: the selected publications from step 5 were read in depth in order to assess if they provide useful value for our SLR or not (section 3.1.2.4, EC3).

7) **Retrieve new publications from references (snowballing)**: as I'm reading publications, we may find useful and interesting to read some referred articles to get deeper insight of a topic (see section 3.1.2.5 and 3.1.2.6).

8) **Classify the papers**: once a paper has been read, classify it in order to then be able to retrieve them quickly as needed in 2 categories: Reviews, and Proposals and Solutions, and then a sub-classification for each category: Least interesting, interesting and very interesting. Some papers may be included in more than one category (see section 3.1.2.7).

Next subsections describe with more detail the strategy process.

3.1.2.1. Information sources

For this Systematic Literature Review, initially we considered using 2 paper database sources in order to find literature which could help to answer our review question:

Inspec Database

Inspec is a major indexing database of scientific and technical literature, published by the Institution of Engineering and Technology (IET). Inspec coverage is extensive in the fields of physics and computer, control, and mechanical engineering. Its subject coverage includes astronomy, electronics, communications, ergonomics, computers & computing, computer science, control engineering, electrical engineering, information technology, and physics.

Inspec covers the following databases:

- ACM Digital Library
- Compendex*Plus
- IEEE Xplore
- MathSciNet
- Web of Science

ISI Web of Knowledge

ISI Web of Knowledge is an academic citation indexing and search service, which is combined with web linking. Web of Knowledge covers areas like sciences, social sciences, arts and humanities, and provides bibliographic content and tools to access, analyse, and manage research information. Moreover, multiple databases can be searched simultaneously, being Web of Science and Inspec among them.

However, both databases (specially the ISI Web of Knowledge) use non-specialized databases which provided too many irrelevant hits with our search expression (refer to section 3.1.2.3), so we finally decided to use the following databases which are specialized on the domain of computing: **IEEE Xplore**, **ACM Digital Library** and **Springerlink**.

IEEE Xplore

IEEE Xplore is a research database that indexes, abstracts, and provides full-text for articles and papers on computer science, electrical engineering and electronics. The database mainly covers material from the Institute of Electrical and Electronics Engineers (IEEE) and the Institution of Engineering and Technology. The *IEEE Xplore* database contains over three million records.

ACM Digital Library

The ACM Digital Library contains full-text collection of all articles published by the Association for Computing Machinery (ACM), and includes almost 200.000 records of full-text access to conference proceedings, magazines, newsletters, and journals, and covers topics from computer technology, online education, software engineering, programming, networking and information to name some areas.

Springerlink

Springerlink develops, manages and disseminates knowledge, publishes books, e-books and peer-reviewed journals in science, technical and medical publishing. They work with the world's best academics and authors in long-standing loyal partnerships based on mutual trust and they are always open to new input. Springer also hosts a number of scientific databases, including

SpringerLink, Springer Protocols, and SpringerImages. Book publications include major reference works, textbooks, monographs and book series; more than 50,000 titles are available as e-books in 13 subject collections.

3.1.2.2. Keyword finding

In order to find good keywords for our searches, and have some clue about where to start and being able in the future to classify correctly search results, we read some papers, publications and books about Cloud Computing security, gathered from Internet and Experts (this Master Thesis's advisors). Table 3-1 describes which are the publications that were used for this purpose as starting point entries.

Document type	References
Books	[37]
Papers	[1], [43], [36]

Table 3-1: Publications used for background setting and keyword finding

The results of this first search were:

- We gathered knowledge about security aspects and requirements in cloud computing which would then come up with keywords to refine my searches. The paper that provided the most benefits for our keyword finding was [36] as it provides more details regarding security than the other publications. This paper identifies 9 sub-areas related to security in cloud computing:
  - Access control
  - Attack Detection
  - Non-repudiation
  - Integrity
  - Security Auditing
  - Physical Protection
  - Privacy and Confidentiality
  - Recovery
  - Prosecution

  Finally, the keywords found useful for finding results for our systematic review are:

Requirement, security, cloud computing, access control, attack, detection, repudiation, integrity, audit, physical, protection, privacy, confidentiality, recovery, prosecution, law

We decided to split the keywords “Attack Detection” and “Physical Protection” in order to increase the amount of papers retrieved for the search expression.

- ○ In the case of Attack Detection, some papers refer to attack detection as “attack protection”, “intrusion detection”, “DDoS attack” and similar. Dividing the two keywords allows for a wider range of results.
- ○ In the case of “Physical Protection”, some papers refer to physical protection as “physical attack protection”, “secure physical data”, or just “hardware protection” to enumerate some of them. Dividing the two keywords allows for a wider range of results.

Also, as explained in section 3.1.2.3, we decided to move to Privacy and Confidentiality. Given that some papers include the word “*law*” as a common word, we decided to include it as a keyword to consider.

- ■ We improved my own background about the topic so we can define better exclusion criteria and provide background information for our Systematic Literature Review.

3.1.2.3. The search expression

As a means to retrieve the papers, we formulated a first query to the paper databases based on the title of this master thesis, my knowledge gathered from reading the documents in Table 1 and the keywords found in section 3.1.2.2. In this subsection we are describing the search expressions used for this literature review as well as the amount of results for each expression.

For this Master Thesis, firstly we considered to perform a general review about security, so the first search results we wanted were papers which could provide me a general background about security requirements. So, the first expression we used was ***(Cloud computing) AND (security OR privacy).*** Table 3-2 shows the results of this query.

Search expression	ACM DL	Inspec	IEEExplore
(Cloud computing) AND (security OR privacy)	1184 hits	2398 hits	419 hits

Table 3-2: Results for the first used search expression

As this search provided too many results, a too broad scope and a good amount of them weren't really focused on cloud computing, we decided, first, to switch and limit the databases used as explained in section 3.1.2.1, and then we modified the expression to ***(Cloud computing) AND (confidentiality OR privacy).*** Table 3-3 shows the obtained results for this query.

Search expression	IEEExplore	ACM DL	Springerlink
(Cloud computing) AND (confidentiality OR privacy)	48 hits	825 hits	1383 hits

Table 3-3: Results for the second used search expression

These results still proved to be not so specific to cloud computing or privacy (specially in the case of ACM), so we decided to focus on one sub-area of security in cloud computing. We performed a search for every sub-area of security as explained in section 3.1.2.2, use the resulting keywords and choose one of them to perform the systematic review based on the amount of results. Table 3-4 outlines the queries we performed and their corresponding results.

Search expression	IEEExplore	ACM DL	Springerlink
"cloud computing" AND security AND "Access control"	146 hits	325 hits	647 hits
"cloud computing" AND security AND Attack AND detection	50 hits	291 hits	316 hits
"cloud computing" AND security AND repudiation	8 hits	19 hits	50 hits
"cloud computing" AND security AND integrity	104 hits	409 hits	674 hits
"cloud computing" AND security AND audit*	104 hits	238 hits	340 hits
"cloud computing" AND security AND physical AND protect*	14 hits	253 hits	391 hits
"cloud computing" AND security AND (privacy OR confidentiality)	66 hits	719 hits	1236 hits
"cloud computing" AND security AND recovery	21 hits	285 hits	494 hits
"cloud computing" AND security AND prosecution	1 hits	7 hits	18 hits

Table 3-4: Results for several search expressions

After analysing the results and based on my personal preference, we choose to focus on privacy and confidentiality and continue with the expression *"cloud computing" AND security AND (privacy OR confidentiality).* However, we decided to perform a last modification on the search expression to get more specific results and ultimately use **"cloud computing" AND security AND privacy AND confidentiality (without duplicates).** Table 3-5 shows the results of this expression for our sources.

Search expression	IEEExplore	ACM DL	Springerlink
"cloud computing" AND security AND privacy AND confidentiality	51 hits	192 hits	134 hits
"cloud computing" AND security AND privacy AND confidentiality (without duplicates)	44 hits	192 hits	120 hits

Table 3-5: Results for the chosen search expression

In total, the searches returned 377 papers, so we decided to hold on this search expression as the final expression used for the systematic review of this Master Thesis. After removing duplicates, for a total of **356 papers**. Figure 3-1 represents graphically the papers based on their source.

ACM Digital Library provided the highest amount of papers. This does not mean, though, that we will find there the highest amount of useful papers for our review as we will see on the next sections.

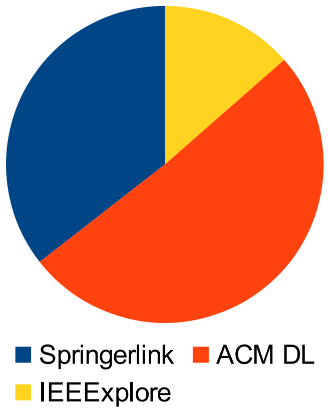

Figure 3-1: Relation of papers gathered from sources

3.1.2.4. Exclusion criteria

In order to select the proper papers for our systematic review, we defined the following exclusion criteria:

- **Selection by title (EC1):** Discarded those whose title points to a conference's workshop, or to a very specific and limited domain of application (like Healthcare, Banking...), or it doesn't make any reference to the cloud computing paradigm, or it doesn't refer to either privacy or confidentiality or specifically to security.
- **Selection by abstract and conclusions (EC2):** Selected all those papers whose abstract and conclusions fulfil the following requirements:

  - ○ Makes a reference to cloud computing.
  - ○ Mentions either privacy or confidentiality as the focus of the paper, or at least providing a state-of-the-art about general security aspects.
  - ○ It's not too technical focused nor too much specific to a certain domain (healthcare, banking...).
  - ○ Mentions some relevant research direction on security, privacy or confidentiality.
  - ○ The conclusions provide an overall idea about what it should have been read in the work.

- ■ **Selection by full-text (EC3):** papers whose text does not aim to provide an explanation to security nor privacy, or are too technical focused or not specific to cloud computing were discarded.

Table 3-6 describes, which have been the results of each applied exclusion step for each database, and the amount of discarded papers from the previous step in %:

Exclusion steps	IEEExplore	ACM DL	Springerlink	Total	% Discarded
0. Starting point	44 papers	192 papers	120 papers	356 papers	-
1. EC1	17 papers	16 papers	34 papers	67 papers	81.17%
2. EC2	11 papers	7 papers	9 papers	27 papers	59.7%
3. EC3	9 papers	1 papers	5 papers	**15 papers**	55.6%

Table 3-6: Results obtained from the initial process after applying the exclusion criteria

As it can be stated, a huge amount of papers were strangely discarded from ACM DL, and after verifying some random papers from the discarded pile, we concluded that ACM searches also through the document, adding to the search results papers who barely make a mention to cloud computing (maybe due to a full-text search).

The rest of steps in the table are trivial to explain. After applying the three filters, there are **15 papers** left from the initial 356, which represents a 4.2% from the total once duplicates were removed. These 15 papers will be included on our SLR. Figure 3-2 represents the whole process of exclusion.

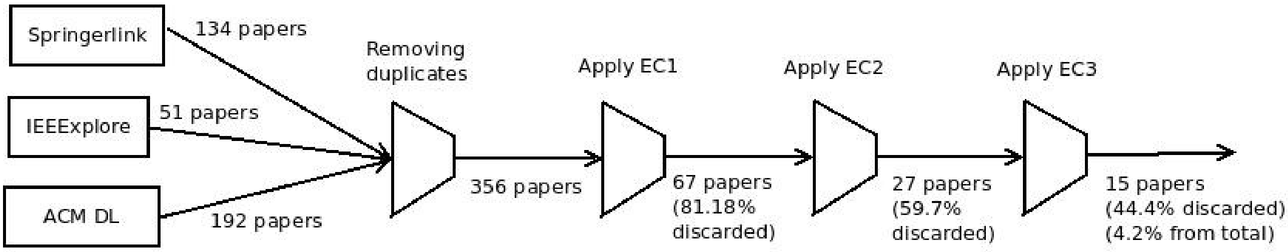

Figure 3-2: The selection process of the initial search

Some conclusions could be extracted from this process. The first conclusion is that, even though IEEExplore provided the smallest amount of papers for our research question, it ended being the one with the highest amount of chosen papers after applying our exclusion criteria. This makes us think on two aspects regarding IEEExplore: either its database is good and contains more specialized content than the other databases, or its search engine is more accurate than the other databases, or both.

Another conclusion that could be extracted is that ACM Digital Library's search engine needs some extra options in order to tweak the search to better fit our search requirements. ACM provided a lot of irrelevant results for our SLR.

We considered that this was a fairly low amount of papers in order to offer a good quality SLR for this topic. One possible reason for this could be an excessively restrictive research question. As we explain in section 4.2 - What is confidentiality? - authors very often use the terms privacy and confidentiality as synonyms when explaining their solutions/reviews about protecting personal information, and although there are some differences between the two words, this line seems pretty blurry in the literature.

Therefore, it's highly probable that if we would have held on the search expression “cloud computing” and “security” and (“privacy” or “confidentiality”) we would have got more good quality papers after applying EC1, EC2 and EC3, as the amount of papers provided by that search was 2,021 as stated in section 3.1.2.3. However, due to time restrictions and practical reasons for this Master Thesis, we will leave this possible continuation or “branching” of this master thesis to future researchers interested in reviewing this topic (see section 6.2 – Future work).

So we had to choose between relaxing the exclusion criteria and include more papers, or apply a snowballing process, selecting the references of the chosen papers. Commonly researchers who write on these sources already did their personal selection of good quality papers, so we are going to take advantage from this and perform the last option, apply a snowballing process to gather and ensure good quality results.

3.1.2.5. Selected publications by references (snowballing)

Some of the selected papers contain references to other papers which can provide added value to the SLR. These papers have been included for the review as well.

We performed the following Exclusion criteria over the set of documents composed by all the references from the selected papers:

- **Selection by title (S-EC1):** Discarded those papers that do not fulfil the original search expression “cloud computing” AND security AND (“privacy” OR “confidentiality”) on the title. As stated in the end of section 3.1.2.4, “cloud computing” AND security AND “privacy” AND “confidentiality” is a too restrictive expression, so we decided to go back to the original search expression and not use the one we are using for the initial search, because none of the papers in the references matched that expression. Afterwards, we also discarded those whose title points to a conference's workshop, or to a very specific and limited domain of application (like Healthcare, Banking...), or don't make any reference to the cloud computing paradigm, or don't refer to either privacy or confidentiality or specifically to security.

However, these conditions were still too restrictive. The search domain on the references is way smaller than the search domain on IEEExplore, ACM DL and Springerlink databases, so using a more unrestrained search expression is more practical and feasible. Hence, we decided to “soften” the restrictions and apply the following exclusion criteria:

- **Selection by title (S-EC1-2):** Discarded those papers that do not fulfil the expression “cloud” AND (security OR privacy OR confidentiality OR law*) on the title. Afterwards, we also discarded those whose title points to a conference's workshop, or to a very specific and limited domain of application (like Healthcare, Banking...), or it doesn't make any reference to the cloud computing paradigm, or it doesn't refer to either privacy or confidentiality or specifically to security.
- **Selection by availability (S-EC2):** Only those documents which are available for free were selected.
- **Selection by source (S-EC3):** Those papers which appear in the databases we use (ACM, IEEExplore, Springerlink) are discarded. We do this because if we select them, we should

have acquired them in our first search.

- **Selection by abstract and conclusions (S-EC4):** Selected all those papers whose abstract and conclusions fulfil the following requirements:
  - Makes a reference to cloud computing.
  - Mentions either privacy or confidentiality as the focus of the paper, or at least providing a state-of-the-art about general security aspects.
  - It's not too technical focused nor too much specific to a certain domain (healthcare, banking...).
  - Mentions some relevant research direction on security, privacy or confidentiality.
  - The conclusions provide an overall idea about what it should have been read in the work.
- **Selection by full-text (S-EC5):** papers whose text does not aim to provide an explanation to security nor privacy, or are too technical focused or not specific to cloud computing were discarded.

Figure 3-3 outlines the steps performed for selecting the papers from the references. As it can be noted, almost every document was available for free. In the end, **6 documents** have been finally selected for our SLR in the snowballing process.

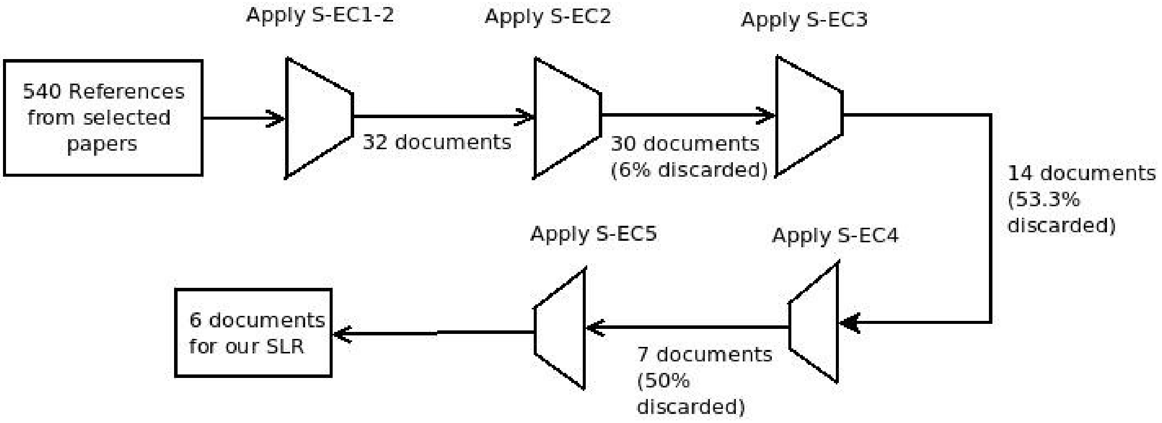
Apply S-EC1-2
Apply S-EC2
Apply S-EC3
540 References from selected papers
32 documents
30 documents (6% discarded)
14 documents (53.3% discarded)
Apply S-EC5
Apply S-EC4
6 documents for our SLR
7 documents (50% discarded)

Figure 3-3: The selection process of the snowballing step

However, it can be seen that almost 50% (16 papers) of the papers were discarded in S-EC3 because they appeared on our initial search engines. That is a pretty high amount of papers, considering that we tweaked a bit the expression search for selecting the papers from references. We are aware that we could have modified the initial search expression **“cloud” AND (security**

OR privacy OR confidentiality OR law*) to keep coherence between the selection method used in the initial search and the selection method used in the snowballing process. But as it can be seen in section 3.1.2.4, **(Cloud computing) AND (confidentiality OR privacy)** provided a high amount of results, and is an expression contained in “cloud” AND (security OR privacy OR confidentiality OR law*) which, logically, would induce to an even greater amount of results.

Therefore, we will include the discarded publications in S-EC3 and Apply S-EC4 and S-EC5 over them (It's obvious that all those papers passed S-EC1 and S-EC2 as they were selected by title in the snowballing process).

16 publications from the references were discarded on the snowballing process at S-EC3. We applied the exclusion criteria in section 3.1.2.4 to them. After applying S-EC4 (selection by abstracts and conclusions), 7 papers were selected. Finally, after applying S-EC5 (selection by full-text), 6 papers have been chosen for our SLR from the pile discarded in S-EC3, which makes up for **a total of 12 publications obtained via the references added to our SLR**. Figure 3-4 complements figure 3-3 with this last addition.

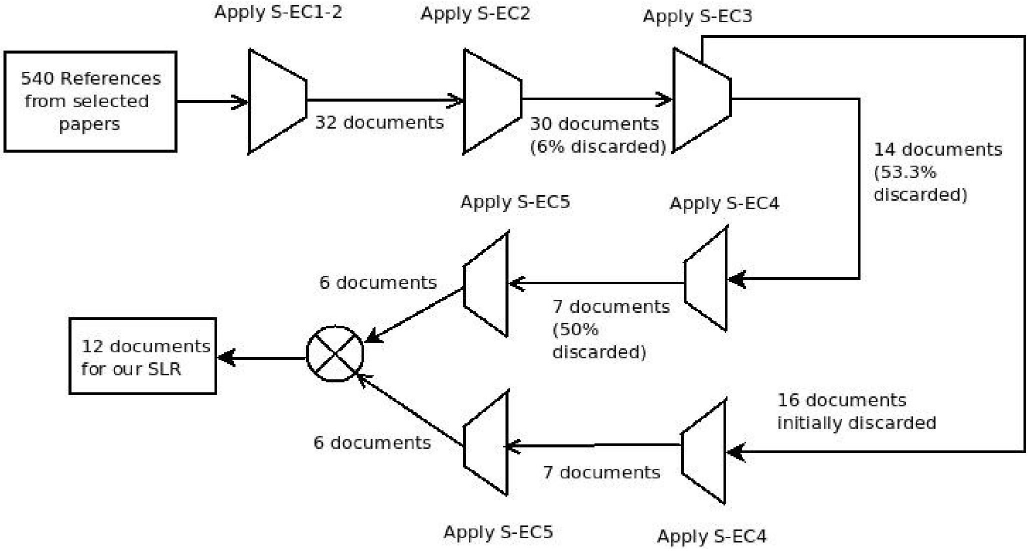
Apply S-EC1-2
Apply S-EC2
Apply S-EC3
540 References from selected papers
32 documents
30 documents (6% discarded)
14 documents (53.3% discarded)
Apply S-EC5
Apply S-EC4
6 documents
7 documents (50% discarded)
12 documents for our SLR
16 documents initially discarded
6 documents
7 documents
Apply S-EC5
Apply S-EC4

Figure 3-4: The final selection process of the snowballing

3.1.2.6. Total publications selected

Figure 3-5 outlines the overall paper selection process in this Master Thesis. From our search in the databases, we selected 15 publications; from the references of those publications, we selected 12 publications, which makes up for a total of **27 publications** that will compose our Systematic Literature Review. Table 3-7 presents the selected literature grouped by its source.

Source	Works
IEEExplore	[SLR5], [SLR6], [SLR7], [SLR8], [SLR9], [SLR10], [SLR11], [SLR12]., [SLR13], [SLR27]
ACM DL	[SLR3], [SLR21], [SLR23], [SLR24], [SLR25]
Springerlink	[SLR1], [SLR2], [SLR4], [SLR14], [SLR15], [SLR22]
Other sources	[SLR16], [SLR17], [SLR18], [SLR19], [SLR20], [SLR26]

Table 3-7: Selected works grouped by their source

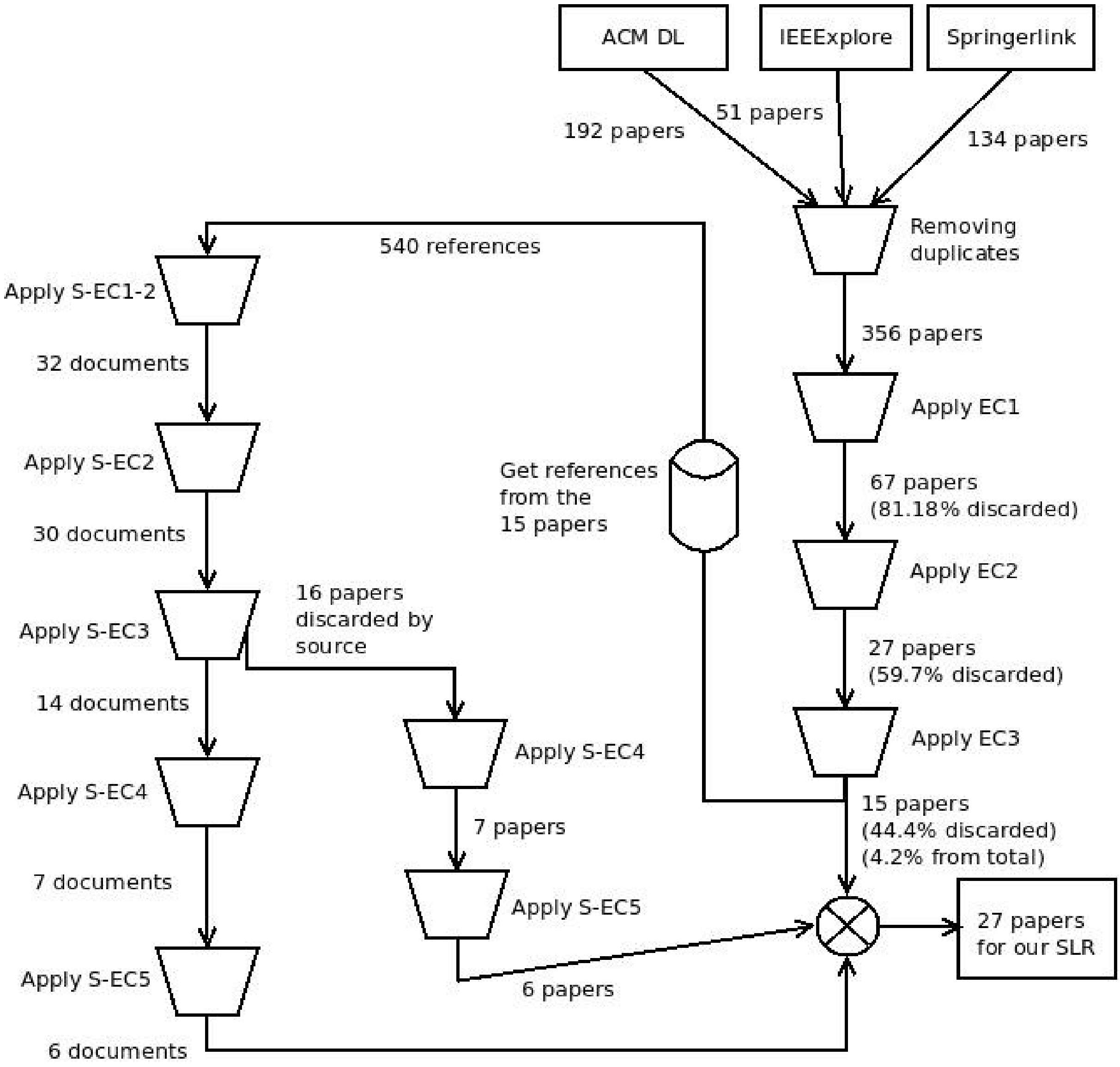

Figure 3-5: The complete selection process of works for the SLR

3.1.2.7. Classification of selected papers

This section will illustrate the classification we have done for the 27 selected publications included in our SLR. The point of classifying the papers is being able to retrieve them quickly later on the SLR. The categorization we followed after reading those papers, We have classified them in four categories as follows:

1) **Reviews:** the content in this category are documents which contain reviews about privacy or confidentiality issues, discussion and general aspects about the topic.
2) **Proposals and solutions**: The content in this category are documents which propose solutions, suggestions or guidelines to solve issues regarding privacy and confidentiality in cloud computing.

Also we did a sub-categorization of papers depending on their “degree of interest” for our SLR:

1) **Least interesting (LI):** the content of the papers and publications from this category are not very suited for this review or contain a small reference to the topic, but at some point they may deserve a mention.
2) **Interesting (I)**: the publications from this category have useful content for this master thesis.
3) **Very interesting (VI)**: the publications from this category deserve special attention as their content may be either broad or their quality is high enough.

Check table 3-9 in section 3.2 for a summary of this classification and how we classified each work.

3.2. Quality assessment of the systematic review

This section contains an analysis of the obtained and selected works in order to illustrate that the publications selected for this review are significant. The process consists in posing several questions which will allow us to assess the quality of works.

Table 3-8 contains the quality assessment filters for the primary studies which will be used to assess the quality of each paper. These filters are based on the exclusion criteria 3 (EC3) we used to select the papers. Also, we are considering too the number of citations papers have received. The number of citations may be an indicator if a paper has a good quality, but not in the inverse order. This is, a high number of citations is an indicator of a good paper, but a low number of citations paper does not mean it is not a good quality paper. They may be too recent to be evaluated, or they just touch a very specific area, reducing the overall focus. Although Cloud Computing has been in the scene for some time, it is still under heavy research focus.

Quality assessment form	
QA1	Is there a clear and understandable statement of the objective of the research?
QA2	Does the paper introduces the concepts and a background of the problem?
QA3	Does the paper include examples of application, case studies or a motivating example?
QA4	Are the conclusions credible and justified?
QA5	Does the study provide value for research and further work?
QA6	Does the paper include a related work section?
QA7	In case of a pure review, does the paper use a Systematic and replicable approach to present results?

Table 3-8: Questions to asses the quality of selected works

The rational of each questions is the following:

1) **QA1** – The work should clearly define and introduce the aim of the work and what it is expected by reading it.

2) **QA2** – Providing background information and concepts helps to understand the problem/solution the are trying to review/propose

3) **QA3** – Including examples helps to illustrate the field of application of a solution or visualize an issue in a real case scenario.

4) **QA4** – Conclusions which clearly synthesize the content of the work help in quickly assessing if the paper fits our interests.

5) **QA5** – A work which poses future work or research directions is useful in order to identify research gaps.

6) **QA6** – Providing a related work section allows readers to get more insight on the matter

and get new points to gather more literature on the topic they are interested, and provide more credibility to the authors' work as it illustrates that there are more researchers who are involved in reaching their same goal.

7) **QA7** – Using a systematic approach to present review results enables the capability to reproduce their results, which adds credibility to the work and allows the expansion of the work to a more recent date.

It is to be noted that these questions are not applicable to every work, and not every posed question is critical. QA1 and QA4 are important because they helped us in selecting papers when applying the exclusion criteria 2 (EC2). QA2 has proven to be common in almost very work we have going through, so it did not contribute very much in assessing the quality of the work. QA3 on the other side helps in making a difference among works, but in purely based review works they are not a mandatory as in works who propose practical solutions. This same rationale applies to QA6 for pure review papers, as they usually intend to point research directions (QA5). Finally, QA7 is a hard-to-assess question as sometimes conference papers have to limit the number of their pages, so there is no space to explain research methodology or information gathering methodology. Also, for papers focused in offering solutions to issues usually there is no systematic review approach. So, for QA7 we put in Table 3-9 a hyphen (-) whenever the methodology used is unknown or not defined, a check (✔) whenever a systematic approach has been used by authors, or a cross (✖) whenever a review has been posed, but the results have not been related to references at any time.

The answer to these questions helped us as well to classify the papers under a Degree of Interest (D.I., see section 3.1.2.7.). Table 3-9 represents the quality assessment, classifying all papers to the specific topic it belongs, the year of its publication, the number of citations, the quality assessment evaluation and the total amount of publications per topic. Each one of the criteria filters on table 3-8 have been graded with a dichotomous “Yes” or “No” scale whether the papers covered them or not. Some papers which provide solutions to certain issues also provide a review to introduce the topic.

In general and according to our assessment, the gathered works provide enough quality to give significant value to our review. It can be noted that in most cases a greater number of citations matches with a higher degree of interest from our point of view. Those works whose citations were not available (NA) were works gathered from heterogeneous sources (works which were not obtained from our selected databases).

Ref.	Cites	Review	Proposals or solutions	D.I.	QA1	QA2	QA3	QA4	QA5	QA6	QA7
[SLR1]	10	✔		LI	✔	✔	✖	✔	✖	✖	✖
[SLR2]	7	✔	✔	I	✔	✔	✖	✔	✔	✔	-
[SLR3]	21	✔	✔	I	✔	✔	✔	✔	✔	✖	✔
[SLR4]	21		✔	I	✔	✔	✔	✔	✔	✔	-
[SLR5]	4		✔	I	✔	✔	✔	✔	✔	✖	-
[SLR6]	85	✔	✔	VI	✔	✔	✔	✔	✔	✔	-
[SLR7]	43	✔		LI	✔	✔	✔	✔	✔	✖	-
[SLR8]	21	✔	✔	I	✔	✔	✖	✔	✔	✖	-
[SLR9]	2		✔	I	✔	✔	✖	✔	✔	✖	-
[SLR10]	95	✔	✔	I	✔	✔	✖	✔	✔	✖	-
[SLR11]	6	✔	✔	VI	✔	✔	✖	✔	✔	✖	✔
[SLR12]	1	✔		LI	✔	✔	✖	✔	✔	✖	✔
[SLR13]	12		✔	I	✔	✔	✖	✔	✔	✔	-
[SLR14]	95		✔	I	✔	✔	✔	✔	✔	✔	-
[SLR15]	2		✔	I	✔	✔	✔	✔	✔	✔	-
[SLR16]	301	✔	✔	VI	✔	✔	✔	✔	✔	✔	-
[SLR17]	NA	✔	✔	VI	✔	✔	✖	✔	✔	✔	✖
[SLR18]	NA	✔	✔	I	✔	✔	✖	✔	✔	✔	-
[SLR19]	NA	✔	✔	VI	✔	✔	✔	✔	✔	✔	-
[SLR20]	NA	✔	✔	VI	✔	✔	✔	✔	✔	✔	✖
[SLR21]	21	✔	✔	I	✔	✔	✖	✔	✔	✖	✔
[SLR22]	NA		✔	I	✔	✔	✔	✔	✔	✔	-
[SLR23]	7	✔	✔	VI	✔	✔	✖	✔	✔	✖	-
[SLR24]	5	✔	✔	I	✔	✔	✔	✔	✔	✖	-
[SLR25]	17		✔	I	✔	✔	✔	✔	✔	✔	-
[SLR26]	NA	✔	✔	LI	✔	✔	✔	✔	✔	✖	-
[SLR27]	25	✔	✔	I	✔	✔	✖	✔	✔	✖	✔

Table 3-9: Summary of the attributes of the selected works

CHAPTER 4

PRIVACY AND CONFIDENTIALITY ISSUES AND CHALLENGES IN CLOUD COMPUTING

SYSTEMATIC REVIEW RESULTS ANALYSIS

Chapter 4: Privacy and confidentiality issues and challenges in Cloud Computing: Systematic Review results analysis

This chapter is the main contribution of this Master Thesis. In it, we will synthesize and describe all the evidence found and collected from the 27 selected papers in the literature regarding privacy and confidentiality, identify issues and challenges and any found solution to those issues, as well as trying to establish a relationship among the found literature in order to find any possible gap or open the door to possible relationships.

The chapter is structured as follows:

1) Section 4.1 will introduce and explain what is privacy and the found definitions in the literature.

2) Section 4.2 will introduce and explain what is confidentiality and what differences it from privacy.

3) Finally, section 4.3 will expose all the issues and challenges related to privacy and confidentiality in cloud computing in a categorized way, listing any remarkable evidence found as well as proposed solutions and suggestions from the different authors.

4.1. What is Privacy?

Privacy is a wide concept that varies among countries, cultures and jurisdictions. Giving a precise definition is difficult if not impossible, and this matter, by itself, poses a problem when trying to establish a consensus.

In the literature, we have found 4 relevant documents related to the definitions about privacy in cloud computing and general information about it: [SLR16], [SLR12], [SLR23] and [SLR27].

Both Mather et al. [SLR16] and X. Ma [SLR12] perform general reviews about privacy concerns in cloud computing (being [SLR16] the one which provides more insight). Pearson and Charlesworth [SLR23] also review some privacy issues (focusing on accountability as a solution, which will be reviewed later in section 4.3.2.1.5 - Audit, monitoring and accountability). They expose background information about the topic and several related concerns which will be discussed and classified through this Systematic review.

Below we list several definitions for privacy which have been identified in the literature

- The definition adopted by the Organization for Economic Cooperation and Development (OECD): *"It is the status accorded to data which has been agreed upon between the person or organization furnishing the data and the organization receiving it and which describes the degree of protection which will be provided"* [SLR16]

- The definition provided in the Generally Accepted Privacy Principles (GAPP) standard: *"The rights and obligations of individuals and organizations with respect to the collection, use, retention, and disclosure of personal information."* [SLR16][SLR27]

- The definition provided by Oxford Dictionary: *"privacy is defined as "a state in which one is not observed or disturbed by other people" and "a state of being free from public attention." More specifically, privacy rights or obligations are related to the collection, use, disclosure, storage, and destruction of personal data or personally identifiable information."* [SLR12]

- *Privacy is a fundamental human right that encompasses the right to be left alone [...]. In the commercial, consumer context, privacy entails the protection and appropriate use of the personal information of customers, and the meeting of expectations of customers about its use. For organisations, privacy entails the application of laws, policies, standards and processes by which Personally Identifiable Information of individuals is managed* [SLR23].

Some differences can be noted in the definitions. While OECD defines privacy as a direct link between an individual and its related data, GAPP, Oxford Dictionary and [SLR23] go one step further, including to the definition the right management of that information. Given the concerns arisen on privacy, a deep and complex definition of Privacy is necessary to correctly define it, but the general idea is that privacy is related to the collection, use, disclosure, storage and destruction of personal data.

Also, different legislations rise different definitions and conceptions of privacy. According to Mather et al. [SLR16] in EU privacy is a basic right, whereas in U.S. it is more centred on avoiding harm. However, the broad principles and definitions described above would apply to most countries.

Privacy is a key business risk and compliance issue and it is, basically, what keep customers away or worried about the cloud. Pearson and Charlesworth [SLR23] suggest that, in order to conform to legal privacy requirements regarding personal information in the cloud, corporations need to demonstrate an appropriate level of control over the data at all stages of its processing, from collection to destruction.

As a background information on privacy and related to the collection and destruction of data mentioned by [SLR23], X. Ma, Mather et al. and Chen et al. [SLR12][SLR16][SLR27] define this process as "**the Data Life Cycle**". Figure 4-1 illustrates this process with its phases.

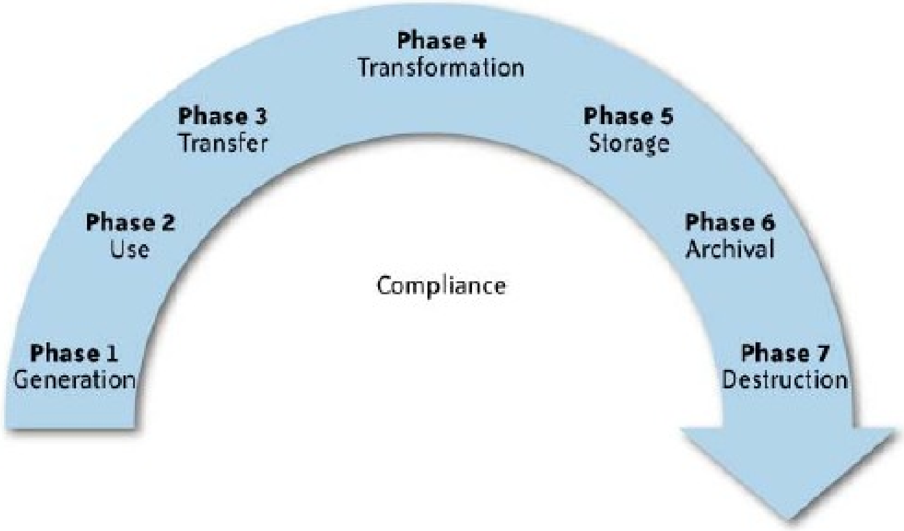

Figure 4-1: The data life-cycle

Table 4-1 describes each one of this phases and which are its main responsibilities.

According to Ma [SLR12], due to poor management or malicious employees of the service providers, data may be leaked to third parties. Such leakage might happen at any stage of the data life cycle. Therefore it is important to investigate what happens with the data in each stage of data life cycle, even if this investigation may raise some issues as explained later in section 4.3.2.1.5.

Besides acknowledging the data life-cycle process, its also necessary to acknowledge what kind of information needs to be managed and protected. On its paper, S. Pearson [SLR17] makes an analysis over several privacy issues and challenges and describes three types of privacy sensitive information:

- Personally identifiable information (PII): any information that identifies an individual (name, address, IP address, credit card number, biometric information...)

- Sensitive information: information on religion, race, health, financial information... any type which is considered private.

- Usage data: usage data collected from the use of computer devices or actions performed (use a printer, visit a web page, etc...)

Phase	Description: This phase deals with...
Generation of the information	- Who owns the data - How this ownership is maintained - How and when personal information is classified - Defines the governance structure that manages and protects this data through the cycle.
Use	- How and where personal information is used - If information is shared with third parties (i.e. CSP subcontractors) - If the information is used with the purpose it was collected - If information enables organization to comply legal requirements.
Transfer	- How information is transferred to the cloud (transport protocols) - If information is protected appropriately - If the transfer required encryption - If there are appropriate access controls
Transforming and sharing	- What happens when personal information is transformed and shared in the cloud. - If transformed and shared data maintains original protections and restrictions - If integrity is maintained - Isolation of sensitive information from original data
Storage	- How is data stored - If there are appropriate controls over personal information - How is integrity, availability and confidentiality maintained - If the CSP supports encryption and key management
Archival	- What happens with data already stored in the cloud - In which media is this data stored and who controls it - How long should data be retained
Destruction	- The removal of data from the cloud and make sure if the data is destroyed in a secure manner or can be recovered somehow

Table 4-1: Data life-cycle phases with descriptions

Finally, Pearson summarizes several key requirements privacy should fulfil in order to keep data protected. Table 4-2 shows these requirements and their descriptions.

The non-fulfilment of these requirements is, in fact, what “unleashes” the issues that will be reviewed in this Master Thesis in Section 4.3.

Requirement	Description
Notice, openness and transparency	- Users must be notified about: • Which data is being collected • How they want to use it • How long they will keep it • With whom will this data be shared • Any other use - Privacy policies must be made available - Personal information should be collected directly from the person unless there is a good reason to not do so.
Choice, consent and control	- Users should decide whether they want personal information to be collected/disclosed or not.
Scope / minimisation	- Only information that is required to fulfil the stated purpose should be collected or shared
Access and accuracy	- Users must be able to • Get access to their personal information • Check the accuracy of data
Security safeguards	- Safeguards must prevent unauthorized access, disclosure, copying, use or modification of personal information
Compliance	- Transactions must be compliant to privacy legislation
Purpose	- Data usage has to be limited to the purpose for which it was collected. - There must be a clearly specified purpose for the collection and sharing of personal information.
Limiting use, disclosure and retention	- Data can only be disclosed to those parties authorized to receive it - Data should only be kept as long as it is necessary.
Accountability	- An organization must appoint someone to ensure that privacy policies and practices are followed. - Audit functions must be present to monitor all data accesses and modifications.

Table 4-2: Privacy requirements and descriptions

In chapter 5, as an additional contribution to this Master Thesis, we will propose a binding for the described data life-cycle phases and privacy requirements to the identified issues in section 4.3. This way we want to try establishing a direct connection between phases, requirements and issues which may serve to direct future research areas.

4.2. What is confidentiality?

Confidentiality is a rather blurry term in the literature and it doesn't seem very clear in the research community which are the key differences between privacy and confidentiality. Very often authors refer to privacy as a global concept when they propose their solutions or perform their reviews whilst, according to the definitions of confidentiality provided below, they could use this term as well.

In the literature, 3 papers make a definition over the term confidentiality: [SLR9], [SLR10] and [SLR11].

Some definitions about confidentiality that have been found in the gathered evidence are:

- *“The property that information is not made available or disclosed to unauthorized individuals, entities, or process.”* [SLR9], definition given by ISO/IEC 2700
- *“Keeping users' data secret in the Cloud systems”* [SLR10].
- *“Confidentiality implies that customer's data and computation tasks are to be kept confidential from both the cloud provider and other customers.”* [SLR11].

These definitions enter in contrast with the ones provided in the previous section about privacy. For example, in [SLR12] X. Ma defined privacy as *“a state in which one is not observed or disturbed by other people”* and *“a state of being free from public attention. More specifically, privacy rights or obligations are related to the collection, use, disclosure, storage, and destruction of personal data or personally identifiable information.”*. As it can be noted, confidentiality is illustrated more specifically related to the domain of “keeping data secret and unavailable” whilst privacy is a more global concept, which entangles rights, obligations and management over this data.

Summarizing the evidence in a single definition, we could say that confidentiality is a characteristic of privacy which focuses on protecting information leakage by establishing measures to prevent the unwanted disclosure of data. Confidentiality, as we will see through section 4.3, is commonly achieved through securing technology with techniques like encryption, access management and virtualization robustness.

In fact, in their review paper about security in Cloud Computing, Xiao et al. [SLR11] explain that confidentiality is an attribute of privacy (saying, therefore and as explained previously, that privacy is the global concept which involves data protection, legislation, and management while confidentiality is more specific to the protection and harm-prevention of data), and regards privacy-preservability as the core attribute of privacy, among confidentiality, integrity and accountability. What's more, the authors consider that privacy-preservability is a stricter form of confidentiality, due to the notion that they both prevent information leakage. If cloud confidentiality is ever violated, privacy-preservability will also be violated. This could be one of the reasons why confidentiality is a term which appears less often in the literature, as most solution that deal with data protection and safety satisfy confidentiality and protects privacy.

4.3. Privacy and Confidentiality concerns, issues and challenges

Whenever an individual customer, a business, a government agency, or another entity uploads data to the cloud, privacy and confidentiality questions may arise [SLR21]. Users expect that cloud providers will protect their data from unauthorized access, and sensitive data will remain private. They also expect that any third parties like governments will not monitor their activity. Even if a Cloud Service Provider takes all the necessary steps to protect this data like using encrypting algorithms, are such efforts sufficient?

Several authors in the literature emphasize in the matter that the security and privacy mechanisms provided by cloud companies are not adequate and result in a big obstacle for users to adopt Cloud Computing technologies [SLR10]. Cloud computing is capable of handling mass data storage and intense computing tasks, so traditional security mechanisms which sufficed on a Desktop environment (e.g. anti-virus, firewalls, etc.) may not be enough in the cloud due to heavy computation or communication overhead [SLR11]. In addition, privacy threats differ depending on the scenario (privacy requirements are different on a banking case than on a mail service, for instance) [SLR17].

This section is the backbone of this systematic literature review. In it we will summarize all the evidence found in the literature about different issues and challenges regarding privacy and confidentiality in Cloud Computing, and research directions/solutions/suggestions authors outline to tackle these problems. More specifically, we provide a classification of privacy issues and the approaches found in the literature to address them or, if no approaches or solutions have been found, mention any posed questions found in the literature to step forward for research in the specific topic/issue.

The literature on this topic is fairly spread and every author mentions several and different issues, some of them which may be connected with each other. For the sake of ease the tracing of this document and clarify the relationships between works, table 4-4 lists the categorization we performed over the concerns about privacy and confidentiality found in the literature. It displays which works have been found for every category, which are the identified issues for each category, and if the issues found relate to privacy (general issues over management and rights over data), confidentiality (measures to prevent data leakage) or both.

Each subsection which contains identified privacy and confidentiality concerns is attached with a figure similar to Figure 4-2 and a table similar to table 4-3. The goal for doing this is to provide and give a quick visual understanding for each identified concern.

Figure 4-2 represents a template to display the interactions of different cloud actors, (most commonly Cloud customer and the Cloud provider, refer to section 2.2 to see the list of Cloud actors) and the questions that they may rise on the usage of the cloud, which, alongside the requirements described in section 4.1, arise the concerns that have been identified. Cloud actors may direct their questions at other actors or themselves (self-questions).

Table 4-3 represents a template to expose which are the identified issues for each category and its respective solutions, alongside a reference to the work that proposed that solution. The works that can be found in these tables are the ones which provide proposals or solutions as noted in the table 3-9 in section 3.2.

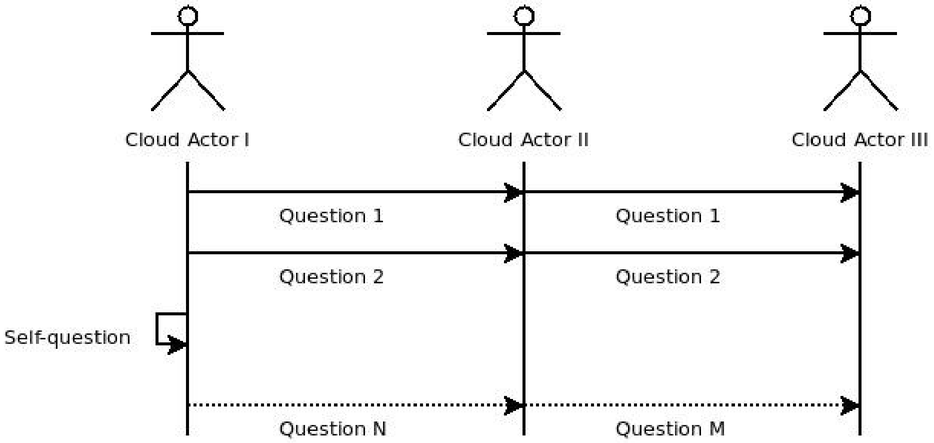

Figure 4-2: Meta-figure which represents the question Cloud actors raise

Issue	Proposed solutions / responses / suggestions
Issue 1	- Proposed solution 1 [#work]
Issue 2	- Proposed solution 1 [#work] - Proposed solution 2 [#work]
Issue 3	- Proposed solution 1 [#work]

Table 4-3: Meta-table which exposes the issues and challenges found alongside works and their solutions

Issue/Challenge			Identified works	Identified issues	Refers to...
Virtualization			[SLR8], SLR11], [SLR13], [SLR19], [SLR20], [SLR23]	- Cross-VM attack - Malicious SysAdmin - VM hopping	Confidentiality
DATA LOCATION	Data Outsourcing	General issues	[SLR17], [SLR19], [SLR20], [SLR21], [SLR23], [SLR26]	- Security: Cloud Computing vs traditional computing - How to build up trust and confidence in customers	Privacy
		Lack of execution controls, Customers' privacy requirements and policies	[SLR3], [SLR6], [SLR8], [SLR9], [SLR18], [SLR19], [SLR20], [SLR26]	- Customer's privacy requirements may not fit to the conditions or policies for a Cloud Service	Privacy
		Data ownership	{SLR16], [SLR20], [SLR27]	- Who owns the data in the cloud	Privacy
		Changeability of terms	[SLR8], [SLR16], [SLR18]	- Bankruptcy of CSP - Change of terms and termination of services	Privacy
		Responsibility for protecting privacy and privacy policy enforcement	[SLR1], [SLR5], [SLR16], [SLR19], [SLR20], [SLR24]	- Who is responsible for protecting privacy - Know if the CSP is fulfilling my requirements	Privacy
		Audit, monitoring and accountability	[SLR10], [SLR11], [SLR16], [SLR17], [SLR20], [SLR23], [SLR26]	- When does information gathering become intrusive - How can I be sure that my CSP fulfils my requirements?	Privacy
		Unauthorized use of data and access control	[SLR7], [SLR16], [SLR18], [SLR19], [SLR20], [SLR22], [SLR24], [SLR27]	- Prevent use of my data in ways I have not authorized - Prevent access to unauthorized parties	Confidentiality
		Multiple service	[SLR24]	- Ensure correct policy integration among CSP	Privacy

Table 4-4: Summary of Privacy and Confidentiality categorization, works and solutions

		composition			
		Privacy breaches notification	[SLR16]	- How to ensure that CSP notifies you when a breach occurs - Who is responsible for managing the breach notification process - If the contract policy include liability	Privacy
	Offshoring	General issues	[SLR10], [SLR14], [SLR16], [SLR18], [SLR19]	- Which jurisdiction applies? - How is compliance impacted by moving to the cloud and where should I store my data	Privacy
		Legal and regulatory issues	[SLR10], [SLR16], [SLR18]	- Different points of view and concepts about privacy. Conflicting regulations - Outdated and inapplicable laws and regulations	Privacy
Data combination and commingling			[SLR10]	- Is my data separated from other tenants	Confidentiality
Encryption			[SLR2], [SLR4], [SLR5], [SLR7], [SLR9], [SLR10], [SLR11], [SLR14], [SLR18], [SLR19], [SLR24], [SLR25], [SLR27]	- Data retrieval efficiency while keeping confidentiality - Confidentiality protection of data - Secure communications and prevent vulnerabilities	Confidentiality
Storage, retention and destruction of data			[SLR8], [SLR12], [SLR16], [SLR20]	- Who enforces retention policy - In which media is data stored - Prevent data concentration vulnerabilities and the stored data is correctly isolated - How much time is my data kept - How to make sure data is properly destroyed	Privacy and Confidentiality
Economic Cost			[SLR3], [SLR6], [SLR15], [SLR20], [SLR23]	- Which cost models are available while preserving privacy - How to keep privacy costs under control	-

Table 4-4 (bis): Summary of Privacy and Confidentiality categorization, works and solutions

4.3.1. Virtualization

Virtualization refers to the abstraction of computer resources and it is a key feature of cloud computing [SLR13] (see Section 2.5.1 – Technological components). Virtualization technology allows dynamic resource allocation and service provisioning, and allows multiple OSs co-reside on the same physical machine without interfering each other. Many virtualization technologies have been proposed and implemented, such as Xen, VMware.

In the literature, 6 papers have been found about issues or mentions regarding virtualization and data privacy: [SLR8], [SLR11], [SLR13], [SLR19], [SLR20] and [SLR23]. [SLR8], [SLR11], [SLR13] and [SLR19] provide actual solutions to the concerns that will be enumerated and described through this section. [SLR20] and [SLR23] provide reviews that complement explanations and helped identifying and assessing these concerns. Figure 4-3 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-5 summarizes the found issues and their proposed solutions about virtualization in the literature.

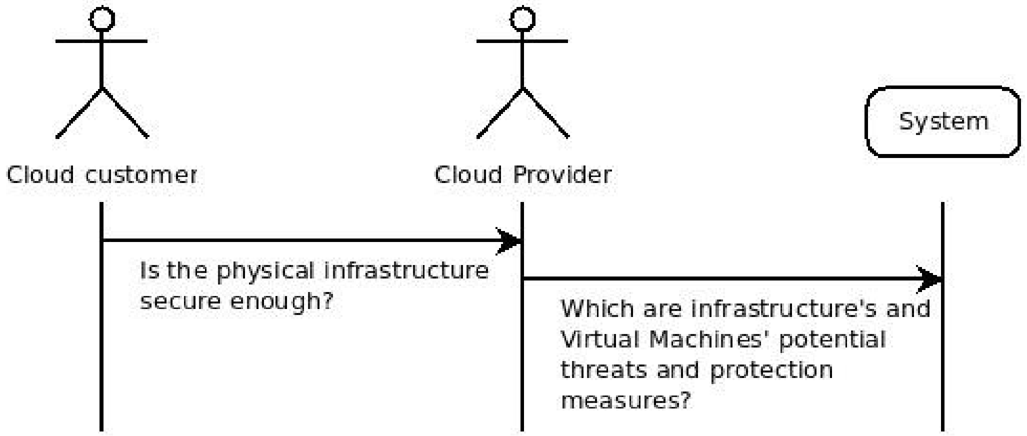
Cloud customer
Cloud Provider
System
Is the physical infrastructure secure enough?
Which are infrastructure's and Virtual Machines' potential threats and protection measures?

Figure 4-3: Questions raised by actors about Virtualization

Issue	Proposed solutions / responses / suggestions
Cross-VM attack	- Placement prevention [SLR11] - Co-residency detection [SLR11] - NoHype [SLR11]
Malicious SysAdmin	- Trusted Cloud Computing Platform (TCCP) [SLR8,SLR11, SLR13] - Combine TCCP with a custom guest VM [SLR13]
VM hopping	- Cover hypervisor security holes [SLR19]

Table 4-5: Summary of identified issues related to virtualization

There are security risks in sharing machines. Basically, and according to Pearson and Charlesworth [SLR23] and Xiao et al. [SLR11] virtualization arises the problem of losing control over data location, who has access to it, and multi-tenancy which, although provides lot of benefits, also arises security issues which may be exploited. NIST [SLR20] adds to the matter that attackers can easily focus their attacks on a single machine to affect multiple victims due to the hypervisor (according to the definition of NIST, the hypervisor is a layer of software between an operating system and hardware platform that is used to operate Virtual Machines like launch or terminating). Hypervisor adds APIs (Application programming interfaces), channels (sockets) and data inputs (like strings) which increase the complexity of a system and increases the surface attack. This point of view seems to not be shared by the ENISA report [SLR19], which say that attacks on hypervisors are less numerous and more difficult to put in practice for an attacker compared to attacks to traditional OS.

Xiao et al. [SLR11] focus on a couple of issues regarding virtualization: **Cross-VM attack** (which may be performed by other tenants in the cloud) and **Malicious SysAdmin** (which may only be performed by an inside attack from cloud vendor). In the ENISA report [SLR19], one more vulnerability is described: **VM hopping**.

Cross-VM attack is a vulnerability that allows a malicious adversary steal information without leaving a trail or raising alarms. To do this, a malicious VM is placed on the physical server where the target client's VM is located. To achieve this, an adversary should first determine where the target VM instance is located (this can be done with network probing tools such as nmap, hping, wget, etc.); then, the adversary should be able to determine if there are two VM instances by comparing IP addresses to see if they match, and measuring the small packet round-trip time.

CSPs like Amazon EC2 launch new instances of VMs on the same set of physical machines, so there is a good chance that the malicious user is placed on the same physical server, and once the malicious VM has been created alongside the victim VM in the same physical machine, both share certain physical resources like data cache, network access, CPU pipelines, etc... which the adversary can employ to attack the VM: measure cache usage to estimate current load of server, estimate traffic rate, keystroke timing that can steal victim's password...

To address this, Xiao et al. [SLR11] propose the following defence strategies against cross-VM:

- Placement prevention: consists in obfuscating co-residence in Virtual Machines by having Dom0 (Dom0 or Domain0 is the superior layer that manages the hardware of Virtual Machines) not respond in traceroute (a diagnostic tool for displaying the route that network packets follow), or by randomly assigning internal IP addresses to launched VMs.

- Co-residency Detection: this approach completely eliminates co-residency. However this goes against the economical benefits and cost-savings of virtualization, so left options are sharing infrastructure with trustworthy customers or VMs owned by the same customer.

- NoHype: NoHype attempts to minimize the degree of shared infrastructure by removing the hypervisor while still retaining the key features of virtualization. This eliminates L1 cache side channel (which may be used for attacks as L1 caches contain potential damaging side-channels, refer to [44] for more information) and retain multi-tenancy properties. However NoHype requires changing hardware, making it less practical. Removing the hypervisor may also deal with the issues previously explained by NIST [SLR20] which relate to the hypervisor, as compromising the hypervisor also compromises all the systems that it hosts.

Malicious SysAdmin is a vulnerability in which a privileged administrator of the cloud provider can perform attacks by accessing the memory of a customer's VMs. For instance, Xenaccess enables an administrator to directly access the VM memory at run time by running a user level process in Domain0. As cloud use increases, chance of malicious activities by cloud provider employees can heavily impact confidentiality, availability and integrity of all data. This is specially important as Cloud Computing architectures need some roles which are very high-risk.

To address this issue, Sengupta et al. [SLR8], Xiao et al. [SLR11] And J. Kong [SLR13] propose the use of **Trusted Cloud Computing Platform (TCCP)** to achieve data confidentiality for guest virtual machines. TCCP is a set of technologies developed by the Trusted Computing Group (TCG) to face the concern of untrusted execution environment. TCCP guarantees that the execution of virtual machines doesn't leak any information and keeps data confidential. TCCP does this by closing the VM execution inside a secure perimeter and making the CSP administrator with root privileges unable to access the memory of a hosted VM. TCCP are usually based on a Trusted Platform Module (TPM), which is integrated on the motherboards, and includes some technologies such as Remote attestation, sealed storage and authenticated booting.

Regarding confidentiality preservation and TCCP, J. Kong [SLR13] proposes an architecture to combine machine virtualization technology with TCCP to achieve a greater degree of confidentiality on virtual machines. This architecture consists in customizing the guest VM operating system, disabling all the unnecessary virtual devices, and disallowing code from the CSP to be executed in the guest VM. By doing this, clients' data confidentiality is greatly enhanced against the service provider, as only valid users are able to boot the target VM. Kong explains this architecture in its paper, which consists of three differentiated parts as it can be seen in figure 4-4: Part #1 contains the trusted part (which includes hardware platform and a trusted Xen hypervisor), part #2 contains the untrusted part (parts controlled by the CSP), and part #3 contains the protected part (which includes the guest VM environment).

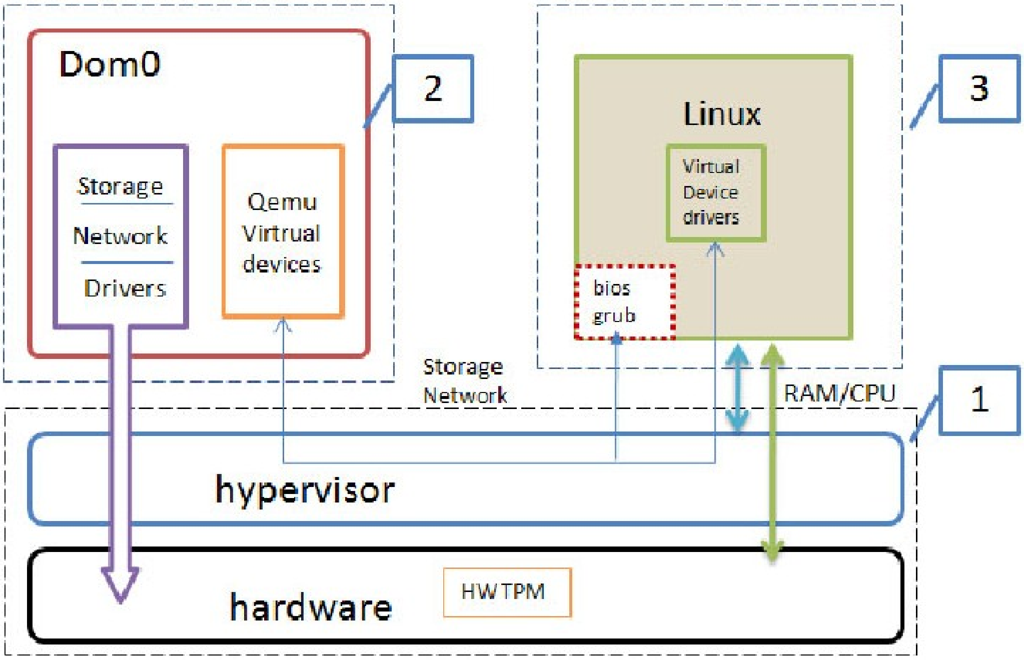

Figure 4-4. Overview of J. Kong proposed the architecture to achieve confidentiality on VMs

VM hopping is a vulnerability in which an attacker hacks a Virtual Machine using some method like the ones described above and then, by exploiting any hypervisor vulnerability, takes the control of other VMs which are under the management of the same hypervisor. However, according to ENISA [SLR19], this kind of vulnerability relies on the supposition that the hypervisor has a security hole. On a robust hypervisor this kind of attacks should not happen.

4.3.2. Data location

When information crosses borders (whether they are international borders or company's boundaries), privacy issues arise a variety of concerns. Usually, customers do not know where the data is [SLR27], and constraints on the flow of personal and sensitive data, as well as constraints in the requirements on the protection of data, have become the subject of national and regional privacy and security laws and regulations.

Data location is one of the most common compliance issues an organization faces with its data and has a significant impact on privacy and confidentiality protections [SLR16][SLR18]. In Cloud Computing, data may be stored redundantly in multiple physical locations and detailed information about the location of the data may be unavailable to the service consumer. This situation makes it difficult to predict whether legal and compliance requirements are being met, as different privacy laws in various countries limit the ability of organizations to transfer some types of personal information.

In the literature, data location issues are classified in three ways. Pearson and Charlesworth [SLR23] divide data location issues into two categories, outsourcing and offshoring, whilst Zhou et al. [SLR10] add one more category, data combination and commingling. Next subsections will review these 3 categories.

4.3.2.1. Data Outsourcing

Data outsourcing refers to the fact that data crosses company's boundaries and is hosted on another company (the CSP). This also means that customers physically lose control on their data, and this loss of control is one of the main causes of cloud insecurity and raises governance and accountability questions [SLR11][SLR12][SLR16][SLR17][SLR23].

Data Outsourcing is one of the issues that arise more concerns about privacy in Cloud Computing. As condensing all the issues related to Data Outsourcing in one section would be rather long, we decided to split them into subsections which will allow us to present them in a more organized way.

4.3.2.1.1. Data Outsourcing: General issues

This section will review general issues found in the literature which regard to the outsource of data. In the gathered evidence, there are 6 papers which deal with issues regarding data location, [SLR17], [SLR19], [SLR20], [SLR21], [SLR23] and [SLR26]. These papers, alongside reviews on data outsourcing, provide actual solutions to some identified issues for this area. Figure 4-5 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-6 summarizes the issues found and solutions proposed about general issue on data outsourcing in the literature.

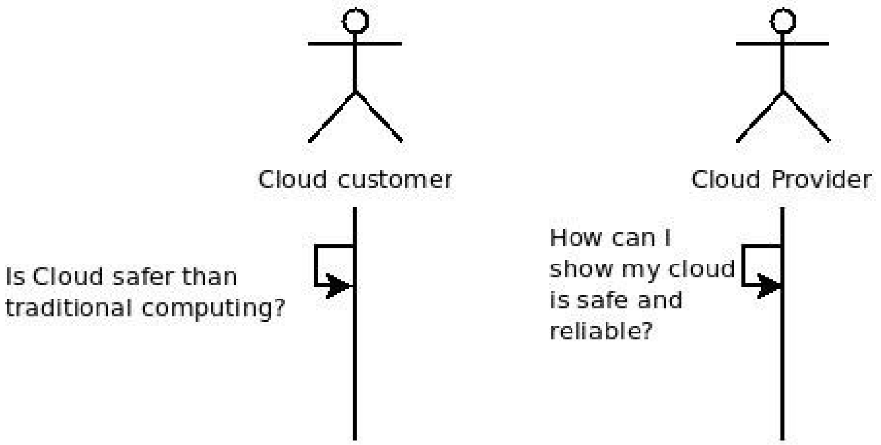
Cloud customer
Cloud Provider
Is Cloud safer than traditional computing?
How can I show my cloud is safe and reliable?

Figure 4-5: Questions raised by actors about general issues on data outsourcing

Issue	Proposed solutions / responses / suggestions
Security: Cloud Computing vs traditional computing	- Cloud Computing is more secure than traditional PC due to CSP expertise in security [SLR23] - Levels of privacy and anonymity will be lower [SLR26]
How to build up trust and confidence in customers	- Organisations should follow risk assessment frameworks and practices like Privacy Impact Assessment (PIA) [SLR20] - Standard organisations should develop cloud certifications and privacy templates [SLR17, SLR19, SLR21] - Cloud Service Providers should follow sets of best practices like the use of Privacy-Enhancing Technologies and techniques like obfuscation and maximise user control [SLR17, SLR19]

Table 4-6: Summary of identified issues related to general issues of Data Outsourcing

On a traditional PC or servers owned by a company there is control over how the data is stored and there are controlled restrictions on who can access it. For cloud computing, the data is stored on the server and the third-party company is responsible for deciding the details of data storage. Due to the fact that sensitive data is out of the owners' control, outsourcing will potentially incur privacy violations.

Nevertheless, and according to Roberts et al. [SLR1], Pearson and Charlesworth [SLR23] and ENISA [SLR19], outsourcing the data to the Cloud may (arguably) be even a safer option than storing it on a traditional PC. The Cloud Provider is usually comprised by experts responsible of maintaining the security of information, whilst most individuals do not have the expertise nor will to implement bleeding edge security features on their home PCs. This fact leads to the conclusion that storing data on the cloud may be actually safer than storing it indoors. However this point of view is not shared by Jaeger et al. [SLR26] who state that the levels of privacy and anonymity for a cloud user will be lower than a desktop user. Some consensus on this topic is required as future research.

In order to provide more confidence to customers when outsourcing data to the cloud, NIST Guidelines [SLR20] provides information and proposes a framework outlining which activities and steps an organisation should follow in order to ensure that the outsourcing of data is done in a secure way and in compliance with all organisational policies while maintaining privacy. The recommended steps to follow are: requirements specification (personnel requirements, access controls...), security and privacy risks assessment, and assess the competency of the Cloud Provider (experience, quality, track records...).

Similar to the NIST guidelines, in the ENISA report [SLR19] and the paper of S. Pearson [SLR17], the authors do some recommendations and provide a set of practices to build up trust which include:

- The definition of standard certifications and creation of Cloud computing security life-cycle standards. According to Subashini and Kavitha [SLR21], the Cloud Security Alliance (CSA) is an organization focused on developing standards and security solutions for the cloud.

- The definition and creation of privacy templates and patterns to fit several kinds of scenarios and help engineers, and the application of the Privacy Impact Assessment (PIA) which aims to help organizations assess the impact of operations on personal privacy, define the privacy requirements and identify problems related to proposed privacy solutions.

- Use Privacy-Enhancing Technologies (PET) like secure online access mechanisms, privacy management tools which allow server-side inspection of policies, and pseudonymisations tools to hide true identity.

- Minimise personal information sent and stored in the cloud and apply anonymisation techniques like obfuscation.

- Protect personal information in the cloud with tamper-resistant hardware during transfer and storage of data and protect information with access controls.

- Maximise user control to increase trust and confidence in the cloud. Approaches that can be used are the definition of personal information management preferences.

- Specify and limit data usage to the purpose it was collected. Mechanisms like Digital Rights Management (DRM) techniques can be used to achieve this.

- Provide feedback and design human interfaces to inform users about privacy functionality.

4.3.2.1.2. Lack of execution controls, customers' privacy requirements and policies

Customers do need different degrees of privacy on its data. For example, Healthcare businesses do need specific necessities whilst a SME (Small or Medium Enterprise) or a startup may need other ones which do not require a high degree of confidentiality and protection measures. Usually CSP offer their clients identical services with rigid Service Level Agreements (SLA), even though their confidentiality requirements deviate from each other [SLR9]. According to ENISA report [SLR19], a poor provider selection can affect company reputation, customer trust and service delivery. If CSP could include better policies and practices, users could be able to better assess privacy and confidentiality risks they face [SLR18].

What's more, cloud customer does not have a complete control over remote execution environment (memory, access to external shared utilities, etc.) [SLR8] nor its data [SLR20], and it can be difficult for the customer to check if the data processing is done in a lawful way [SLR19], so customers would want to inspect the execution traces to ensure that illegal operations are not performed on his cloud environment.

In the literature, 8 papers have been identified which provide solutions to address these issues and suit customers' privacy requirements or make a mention about it: [SLR3], [SLR6], [SLR8], [SLR9], [SLR18], [SLR19], [SLR20] and [SLR26]. [SLR3], [SLR6], [SLR9] and [SLR18] provide actual solutions to the concerns that will be enumerated and described through this section. [SLR8], [SLR19], [SLR20] and [SLR26] provide reviews that complement explanations and helped identifying and assessing these concerns. Then, figure 4-6 symbolizes the questions that the main Cloud actors arise on the matter. Finally, table 4-7 summarizes the evidence found about the topic of fitting customer's privacy requirements and issues related to the lack of control.

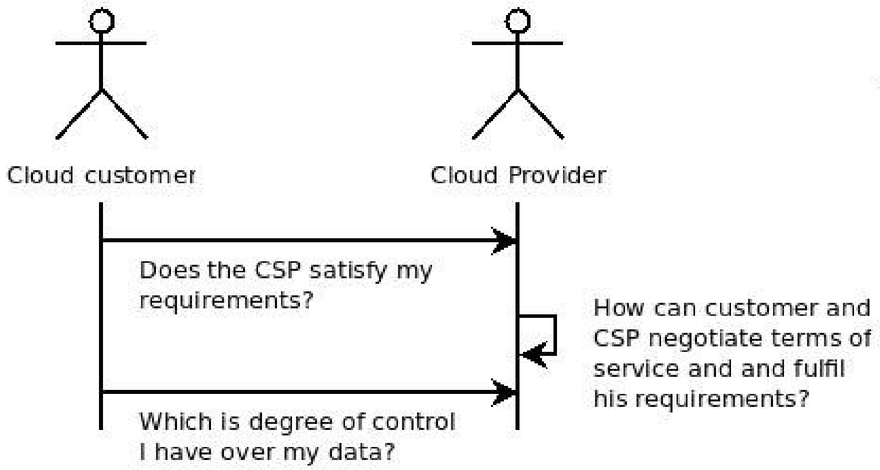
Cloud customer
Cloud Provider
Does the CSP satisfy my requirements?
How can customer and CSP negotiate terms of service and and fulfil his requirements?
Which is degree of control I have over my data?

Figure 4-6: Questions raised by actors about lack of execution controls, customers' privacy requirements and policies

Issue	Proposed solutions / responses / suggestions
Customer's privacy requirements may not fit to the conditions or policies for a Cloud Service	- Policy-driven frameworks to establish integrated policies and identify best providers [SLR3] - ScoRiM [SLR9] - Privacy as a Service [SLR6] - Establish standards in cloud computing paradigm to ease the analysis of different CSP [SLR18]

Table 4-7: Summary of identified issues related to lack of execution controls, customers' privacy requirements and policies

As stated previously, customers deal with large amounts of cloud service providers, which overwhelms customers as they have to manually identify which of them meets his privacy requirements [SLR26]. Gellman [SLR18] suggests to cloud computing industry to establish standards in the paradigm, as the current lack of them makes it difficult to users to analyse and assess the differences between CSPs. Commonly, when a cloud service provider is chosen, the established privacy policy between the user and the CSP may not exactly meet or are incompatible with user's privacy requirements. According to NIST [SLR20], inadequate policies may lead to undetected intrusion and violation policies due to insufficient auditing and monitoring and loss of privacy because the CSP handles sensitive data in a less rigorous way than organisation's.

New policy requirements could be written for all participating parties, but this can be very time consuming as communication and negotiation efforts are required. It may even become more problematic if users are using multiple cloud providers (even if they are not aware of that, see section 4.3.2.1.8 for more information).

To deal with this problem, Lin et al. [SLR3] propose a **policy-driven framework** which integrates three key functions: (1) policy ranking that helps users quickly identify a suitable policy provider; (2) automatic policy generation that takes policies and requirements from the user and the service providers to automatically generate an integrated policy to be adopted by the participating parties; (3) policy enforcement that enforces privacy policies across multiple parties.

Figure 4-7 illustrates this framework with an example: a user joined the cloud and faces six cloud service providers, each of them able to provide the service that user's needs. In order to find the service provider whose privacy policies best fit user's privacy requirements, the user's privacy requirements and policies from service providers are sent to the **policy ranking module** together. The ranking module helps select service provider S2 for the user by comparing the customer requirement with the policies of multiple service providers and subsequently picking the one with the highest rank. The second step is to send their policies to the **policy integration module** which will automatically generate an integrated policy as agreed by both parties. Throughout the service, user's data privacy will be protected by the executable policy and the executable policy may also travel among contractors associated with service provider S2.

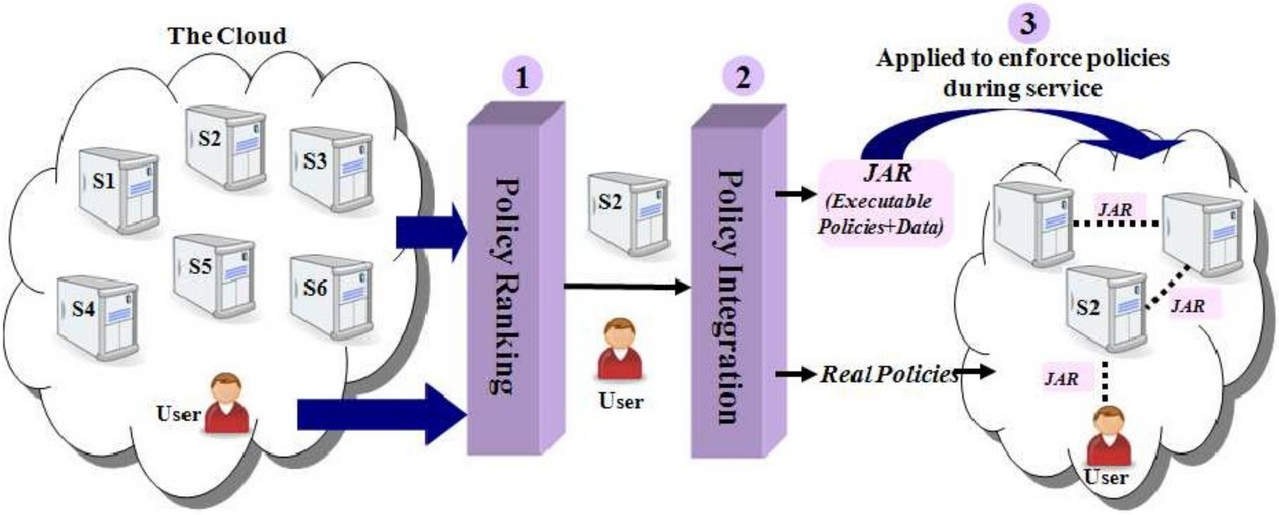

Figure 4-7: An Overview of the Lin et al. Policy-driven framework

Another solution proposed by Chou et al. [SLR9] is **SaaS Confidentiality Risk Management (ScoRiM) Framework**, which improves the client side confidentiality management in a public SaaS and focuses on small to medium sized enterprises (SMEs), which are often confronted with rigid contracts enforced by CSP and cannot negotiate Service Level Agreements (SLAs). Figure 4-8 presents and outline of this framework.

It is not a technology solution, but rather a handful set of steps that businesses should follow to achieve a fitting policy. Basically, the objective of this framework is to help customers to determine which public SaaS provider best suits their confidentiality needs. These steps are, in order of execution: *Preparation Phase* (definition of IT goals, establish security policies and understand legal requirements), *Conception Phase* (establish confidentiality requirements over data), *Identification Phase* (identification of the current system, potential threats and vulnerabilities), *Confidentiality Risk Assessment* (check if current security controls fulfil confidentiality requirements), *Provider decision Phase* (decide the most appropriate Cloud Provider), *Treatment decision and Implementation Phase* (give treatment to high risk threats and estimate time and cost) and *Management Phase* (review that the designed treatments are followed)

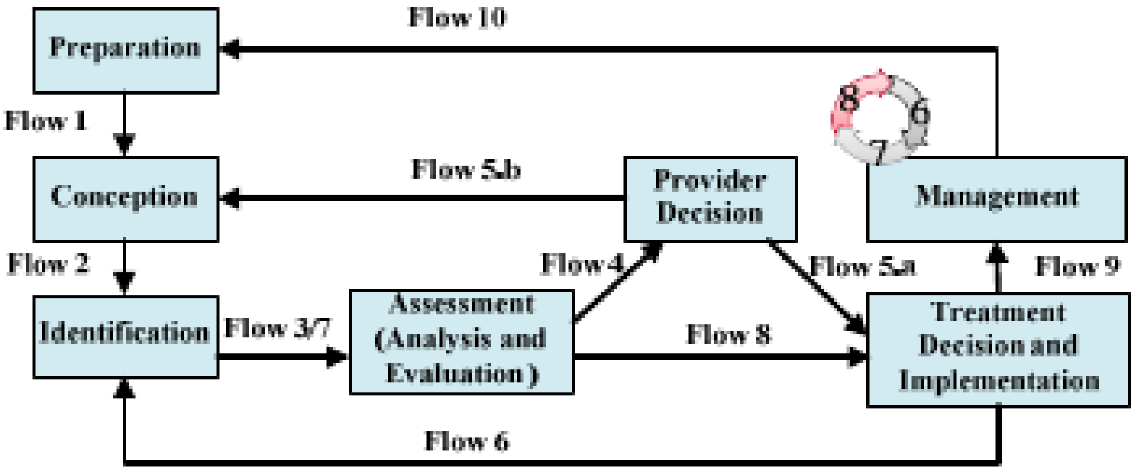

Figure 4-8: The outline of ScoRiM Framework

Itani et al. [SLR6] present a completely different approach compared to the previous ones presented to provide customers with a loose control over privacy: **PasS (Privacy as a Service)**, which basically is a set of security protocols for ensuring the privacy and legal compliance of customer data in cloud computing architectures. PaaS is based on the idea that privacy should be provided to cloud customers as a service with low additional cost, and that this privacy settings should be configurable, flexible, have control over the stored data, better protection of sensitive data and achieving legal compliance.

The goal of PasS is to maximize users' control in managing the several aspects related to the privacy of sensitive data, and achieving secure storage, processing, and auditing of customers' confidential data by taking advantage of the tamper-proof capabilities of cryptographic coprocessors, which provide a secure and trusted execution domain in the cloud protected from unauthorized access. This is achieved by implementing software protection and data privacy categorization mechanisms. The authors mention that they implemented a prototype of the PasS protocols on a simple banking application. The prototype included, in addition to a sample implementation of the privacy protocols, an implementation of the data privacy categorization and software division mechanisms.

4.3.2.1.3. Data ownership

When moving to the cloud, data storage and management changes its hands. This raises issues about who owns the data in the cloud, which must be clearly defined to avoid future litigations.

In the literature, 3 works have been identified which provide explanations about the issue of data ownership: [SLR16], [SLR20], and [SLR27]. These three papers provide suggestions on the matter. Figure 4-9 symbolizes the questions0 that the main Cloud actors arise on the matter, and table 4-8 summarizes the evidence found about the topic.

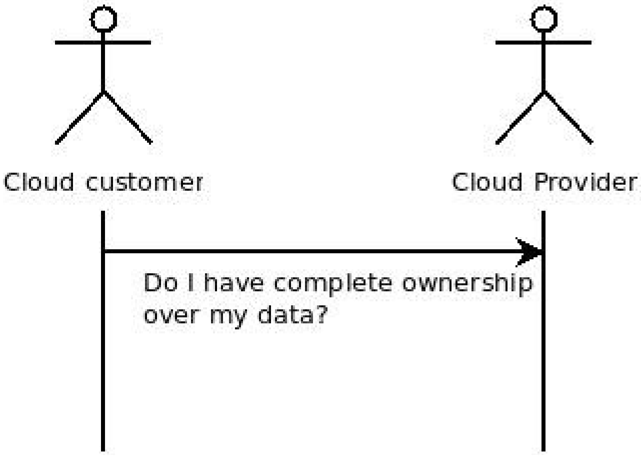

Figure 4-9: Questions raised by actors about data ownership

Issue	Proposed solutions / responses / suggestions
Who owns the data in the cloud	- Establish clear policies and SLAs [SLR16, SLR20, SLR27]

Table 4-8: Summary of identified issues related to data ownership

Mather et al. [SLR16] and NIST [SLR20] suggest, to deal with, that organisation's ownership over data has to be firmly established in the service contract and SLAs, clearly stating that the organisation retains exclusive ownership over the data and that the CSP does not acquire rights nor licenses over it.

Chen et al. [SLR27] exemplify data ownership issues with question that clearly defines the issues about data ownership: “If you move to a competing service provider, can you take your data with you? Could you lose access to documents if you fail to pay a bill?”. According to the authors, right now there is no way to control these issues besides assured SLAs or by keeping the cloud private. However, this kind of policies enter in contrast with what is explained in Section 4.3.5 – Economic Cost – where some cloud models may allow CSP to use data for several purposes like advertisement.

Given data ownership issues are loosely related with trust and policies, solutions to build up trust explained in Section 4.3.2.1.1 and solutions to define and accommodate policies explained in Section 4.3.2.1.2 may prove useful.

4.3.2.1.4. Changeability of terms

In Cloud Computing, terms of service may be the most important feature for an average user not subject to legal or professional obligation. When the data is stored on the cloud, it depends on a third-party company which may reserve the right to change/terminate established terms of service or privacy policies. A lack of completeness and transparency in the terms of use arises several concerns to the customers as, depending on the terms, privacy and confidentiality risks vary significantly [SLR18].

In the literature 3 papers have been found which explain or mention issues related to the changeability of terms: [SLR8], [SLR16] and [SLR18]. [SLR8] and [SLR18] provide actual solutions or suggestions on this concern, while [SLR16] provides review which helped identifying and assessing it. figure 4-10 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-9 summarizes all the concerns found and their respective solutions proposed by authors.

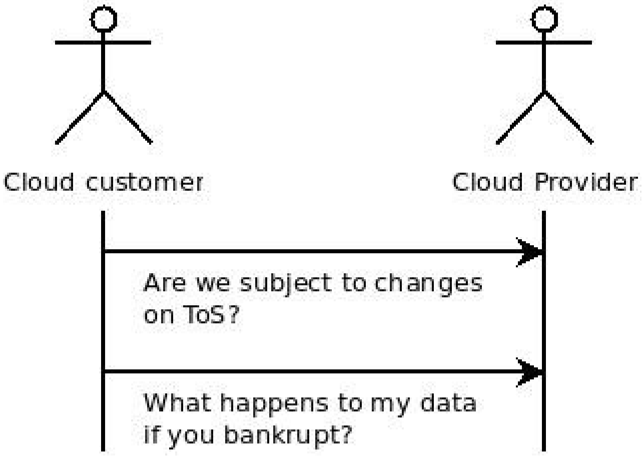
Cloud customer
Cloud Provider
Are we subject to changes on ToS?
What happens to my data if you bankrupt?

Figure 4-10: Questions raised by actors about changeability of terms

Issue	Proposed solutions / responses / suggestions
Bankruptcy of CSP	- Publish and maintain data interfaces and transformation logic [SLR8]
Change of terms and termination of services	- Perform adequate legal review of ToS [SLR18]

Table 4-9: Summary of identified issues related to changeability of terms

One possible way a CSP changes its services is, obviously, bankruptcy of the CSP. Mather et al. [SLR16] rise the concern about if the data is transferable to other third parties upon bankruptcy. In addition, Gellman [SLR18] explains that bankruptcy laws provides limited procedural protections for customers if the CSP had a privacy policy prohibiting the transfer of personal information to unaffiliated people (which, once more, suggest that customers carefully review privacy policies).

To deal with this issue, one solution proposed by Sengupta et al. [SLR8] is that, in order to prevent data getting locked in the case of a CSP bankruptcy, to publish and maintain a standard set of data interfaces and transformation logic. This way data can be transferred to other companies with ease by avoiding vendor lock-up.

Another way a CSP changes its services is if it reserves its right to do so any time. This reduces the comfortability of the customer, so Gellman [SLR18] suggests to perform an adequate legal review of the terms of service.

Finally, a radical change of terms explained by Gellman is the termination of the services from the CSP if it reserved the right to do so; a customer who did not keep a backup of his data stored in the cloud may lose the information permanently, with the subsequent troubles for customer's company.

To deal with the issue of changing terms, probably the solutions explained in Section 4.3.2.1.2 - Lack of execution controls and customers' privacy requirements and policies – may be of use as they deal with finding a proper Cloud Service Provider which fulfils customer's requirements. This way a more trustworthy relationship between customers and CSP can be established, providing more comfortability to cloud customers.

4.3.2.1.5. Responsibility for protecting privacy and privacy policy enforcement

The literature is fairly spread about the who is responsible about protecting data privacy and keep it confidential and protected. Conflicting opinions have been found in the literature about liability on privacy. So, in order to avoid disputes over responsibility issues, it needs to be a clear definition and understanding of roles and responsibilities between the customer and the provider.

In the literature, 6 papers have been found which deal with concerns related with who is responsible for ensuring data privacy: [SLR1], [SLR5], [SLR16], [SLR19], [SLR20] and [SLR24]. [SLR19], [SLR20] and [SLR24] provide actual solutions on the concern, while [SLR1], [SLR5] and [SLR16] provide reviews that helped identifying and assessing it. Figure 4-11 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-10 summarizes the evidence found on this topic as well as their related solutions.

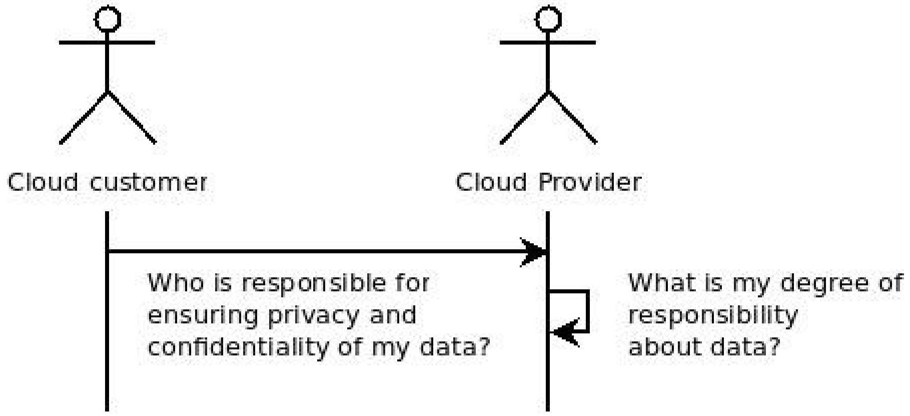
Cloud customer
Cloud Provider
Who is responsible for ensuring privacy and confidentiality of my data?
What is my degree of responsibility about data?

Figure 4-11: Questions raised by actors about responsibility for protecting privacy and privacy policy enforcement

Issue	Proposed solutions / responses / suggestions
Who is responsible for protecting privacy	- Shared responsibility between provider and customer [SLR24] - Responsibility depends on the cloud deployment model [SLR19], [SLR24] - Liability frameworks
Know if the CSP is fulfilling my requirements	- Deploy audit mechanisms to enforce policies [SLR20]

Table 4-10: Summary of identified issues related to the responsibility for protecting privacy and privacy policy enforcement

In their guidelines document, NIST [SLR20] explain that when outsourcing data, part of the responsibility and control is transferred to the CSP, which generates a dependency on the cooperation of the CSP to perform activities which imply both entities, which at the same time increases the complexity of carrying on activities like monitoring, incident response, compliance with data protection laws, requiring extra coordination efforts between organisation and CSP.

According to Mather et al. [SLR16], there are conflicting opinions regarding who is responsible for security and privacy; some authors delegate the responsibility to providers (without transferring accountability). Other authors like Roberts et al. [SLR1], Mishra et al. [SLR5], Takade et al. [SLR24] argue that from the public and the law point of view, the user organization is responsible of data security and privacy. In this sense, security and responsibility are shared between the client and the service provider and mutual cooperation is necessary. They even say that full reliance on CSP to protect personal data is irresponsible and can lead to negative consequences.

Nevertheless, Takade et al. [SLR24] and the ENISA report [SLR19] provide more in-depth information about the responsibility of privacy, adding the deployment model as a variable to determine the responsibility:

- in SaaS, providers have the most part of responsibility due that they provide the full kit of features. Customers and providers are responsible of compliance with data protection laws of customers who use SaaS services while CSP is ultimately responsible of providing and maintaining physical security and monitoring.

- in PaaS, as developers build their own applications, customers are responsible for protecting the data from their applications, whilst the CSP is responsible for ensuring isolation among other tenants.

- in IaaS, like in PaaS, customers are responsible for securing the OS and its content (monitoring, OS security patches...), whilst the CSP is responsible of keeping isolation between virtual machines and maintaining the physical infrastructure. (see section 4.3.1 – Virtualization). The definition of clear Service-Level Agreements also helps in setting-up priorities and warranties in Cloud Computing.

This categorization of responsibilities is related to the level of control depending on the delivery model, as explained in section 2.3, where as the amount of customer's control over infrastructure grows, so does its responsibility. Cloud customers need to pay attention on the roles and responsibilities between the organisation and the CSP to ensure that requirements are met. What NIST suggests in order to reduce risks is to deploy audit tools and mechanisms to determine how data is stored, protected and used, validate the service and verify that policy is enforced [SLR20].

Whatever responsibility assignment is done, ENISA [SLR19] recommends, when contracting cloud services, to carefully review terms of use and SLA clauses for clear role definition and responsibilities.

4.3.2.1.6. Audit, monitoring and accountability

Audit and monitoring deals with how organizations can monitor and control the activities of their Cloud Service Providers over their data. More specifically, the purpose is to watch what happened in the Cloud system in order to ensure that privacy requirements, SLAs and compliance with laws are enforced when their personal information is in the cloud [SLR10][SLR23][SLR24]. However, audit and monitoring raises some questions in regard of the matter of collecting usage data, i.e. profiling users to ensure high quality service. Customers are likely to not want sometimes that their actual content is monitored or give personal information against their will [SLR17][SLR26]. Even more, under some jurisdictions, Cloud Providers may be obliged to disclose and report information to authorities and governments (i.e. terrorism, missing children, etc.) [SLR18].

In the literature, 8 papers have been identified which deal with audit, monitoring and accountability issues: [SLR10], [SLR11], [SLR16], [SLR17], [SLR19], [SLR20], [SLR23] and [SLR26]. [SLR11], [SLR16], [SLR17], [SLR23] and [SLR26] provide solutions to the identified issues on the concern, while [SLR10], [SLR19] and [SLR20] provide reviews that helped identifying and assessing it. Figure 4-12 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-11 summarizes all the concerns found and their respective solutions proposed by authors.

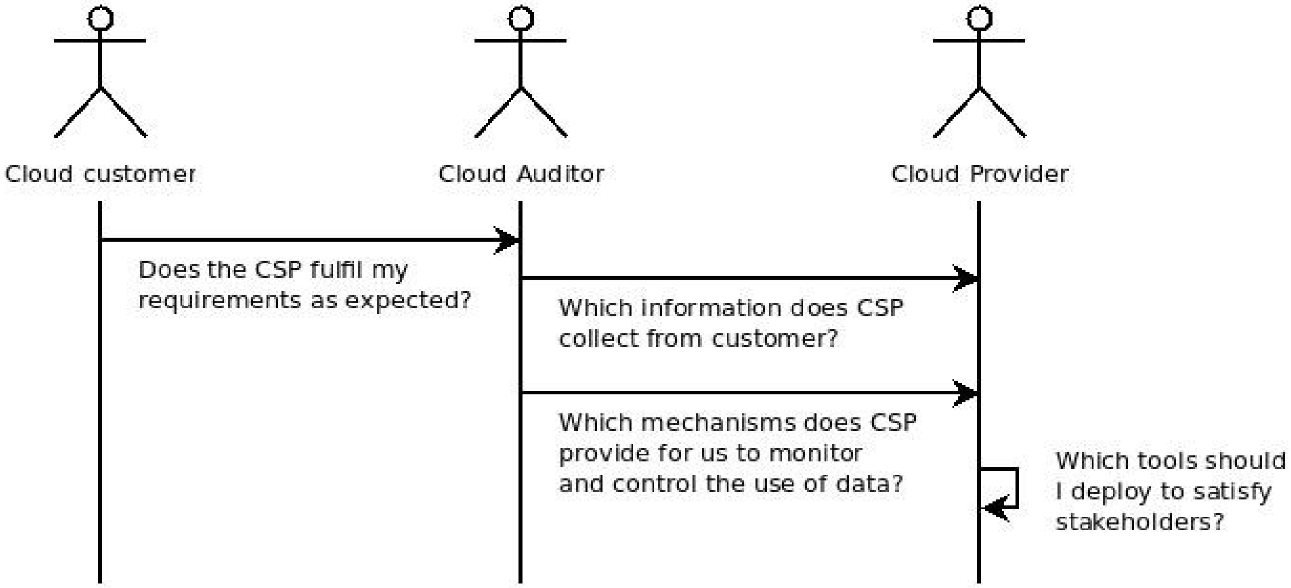
Cloud customer
Cloud Auditor
Cloud Provider
Does the CSP fulfil my requirements as expected?
Which information does CSP collect from customer?
Which mechanisms does CSP provide for us to monitor and control the use of data?
Which tools should I deploy to satisfy stakeholders?

Figure 4-12: Questions raised by actors about responsibility for protecting privacy and privacy policy enforcement

Issue	Proposed solutions / responses / suggestions
When does information gathering become intrusive	- Terms and conditions. Balance between collection of information and safety of data [SLR11][SLR26]
How can I be sure that my CSP fulfils my requirements?	- Accountability [SLR16][SLR17][SLR23]

Table 4-11: Summary of identified issues related to audit, monitoring and accountability

Inadequate monitoring mechanisms can lead to issues and complicate the assignment of responsibilities. In their guidelines, NIST [SLR20] suggests organisations to have knowledge of cloud provider's security measures in order to be able to conduct risk management by identifying vulnerabilities, analyse system security features and ensuring that privacy and security controls are implemented correctly in order to meet government security requirements. Related to this, the solution provided by Chou et al. [SLR9] in section 4.3.2.1.1 - Lack of execution and customers' privacy requirements, ScoRiM, may prove useful as one of the phases of their proposed framework includes risk assessment.

Another solution proposed by Pearson [SLR17][SLR23] and Mather et al. [SLR16] is accountability. Accountability is defined in the literature as the capability of identifying which party, with undeniable evidence, is responsible for specific events, and can be obtained via a combination of regulations, contracts and the use of privacy technologies like system architectures, access controls and machine readable policies attached to data in order to enforce the fulfilment of these policies to all stakeholders [SLR11][SLR16][SLR23].

In [SLR23], the authors explain how to implement accountability on the cloud and defend the idea that, with adequate accountability mechanisms, accountability can solve issues related to security breaches notifications (see section 4.3.2.1.9 – Privacy breaches notification), how data is managed in the cloud (see section 4.3.2.1.7. - Unauthorized use of data and access control), builds up trust on the CSP because processes become clearer, helps in making sure that the cloud service is compliant with laws and policies (see section 4.3.2.2.1 – Legal and regulatory issues) and, ultimately, location of data becomes a less relevant issue because it is assured that data is treated as described regardless of jurisdiction. In fact, accountability is included in some privacy frameworks, like APEC (Asia-Pacific Economic Cooperation, a framework that provide a set of best practices of privacy on Asia-Pacific economic areas [40]).

Another way to implement accountability is explained by Zhou et al. [SLR10], which consists on adding auditability as an additional layer above the virtualized operating system hosted in a Virtual

Machine, arguing that it is much more secure to implement auditability over a VM than an application or software. The reason is that the system then is able to watch the entire access duration when implementing auditability over the VM.

Despite the benefits of accountability, Xiao et al. [SLR11] argue that privacy and accountability may conflict with each other. In their paper, the authors explain some of the vulnerabilities and corresponding solutions to accountability. In regards of privacy, one of these issues is that the enforcement of accountability will violate privacy in some degree due to the data collection, and extreme privacy protection (i.e. full anonymity) will make accountability more difficult and harder to apply. Some accountability operations like logging or tracing cannot be performed without disclosing some private information (like IP addresses). The authors advocates for a balance in order to offer the best efficient service. Other issues relate to cloud integrity and availability, which are out of the scope of this Master Thesis.

In symphony with Xiao et al, Jaeger et al. [SLR26] mention that many companies provide contextualized ads based on keywords, Web sites viewed (like Amazon ads), etc. The authors ask themselves if users really want this to happen, as corporate users may be more concerned about monitoring of information than regular customers who just use e-mail solutions like Gmail. Sometimes the barrier that separates gathering information to increase service quality and data mining is not clear enough.

Another issue explained by the ENISA report [SLR19] regards compliance and certifications. When migrating to the cloud, some companies may have made investments in achieving certifications for meeting standards, competitive advantage or compliance with laws and regulations. However, if the Cloud Provider does not provide audit tools to customers, or provider cannot deliver evidence of their compliance to requirements, companies' certification and compliance requirements can be affected as well. Related to this, Mather et al. [SLR16] suggest Cloud Service Providers to apply **compliance frameworks and standards** to manage risks and satisfy customers while providing higher degree of trust. Some of the frameworks and certifications mentioned are ITIL (Information Technology Infrastructure Library, which basically is a set of best practices), NIST guidelines or ISO 27001 [41] and, as mentioned previously, APEC.

4.3.2.1.7. Unauthorized use of data and access control

Another issue found in the literature is the unwanted usage of the data from the CSP, how to make sure this doesn't happen and restrict/limit the access to personal information.

In the literature, 8 papers have been identified which deal with the issue of unauthorized use of data: [SLR7], [SLR16], [SLR18], [SLR19], [SLR20], [SLR22], [SLR24] and [SLR27]. [SLR7], [SLR16], [SLR19], [SLR22], [SLR24] and [SLR27] provide actual solutions on the concern, while [SLR18] and [SLR20] provide reviews that helped identifying and assessing it. Figure 4-13 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-12 summarizes all the concerns found and their respective solutions proposed by authors.

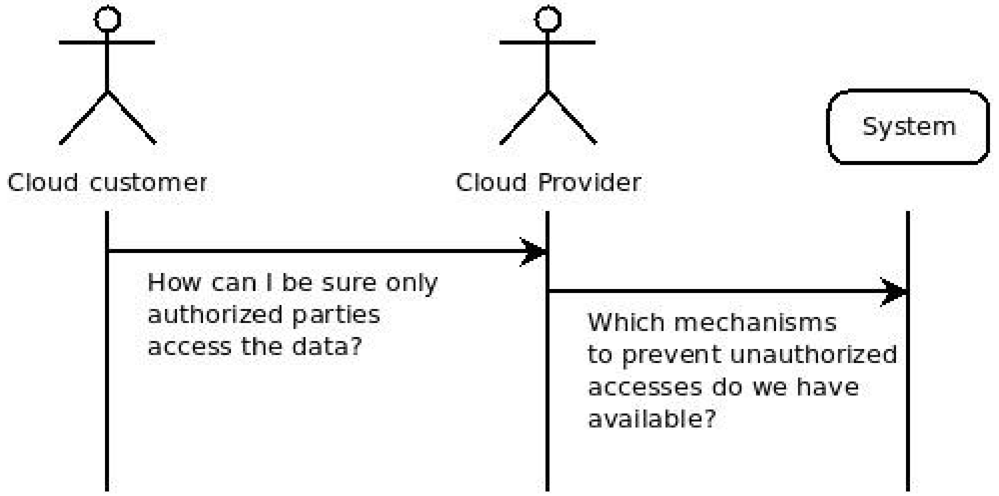
Cloud customer
Cloud Provider
System
How can I be sure only authorized parties access the data?
Which mechanisms to prevent unauthorized accesses do we have available?

Figure 4-13: Questions raised by actors about unauthorized use of data and access control

Issue	Proposed solutions / responses / suggestions
Prevent use of my data in ways I have not authorized	- Trust in the CSP [SLR24] - User-centric trust-model privacy management tool [SLR27]
Prevent access to unauthorized parties	- Identity and Access Management Protocols (IAM) [SLR7] [SLR8][SLR16][SLR20] - User-centric Identity Management (IDM) [SLR22][SLR24]

Table 4-12: Summary of identified issues related to unauthorized use of data and access control

One issue found in the literature is how to ensure that no malicious CSP employees access the data. Concerns could arise from secondary use of information by CSP, like the selling of personal information (i.e. a private photography of a user stored in the cloud may be sold for marketing purposes) [SLR18]. Takade et al. [SLR24] explain that it is improbable that there is any full-technical mean to prevent CSP to use the data in ways that have not been agreed with the customer, so combination of technical and non-technical approaches are necessary. The authors regard trust in the CSP as the most important attribute to guarantee data safety. CSP can build up trust using one of the ways explained in section 4.3.2.1.6 - Audit, monitoring and accountability - as accountability and auditability solutions can help enforcing a correct usage of the data.

Even though Takade et al. [SLR24] suggest that there is no way to prevent malicious use of data from the CSP, Chen et al. [SLR27] explain and propose a user-centric, trust model client-based privacy management tool developed by Mowbray [42], which can help users control the storage and use of their sensitive data in the cloud. However, no more information has been found in the literature for this SLR about this approach.

Regarding issues about who access data and how, one approach discussed and proposed by Almulla and Yeun [SLR7], NIST [SLR20], Sengupta et al. [SLR8] and Mather et al. [SLR16] to prevent unauthorized use and access of data are Identity and Access Management (IAM) protocols. More specifically, in their paper explain Security assertion Markup Language (SaML), and Open Authentication (OAuth) protocol. SaML is an authentication protocol based on XML standards and it is used to exchange authorization and authentication information between the Identity provider and Cloud Service Provider. OAuth is an authentication protocol which allows users to share their private data located on one CSP with another CSP without exposing the personal identity information like usernames and passwords.

However, before applying the explained IAM protocols, it is mandatory that the organisations checks if they are suitable for it [SLR20]. ENISA [SLR19] suggests defining good key management procedures, as the loss or corruption of keys for digital signatures and file encryption can hinder personal data. Another suggestion made by Mather et al. [SLR16] to deal with decentralised and inconsistent IAM application consists in finding technology solutions to enable centralized and automated user access management. However this kind of implementations can last years and incur considerable costs.

Another solution to prevent unauthorized use of data or security breaches proposed by Takade et al. [SLR24] and A. Cavoukian [SLR22] is User-centric Identity Management (IDM). IDM is an approach for handling private identity attributes, whose goal is to allow users control their digital identities, determine what information will be revealed to who and for what purpose. With User-centric IDM part of personal data management is relegated from companies and the CSP. One example explained by Cavoukian [SLR22] is the use of identity services (he mentions OpenID as an example), which can provide greater control over personal information. These identity services allow customers to store their credentials on a service and reuse them all around the Web without having to manage multiple accounts and passwords.

However, and according to Cavoukian, User-centric IDM requires a clear framework of agreed-upon rules between the customer and the CSP which includes policies that describe what information is requested and why, and include a machine-readable policy that travels with the personal information to ensure that this data is only used in accordance to this policy. Some suggestions made by the author in order to exploit the full potential of this kind of personal information management in Cloud Computing are:

- Identity management service companies require to build up trust in order to incite companies use them and take advantage of them.

- Standards organisations and governments should develop and promote standards for identity systems and privacy-enhancing technologies, and software developers should make use of them.

Finally, it seems reasonable that the proposals in Section 4.3.3 – Encryption could deal with the problem of unauthorized use of data as both issues pursue the same goal of protecting data confidentiality from third-parties. Also, accountability (section 4.3.2.1.5 – Audit, monitoring and accountability) could serve as a mean to track how is data managed and who access what data.

4.3.2.1.8. Multiple service composition

Sometimes, service providers cooperate with each other to provide bigger application services to customers. However, this raises several issues which are discussed below.

In the literature, 1 paper has been found related to the issue of multiple service composition: [SLR24], which identifies the concern and provides a solution to it. Figure 4-14 symbolizes the questions that the main Cloud actors arise on the matter, and Table 4-13 summarizes the found concerns about the matter.

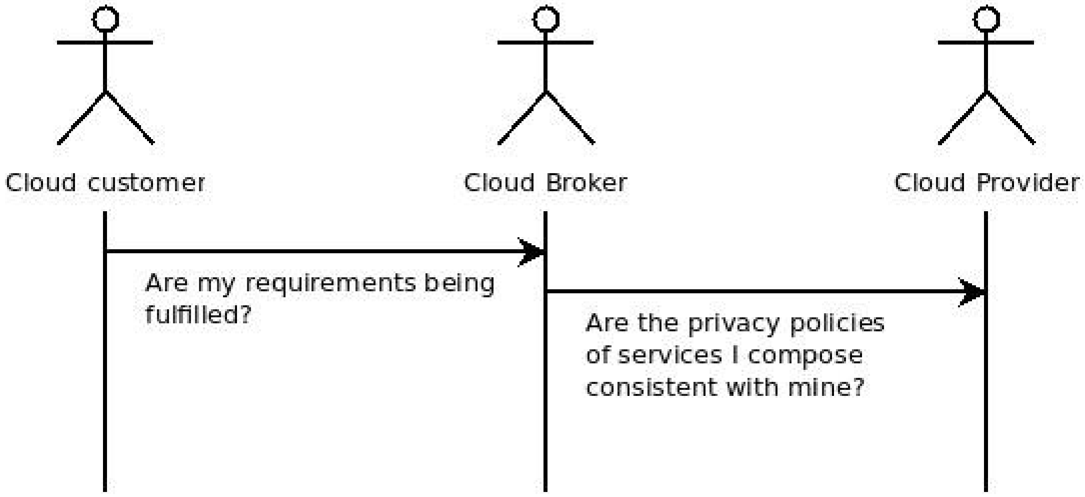
Cloud customer
Cloud Broker
Cloud Provider
Are my requirements being fulfilled?
Are the privacy policies of services I compose consistent with mine?

Figure 4-14: Questions raised by actors about multiple service composition

Issue	Proposed solutions / responses / suggestions
Ensure correct policy integration among CSP	- Trust-based interoperation frameworks [SLR24]

Table 4-13: Summary of identified issues related to multiple service composition

In their review paper, Takade et al. [SLR24] explain that different service providers may have different security and privacy mechanisms, so it is necessary to ensure that these collaborations are monitored effectively because, even though individual policies are verified, security violations can occur during the policy integration. One suggestion the authors make to solve this issue is the development of trust-based interoperation frameworks to capture the parameters required to manage interaction requirements and policy-evolution management.

The policy integration matter is related with the proposal made by Lin et al. [SLR3], which may help solving the issue explained above. Research should be done to validate this. See section 4.3.2.1.1. – Lack of execution controls and customers' privacy requirements – for more information.

4.3.2.1.9. Privacy breaches notification

Privacy breaches notification deals with how organizations are notified when a data leakage occurs (which may inevitably happen sometimes).

In the literature 1 work has been identified which deals with privacy breaches notification issues: [SLR16], which provides review that helps identifying the issue although no solution has been proposed. Figure 4-15 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-14 summarizes the found concerns about privacy breaches notifications.

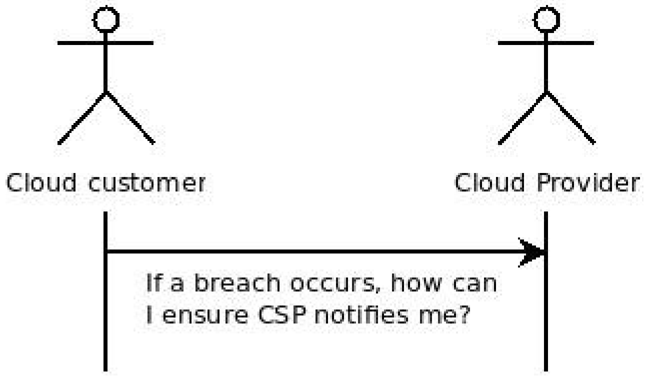

Figure 4-15: Questions raised by actors about privacy breaches notification

Issue	Proposed solutions / responses / suggestions
How to ensure that CSP notifies you when a breach occurs	- No solutions found in the literature
Who is responsible for managing the breach notification process	- No solutions found in the literature
If the contract policy include liability	- No solutions found in the literature

Table 4-14: Summary of identified issues related to privacy breaches notification

Basically and regarding to this topic, Mather et al. [SLR16] poses the question that how a company can ensure that the CSP notifies it when a privacy breach occurs, who manages the notification process and how is the responsibility contract enforced in case of a CSP negligence to determine the fault. These questions have a relationship with the trust of the cloud provider, as a trusted CSP provider has more chances to fulfil the SLAs.

Although no direct evidence has been found in the literature to solve these issues, accountability and audit frameworks may help to detect privacy breaches, as it deals with the tracking of problems. See section 4.3.2.1.5 – Audit, monitoring and accountability – for more information.

4.3.2.2. Offshoring

Offshoring refers to the mobility of data across jurisdictions (your data crosses international boundaries). Offshoring of data increases risk factors and legal complexity, which poses a challenge for companies who serve customers from multiple jurisdictions [SLR10][SLR16] and directly affects and complicates compliance with the laws in Cloud Computing architectures, as various types of security and privacy laws and regulations exist within different countries. One definition of compliance provided by NIST [SLR20] is *"organization's responsibility to operate in agreement with established laws, regulations, standards, and specifications".* Mather et al [SLR16] adds to this definition also that compliance deals with who is responsible for maintaining it.

For the sake of simplicity, we will divide the issues related to Offshoring in two subsections, one focused on general issues, and the other one focused on legal and regulatory issues.

4.3.2.2.1. Offshoring: General issues

In the literature, 5 works have been identified which deal with general issues related to Data offshoring: [SLR10], [SLR14], [SLR16], [SLR18] and [SLR19]. These works provide identify the issues around this concern and provide solutions to them. Figure 4-16 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-15 summarizes the concerns found in the literature about Offshoring in cloud computing and their respective found solutions.

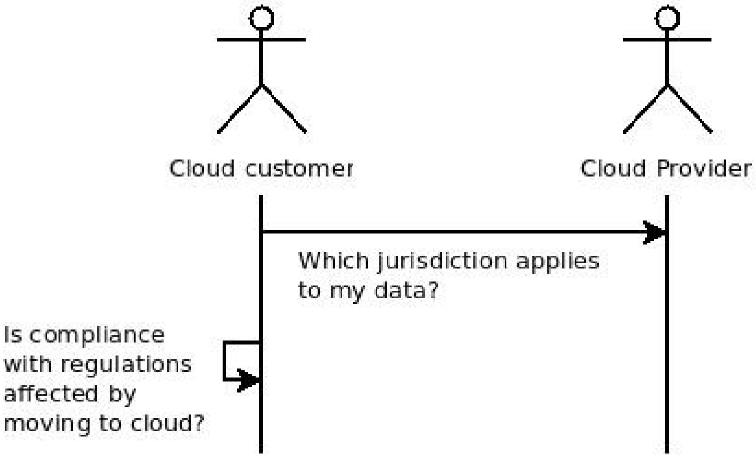
Cloud customer
Cloud Provider
Which jurisdiction applies to my data?
Is compliance with regulations affected by moving to cloud?

Figure 4-16: Questions raised by actors about general issues of offshoring

Issue	Proposed solutions / responses / suggestions
Which jurisdiction applies?	- Depends on the implicated countries and privacy laws: Applies the jurisdiction where data is physically located, or the jurisdiction of the subjects of data, or company's location [SLR10] [SLR18] - CSP can ensure to keep data inside a specific jurisdiction [SLR10] [SLR18] - Ask the CSP for any remarkable detail about data location and legal concerns [SLR19]
How is compliance impacted by moving to the cloud and where should I store my data	- Transfer principles [SLR14, SLR16] - Store data on specific jurisdictions to minimize legal risks [SLR14, SLR16]

Table 4-15: Summary of identified issues related to general issues of offshoring

In Cloud Computing, the location of the data may not be notified to the cloud customer due to the loss of governance over data and processes [SLR19]. One of the issues found in the literature and expressed by Mather et al. [SLR16] is which jurisdiction applies when information crosses borders and how it is determined. Special care has to be taken about the jurisdiction, as the applied jurisdiction to the data may, in some cases, provide unlimited access to the stored data (i.e. China legislation).

A CSP can promise through policy to maintain user data in a specific jurisdiction, reducing some of the location risks that a user may face. However this is not typically the case, and data may be split or mirrored in multiple jurisdictions in the cloud, making it unclear which laws apply to the data. This could lead companies to violate regulations without even noticing it [SLR19]. The point of view of Zhou et al. [SLR10] and Gellman [SLR18] is fairly different from Mather et al. [SLR16] (who say it is not sure the jurisdiction that applies). Zhou et al. and Gellman explain that jurisdiction is determined differently in different countries: some privacy laws are based on the location of the organization, some on the physical location of the data centre, and some on the location of the data subjects.

Another issue exposed by Mather et al. [SLR16] is how the impact on compliance with laws is determined when a customer decides to move his data to the cloud. Some countries may allow personal information to be processed without the awareness of the data subject, while in EU countries, for instance, only allow personal information to be processed if the data subject is aware of the processing and its purpose, and place special restrictions on the processing of sensitive data (for example, health or financial data), and explicit consent of the data owner is necessary [SLR14]. To prevent this kind of issue, Mather et al. propose two solutions:

- One solution is what are called the “Transfer principles”, under a set of privacy principles defined in their book. These principles specify that data should not be transferred to countries that do not provide the same level of privacy protection as the organisation who collected the information.

- The other solution consists that CSP should store data on servers located in specific jurisdictions that minimize legal risks. They empathize that these location may be outside Europe and United States in order to prevent the legal and sometimes contradictory point of view of privacy (when organisation deliver services around the world).

The ENISA report [SLR19] recommends customers to ask any legal questions to the cloud provider in order to ensure the location of data and infrastructure, the outsourced services and which jurisdiction applies.

Finally, Gellman [SLR18] makes a remark on data location and provides a different point of view from the other authors who try to deal with data location issues and ambiguity. Even though in the literature data location is perceived as an issue rather than an advantage, Gellman explains that uncertainty in location can be beneficial for someone who tries to keep data out of the reach of a government or litigant.

4.3.2.2.2. Legal and regulatory issues

As explained in the offshoring general issues, privacy laws vary according to jurisdiction. Currently, most legal issues in Cloud Computing are resolved during the evaluation of contracts, Service Level Agreements (SLA) and User Licensing Agreements (ULA). However, some privacy protection laws are either ambiguous, or out of date and did not keep in pace with technology.

In the literature, 3 works have been identified which deal with issues related to Legal and regulatory issues in Cloud Computing: [SLR10], [SLR16], and [SLR18]. These works identify the several concerns around this topic and provide solutions for it. Figure 4-17 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-16 summarizes the concerns found in the literature about the topic in Cloud Computing and their respective found suggestions.

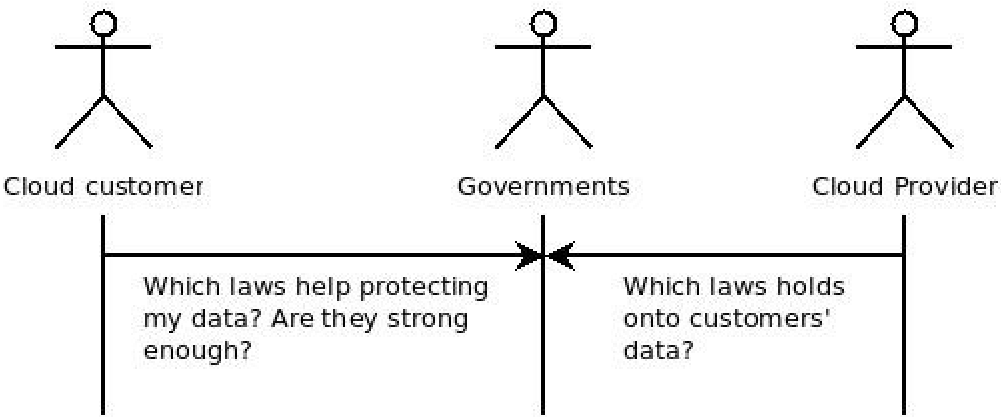
Cloud customer
Governments
Cloud Provider
Which laws help protecting my data? Are they strong enough?
Which laws holds onto customers' data?

Figure 4-17: Questions raised by actors about legal and regulatory issues

Issue	Proposed solutions / responses / suggestions
Different points of view and concepts about privacy. Conflicting regulations	- No solutions have been found in the literature
Outdated and inapplicable laws and regulations	- Apply and discuss changes to adapt them to Cloud Computing paradigm [SLR10][SLR16][SLR18]

Table 4-16: Summary of identified legal and regulatory issues

One issue Mather et al. [SLR16] describe in their book is the different understanding and points of view over privacy, which are the reason of multiple legal battles, political disputes and conflicting regulations. Some examples of conflicting regulations the authors explain are the U.S. Federal Rules of Civil Procedure (FRCP) and EU Directive. In Europe, privacy is considered like a basic human right and processing of personal data is forbidden, whilst in U.S. national security takes precedence over privacy (e.g. the USA Patriot Act, which is explained below, is the most controversial law and has arose several disputes among countries according to Mather et al.)

Zhou et al. [SLR10], Mather et al. [SLR16] and R. Gellman [SLR18] identify issues related to several acts that try to protect data privacy. They explain that, even though many laws have been published to protect users' privacy and businesses secrets, they are out of date and inapplicable to scenarios where more parties enter in action (like Cloud Computing). The authors suggest that these acts require discussion and changes to adapt to Cloud Computing environment.

The following list describes these problematic laws and the issues that have been found on them in the literature:

- Electronic Communications Privacy Act of 1986 (ECPA). ECPA are a set of U.S. laws that provide protections against government to access electronic information stored in devices of third parties (e.g. Internet Service Providers (ISP)), including mail and other computer information. However, Xiao et al [SLR11]. and R. Gellman [SLR18] explain that these regulations are outdated because ECPA relies on a model of e-mail and Internet that is generations behind current technology and practices and they protect privacy between only two parties (in cloud environment more parties besides Cloud Service Provide and Cloud user come into play, like governments), making it difficult to figure out what of those ECPA protections apply to Cloud Computing and when.

- Legally Privileged Information: Some laws establish regulations on the data privacy in the relationship of two or more parties, e.g. doctor-patient, lawyer-client, etc. Gellman [SLR18] argues that when information is legally privileged, sharing that information with a CSP can affect the validity of the privilege. For instance, if a patient shares a record containing a confidential communication with a cloud provider and the cloud provider uses the information in that record to serve an advertisement to the patient, the privilege may be utterly undermined.

- USA PATRIOT Act (UPA). This act allows FBI to access any business record under court order. As Cloud providers are businesses, it allows FBI to access Cloud user's private data as well.

- Health Insurance Portability and Accountability Act (HIPAA): HIPAA regulates the use and disclosure of protected health information by health care providers. However, according to

Mather et al. [SLR16], it does not currently regulate the third-party providers of health care businesses, so a legal demand by a private party to a CSP for disclosure of protected health information would lead users' private information to be disclosed.

- Gramm Leach Bliley Act (GLBA): GLBA restricts financial institutions from disclosing a consumer's personal financial information to a non-affiliated third party. However, disclosure to a CSP is not restricted, thus exposing cloud users' financial records.

4.3.2.3. Data combination and commingling

Data combination deals with the separation of customer's stored private data in the cloud with that of other tenants.

In the literature, 1 paper has been identified which mentions issues about data combination and commingling: [SLR10], which identifies the concern. Figure 4-18 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-17 summarizes the found concerns about data combination and commingling.

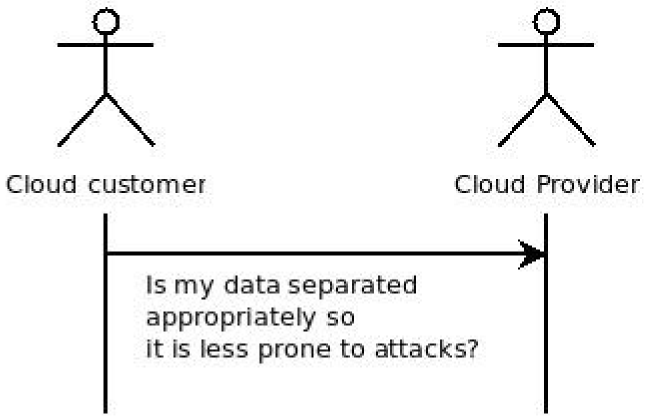
Cloud customer
Cloud Provider
Is my data separated appropriately so it is less prone to attacks?

Figure 4-18: Questions raised by actors about data combination and commingling

Issue	Proposed solutions / responses
Is my data separated from other tenants	- No solutions found in the literature

Table 4-17: Summary of identified issues related to data combination and commingling

According to Zhou et al. [SLR10], cloud customers need to be sure whether its private data is being stored separately from others or not. If they are combined with those of other tenants' data (like the same database), then this data becomes much more unsafe as it becomes more prone to attacks or virus transmission, as one attack to one victim may affect the availability or data integrity of other companies located in the same multi-tenant environment.

Although no direct evidence has been found in the literature, given that this concern is somewhat related to virtualization issues, probably this opens a gate that concerns and solutions provided in Section 4.3.1. – Virtualization – may be applied as well to data combination and commingling. More research in this area should be performed.

4.3.3. Encryption

Encryption, as explained in section 2.5.1, is the process of encoding messages or information in a way that that only authorized parties can read this information, preventing unwanted parties and hackers read it [43]. To protect the confidentiality of sensitive data stored in the cloud, encryption is the widely accepted technique in Cloud Computing architectures. In fact, encryption is the only recognized standard for data protection, like the NIST Federal Information Processing Standards (FIPS) [SLR16]. Nevertheless, encryption is not all-purpose as it alone cannot provide complete solutions to all privacy issues in cloud computing, and it complicates query processing on the data.

Most of the evidence found in the literature agrees that the biggest problem regarding encrypted data in the cloud is data access efficiency while preserving confidentiality.

In the literature, 14 papers which deal with encryption issues and proposals have been found: [SLR2], [SLR4], [SLR5], [SLR7], [SLR9], [SLR10], [SLR11], [SLR14], [SLR18], [SLR19], [SLR20], [SLR24], [SLR25] and [SLR27]. [SLR2], [SLR4], [SLR5], [SLR7], [SLR9], [SLR11], [SLR14], [SLR18], [SLR20], [SLR24] and [SLR25] provide actual solutions to the identified issues on this concern, while [SLR10], [SLR19] and [SLR27] provide reviews that helped identifying and assessing it. Figure 4-19 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-18 summarizes the issues found and solutions proposed about the topic in the literature.

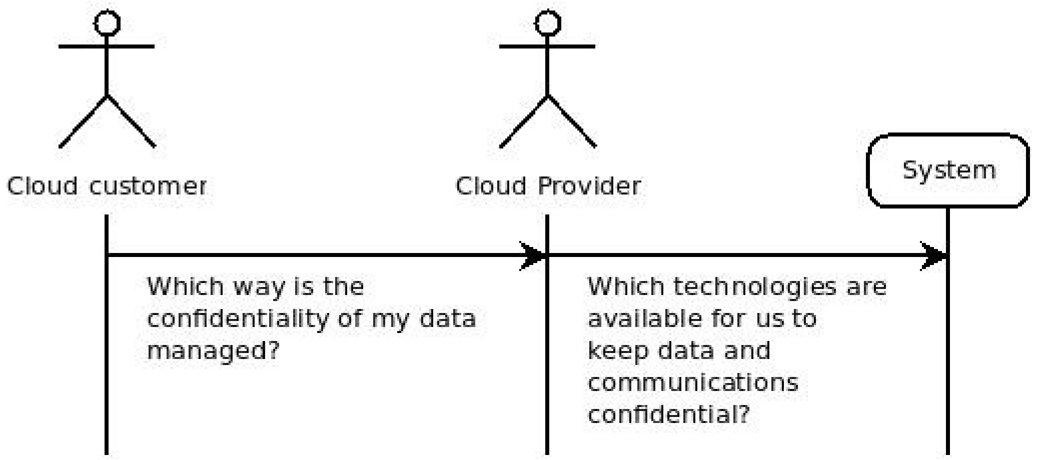
Cloud customer
Cloud Provider
System
Which way is the confidentiality of my data managed?
Which technologies are available for us to keep data and communications confidential?

Figure 4-19: Questions raised by actors about encryption

Issue	Proposed solutions / responses
Data retrieval efficiency while keeping confidentiality	- Homomorphic encryption [SLR2][SLR9] - Private Information Retrieval (PIR) [SLR2][SLR4] - Obfuscation [SLR4][SLR14][SLR25] - Secure co-processors [SLR2] - Secure index based framework [SLR2] - Trusted hardware token with Secure Function Evaluation [SLR11] - Privacy-preserving repository [SLR5]
Confidentiality protection of data	- NIST encryption standards [SLR7][SLR20] - Data-centric security approach [SLR24]
Secure communications and prevent vulnerabilities	- Anonymous communication through Onion Routing [SLR18]

Table 4-18: Summary of identified issues related to encryption

One important issue mentioned in the literature is the process of encrypted data. According to Chen et al. [SLR27] in traditional computing the data being treated is almost not encrypted for any program to use. However, in the cloud, due to the multi-tenancy feature, the data being processed by cloud-based applications is stored together with the data of other users, and unencrypted data in the process is a serious threat to data security. So solutions to process data while in encrypted state become necessary.

Agrawal et al. [SLR2] and Pearson et al. [SLR4] review several techniques for supporting the process of encrypted data. Regarding confidentiality protection, they mention a solution called **Private Information Retrieval (PIR)**, which can hide any query done on encrypted data. However, and according to the authors, PIR is too expensive in terms of communication and computation, solutions based on query processing over encrypted data are not practical yet, as they are not flexible (designed for a specific query), or they do not support another type of queries (like updates), or they trade confidentiality by functionality.

Chou et al. [SLR9] add to the previous statement that most of the current encryption solutions do not allow the processing of data in encrypted state. [SLR9], [SLR2], [SLR19] and [SLR27] mention as a possible existing solution the **Homomorphic encryption**, developed by IBM in 2009, which enables secure calculation on encrypted data; however homomorphic encryption comes with the drawback that it takes high computation cost and time to run operations over encrypted data.

In order to keep data confidential while having efficient data access, one solution Pearson et al. [SLR4], [SLR14], [SLR25] propose is a privacy manager that relies on obfuscation techniques to enhance confidentiality. The privacy manager can provide obfuscation and de-obfuscation service to reduce the amount of sensitive information stored in the cloud and reduce the risk for data leakage and loss of privacy. The main idea is to store the clients' data in a encrypted form in the cloud, and the data process is directly performed on the encrypted data. One limitation is that CSP may not be willing to implement additional services for privacy protection. So, without provider's cooperation, this scheme will not work.

Another solution proposed by Agrawal et al. [SLR2] for efficient data access while keeping confidentiality is the use of a secure co-processor on the cloud server side and put all sensitive data process inside that co-processor. However this comes with the drawback that every client in the CSP cloud should trust the co-processor. So, in the end, to solve these issues, the authors propose a **secure index based framework** to support efficient processing of multiple database queries and practical secure data management while preserving both confidentiality and functionality. This framework is based on not processing encrypted data directly like previous mentioned solutions, instead it uses an encrypted index which allows to locate and retrieve the data faster (in a small number of rounds) and **does not rely on the CSP**. This solution may solve the problem of the solution proposed by [SLR4] as their solution required the cooperation of CSP.

In their paper, Mishra and Dash [SLR5] propose a repository for preserving privacy of data in cloud, facilitating the integration and sharing of data across cloud while preserving data confidentiality, and delegates most computation intensive like encryption tasks to the server. The process consists as follows: when a user makes a query, this repository analyses user's integration requirements and constructs and decomposes the query plan for his query, discovers and fetches data from the cloud service, assimilates all data together, and returns the final results to users. All this process is managed under encryption. This solution is somewhat related to that of [SLR2] as they both deal with query processing, although in [SLR5] it is not clear whether their solution is practical in terms of computation cost or not as there is not any demonstration. The authors say that the benefits that this repository provides are: in the process, it can be known who is using the data and in what way (certain degree of accountability, see section 4.3.2.1.6 – Audit, monitoring and accountability), and the dispatched information is adequate to support clients' integration requirements, but carries no extra information of the data and that it is cost-effective and robust.

Another solution explained by Xiao et al. [SLR11] is a proposal made by Sadeghi et al. They propose to combine a trusted hardware token with Secure Function Evaluation (SFE) in order to compute arbitrary functions on data when it is still in encrypted form. The goal of this work is to minimize the computation latency to enable efficient and secure data outsourcing in cloud computing. The computation leaks no information and is verifiable.

Regarding the way data is stored in the cloud, Almulla and Yeun [SLR7] suggest that encryption techniques like symmetric or asymmetric encryption algorithms and key management of the symmetric cipher should be taken into consideration in order to ensure that data in the cloud cannot be accessed by unauthorized parties. Zhou et al. explain in [SLR10] that this kind of approach was successfully used by TC3, a healthcare company with access to sensitive patient records and healthcare claims, when moving their HIPAA-compliant application to Amazon Web Services. [SLR7] also suggest that NIST encryption standards [SLR20] should be implemented in order to protect confidentiality.

Another issue found in the literature explained by Takade et al. [SLR24] is that data owners should have full control over who has the right to use the data in the cloud and what they are allowed to do with it. To provide this data control, authors suggest the use of what they call "data-centric security approach", based on encryption and usage policy rules, so when someone tries to access the data, the system checks its policy rules and only provides the data if the policy is satisfied.

Regarding encrypted communications, the ENISA report [SLR19] makes a mention on several vulnerabilities which can affect personal data while being transferred: MITM (man-in-the-middle) attacks, poor authentication and acceptance of self-signed certificates. No more evidence has been found in the literature about these issues. On the other hand and related to communications, Gellman [SLR18] opens the gate for further research on this area with the envisioning of "onion cloud providers" which may make the pursue and finding of data very difficult. More specifically, he explains a technique for anonymous communication over the network called **Onion Routing**, which consists on repeatedly encrypt a message and send it through multiple network nodes called onion routers. Each onion router removes a layer of encryption to uncover routing instructions and sends the message to the next router. This process is repeated until the message reaches its destination.

4.3.4. Storage, retention and destruction of data

Retention deals with how long personal information transferred to the cloud is retained and which retention policy applies to this data; Destruction deals with how cloud providers destroy personal information when the retention period ends and how organizations ensure that their data is destroyed and is not available to other cloud users. In cloud computing data is outsourced and the lack of control on this data raises issues regarding the maintenance of this data.

In the literature 4 papers make references about retention and destruction of data: [SLR8], [SLR12], [SLR16] and [SLR20]. [SLR8] and [SLR20] provide actual solutions to the issues identified on this concern, while [SLR12] and [SLR16] provide reviews that helped identifying and assessing it. Figure 4-20 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-19 summarizes the issues found and solutions proposed about storage, retention and data destruction in the literature.

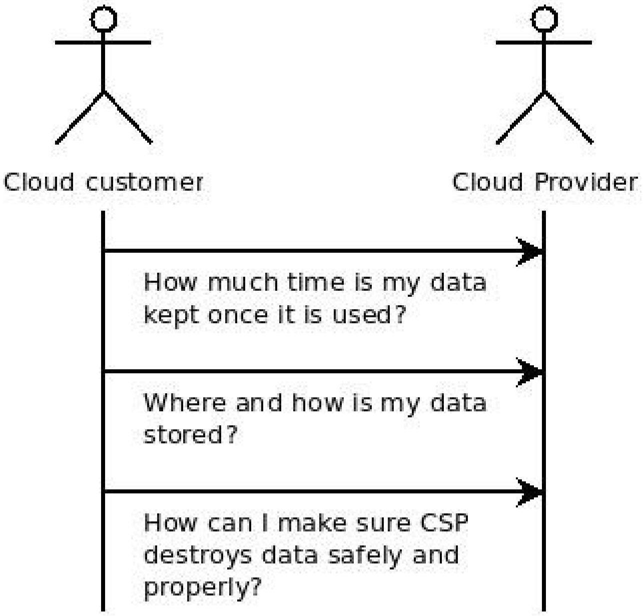
Cloud customer
Cloud Provider
How much time is my data kept once it is used?
Where and how is my data stored?
How can I make sure CSP destroys data safely and properly?

Figure 4-20: Questions raised by actors about storage, retention and destruction of data

Issue	Proposed solutions
Who enforces retention policy	- No solutions found for this concern in the literature
In which media is data stored	- No suggestions nor proposed solutions found in the literature
Prevent data concentration vulnerabilities and the stored data is correctly isolated	- No suggestions nor proposed solutions found in the literature
How much time is my data kept	- Review policies [SLR20]
How to make sure data is properly destroyed	- Encrypt unwanted data and destroy the key [SLR8] - Review SLAs and policies [SLR20]

Table 4-19: Summary of identified issues related to storage, retention and destruction of data

Mather et al. [SLR16] and ENISA report [SLR19] pose some of the issues that arise regarding retention and destruction of data: Who enforces the retention policy in the cloud, and how are exceptions to this policy managed? Is my unwanted data truly destroyed in the cloud?. These questions are closely related to data outsourcing (see section 4.3.2.1 – Data outsourcing) and responsibility of data (see section 4.3.2.1.5 - Responsibility for protecting privacy and privacy policy enforcement).

Related to retention and according to Ma [SLR12], the kind of storage used by the CSP (RAIDs, portable media, etc.) is an important factor, as there are risks associated depending on the device used, and that the archival of data fulfils legal requirements (usually it is more unsafe to put data on portable devices than RAIDs, for example). However this information is not typically disclosed to customers, which in some cases may be a problem.

In reference to the storage of information, NIST Guidelines [SLR20] explain that depending on the CSP deployed infrastructure, data concentration is an issue as there is more data concentrated in a single point, making it valuable attack targets, so isolation is necessary to minimize risks. In special, NIST places special concerns on which data should be located alongside high-importance data present in a device, as it is more likely that attackers may insist more over it.

Another issue regarding the destruction of data explained by Sengupta et al. [SLR8] is that some services (like Google Gears) cache data on their devices, and if this data is not secured and purged on a regular basis, the data is prone to attacks and pose a serious risk to privacy. On the same matter, Ma [SLR12] and NIST [SLR20] explain that when a file is deleted, the file name is removed from the directory, but the actual data still remains on the disk, and attackers still can use special techniques to get the deleted data back (so, data must be destroyed in a secure manner).

So, in order to be sure that data cannot be accessed any more, NIST [SLR20] suggests that SLAs should define the measures that are taken to ensure that data is completely destroyed. Another approach Sengupta et al. [SLR8] explain is to encrypt the data and then destroy the key so it is guaranteed that no one can access that information.

4.3.5. Economic cost

In the literature we have seen many technologies and solutions to solve privacy and confidentiality in the cloud issues and meet users' requirements. However, at the same time, with the increased amount of processed data, the cost for privacy protection also increases making some of those solutions not feasible for a commoditized environment like Cloud Computing [SLR15], which, in the end, becomes an issue.

In the literature, 5 works have been identified which deal or mention privacy costs on implementing Cloud Computing solutions: [SLR3], [SLR6], [SLR15], [SLR20] and [SLR23]. [SLR6], [SLR15], [SLR20] and [SLR23] provide suggestions for the different issues, while [SLR3] makes a review that helped identifying and assessing the concern. Figure 4-21 symbolizes the questions that the main Cloud actors arise on the matter, and table 4-20 summarizes the issues found and solutions proposed in the literature about economic costs on privacy and confidentiality solutions.

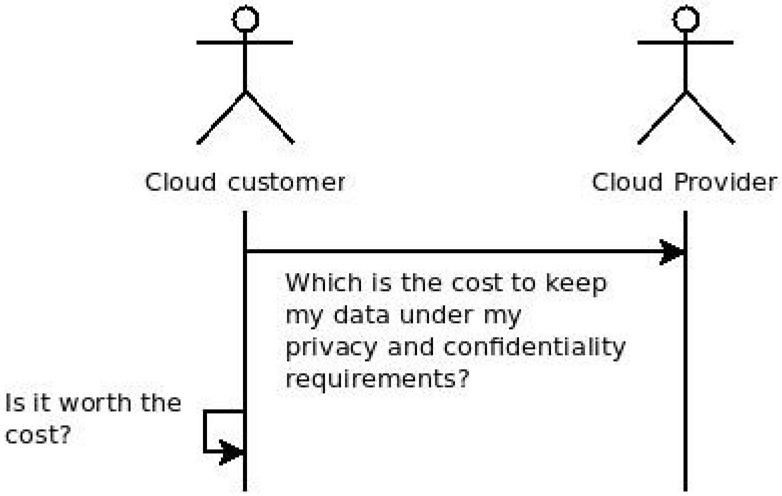
Cloud customer
Cloud Provider
Which is the cost to keep my data under my privacy and confidentiality requirements?
Is it worth the cost?

Figure 4-21: Questions raised by actors about privacy economic cost

Issue	Proposed solutions
Which cost models are available while preserving privacy	- Allow the usage of analytics over data and allow contextual advertisement [SLR20][SLR23] - Low fee, non-negotiable terms services [SLR20] - Fee-based negotiable terms services [SLR20]
How to keep privacy costs under control	- Privacy-aware inter-cloud data integration system [SLR15] - Privacy-as-a-Service (PaaS) [SLR6]

Table 4-20: Summary of identified issues related to privacy economic cost

One method explained by Pearson and Charlesworth [SLR23] for cost-saving data process consists on providing the CSP with permission to use analytics or even analyse data stored and demand advertisers, in order for the user to receive free services. They mention as example of this the Google's business model, centred on providing free service to users while using information obtained to benefit advertisers. However, organizations and individuals will not want their data used for information for advertisers, so this will likely require the users to choose services that include fees but offer higher confidentiality and do not sell data to advertisers.

Related to what Pearson and Charlesworth explained, NIST [SLR20] provides a classification of types of cloud services models depending on the cost with the aim of advising customers that, when moving to cloud, they should choose carefully the destination of their data:

- The first one are services provided with no cost to consumers, supported through advertisements (like mail). However this presents the issue that these services usually collect information from the usage and customer's personal information and encrypted communications may be unavailable.

- The second one provides services to customer's at a low fee, with similar capabilities as those of the first classification, and more safety and less private data collection. However, the terms of service in these cases are usually non-negotiable and CSP can modify them unilaterally (see section 4.3.2.1.4 - Changeability of terms -).

- The third one provides fee-based services with negotiable terms of service between the organization and the CSP. NIST explains that the costs generally depend on the deviation from non-negotiable-term services.

In their paper, Tian et al. [SLR15] explain some solutions from other authors like privacy preserving repositories to accept integration requirements from users (like the privacy repository explained in [SLR5]), help data sharing and return the data integration results to users, keeping the processing of data secure by randomizing the data before sending it to the repository, and using encryption algorithms. However, they did not consider the practicability of uploading all the unprocessed data to the repository and the high cost for transmitting the data to the cloud. Related to this issue, Lin et al. [SLR3] comment that privacy protection techniques should not add a significant communication overhead, and therefore the communication cost before and after using policy enforcement techniques should be compared.

Tian et al. suggest that privacy may be offered as a new service in the cloud environment, in which customers pay a different price for those privacy services with different protection assurances according to the importance of data. Based on this analysis, the authors present a privacy-aware inter-cloud data integration system, which considers the trade-off between the privacy requirements from users and the charging for those data protection and processing. This idea is very much related to the Privacy as a Service approach proposed by Itani et al. [SLR6] (see section 4.3.2.1.2 – Lack of execution controls, customers' privacy requirements and policies), where they both suggest providing privacy as a low cost service.

The aim of the system proposed by Tian et al. is to mediate between users' privacy preferences and the cost for privacy protection. A schema of this system can be seen in Figure 4-22. Summarizing how it works, users send their query to the repository cloud, and then, after the query plan executor processes query, given user's preferences (by setting risk values to decide the importance of data) and estimating the cost of that query in terms of encrypting, uploading and downloading of data, the query plan processor decides whether to execute that query or not. This way privacy is maintained based on users' requirements and cost.

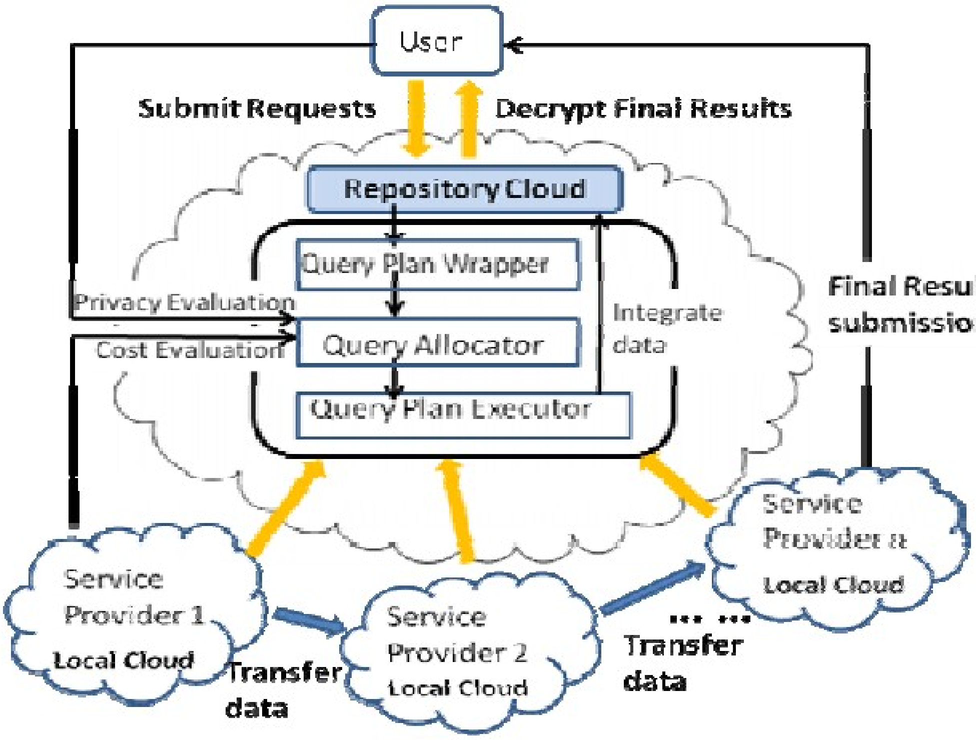

Figure 4-22: Privacy-aware inter-cloud data integration system architecture

CHAPTER 5

DISCUSSION

Chapter 5: Discussion

5.1. Conclusions from the Review

In Chapter 4 we have seen that privacy is probably the most common and important problem any organization faces when it decides to move to Cloud Computing, even more than data integrity or availability. Information is the currency of 21st century, and depending on the type of this information, it can be priceless. Many companies heavily doubt on moving their data to the cloud, as sometimes a simple leakage of information or a bad usage of it can lead to bankruptcy of the company.

However, the so-claimed cost-savings of leverage on Cloud Computing technologies and approaches heavily attract customers as explained in Chapter 2. However, as we have seen in the SLR, the amount of issues and concerns raised by privacy and confidentiality matters is not small at all. Depending on the size of the company who decides to move to the cloud and the kind of data they wish to move there for storage or processing, special measures and assessments should be made to minimize any possible risk and comply with laws and regulations. For this reason, industry and researchers put lots of efforts to address these issues and build-up trust in the cloud, so more companies can benefit from its economic benefits without worrying too much.

Based on the descriptions and definitions found in the literature on the different privacy and confidentiality issues, requirements and data life-cycle phases (see section 4.1 and section 4.2), we establish in table 5-1 a relationship between each privacy requirement phase of the data life-cycle and their associated phases and identified issues in the literature. The objective of doing this is raising discussion through finding similarities, differences and gaps, help us extracting more elaborated conclusions, assess whether requirements are being fulfilled and suggest research directions to any uncovered gap we may find.

Requirements	Phases	Related Issues
Notice, openness and transparency	- Generation of the information - Use - Transfer - Transforming and sharing - Storage - Archival - Destruction	- Lack of execution controls, Customers' privacy requirements and policies - Data ownership - Changeability of terms - Responsibility for protecting privacy and privacy policy enforcement - Privacy breaches notification
Choice, consent and control	- Generation of the information - Use - Archival	- Unauthorized use of data and access control - Multiple service composition
Scope / minimization	- User - Archival - Destruction	- Unauthorized use of data and access control - Storage, retention and destruction of data
Purpose	- User - Archival - Destruction	- Unauthorized use of data and access control - Storage, retention and destruction of data
Access and accuracy	- Storage - Transforming and sharing	- Unauthorized use of data and access control
Security safeguards	- Generation of the information - Use - Transfer - Transforming and sharing - Storage - Archival - Destruction	- Virtualization - Encryption
Compliance	- Generation of the information - Use - Transfer - Transforming and sharing - Storage - Archival - Destruction	- Offshoring - Legal and regulatory issues
Limiting use, disclosure and retention	- Transforming and sharing - Archival - Destruction	- Unauthorized use of data and access control - Audit, monitoring and accountability - Storage, retention and destruction of data
Accountability	- Generation of the information - Use - Transfer - Transforming and sharing - Storage - Archival - Destruction	- Audit, monitoring and accountability - Lack of execution controls, Customers' privacy requirements and policies

Table 5-1: Relationship among privacy requirements, data life-cycle phases and issues

Notice, openness and transparency

This requirement affects all phases of data management. This is in line with the importance that several authors give to the necessity of users to be aware what is done and what happens with their data at every moment and that their privacy requirements are met. Transparency is vital for generating trust, and providing customers with complete awareness of cloud details and functions definitely helps to achieve this goal.

Customers should carefully review privacy policies and SLAs to make sure that no privacy gaps are uncovered, and ensure which information is collected from the vendor. Guidelines, frameworks and suggestions described in section 4.3.2.1.1 – Data Outsourcing: General issues – like the Privacy Impact Assessment (PIA) or APEC, and the custom privacy policy generation and negotiation like those in section 4.3.2.1.2 – Lack of execution controls, customers' privacy requirements and policies – should help customers identify and select cloud providers whose policies ensure transparency.

However, the lack of standards and compliance issues require customers to put extra efforts on assessing privacy risks and compare different cloud vendors. The proposed solutions may ease this task but, as suggested by Gellman [SLR18] in section 4.3.2.1.2, further research is necessary on this area. This way, both customers and cloud providers could benefit from more transparency and safety.

Note that without transparency and event notification, other requirements may be affected or violated. Even though thanks to Cloud Computing customers are abstracted of implementation details and complexity, hiding certain aspects like which cloud providers participate in their architecture and service composition (see section 4.3.2.1.8 - Multiple service composition) or the location of data can raise serious problems. Accountability could prove useful to make sure that everything works the way it is expected, and Policy integration frameworks like the one proposed by [SLR3] could help in making sure that in the service chain policies are not violated at any step. (see section 4.3.2.1.1. – Lack of execution controls and customers' privacy requirements – for more information).

Compliance

Legislation is an entity that should be satisfied at every moment, so it is logic that this requirement affects every phase of the data life-cycle. However, as it can be noted by reading section 4.3.2.2 – Offshoring – legislation issues basically raise ethical and cross-jurisdictional matter which may conflict with companies' and customers' interests and rights.

As explained by several authors, revision of several privacy laws is necessary to cover the privacy gaps that generate these conflicts.

Accountability

This requirement affects all phases of data life-cycle as every data access or modification should be tracked. As we could see in section 4.3.2.1.6 – Audit, monitoring and accountability -, accountability is shown as a solution to provide more control and awareness to customers, favouring transparency, solving the related issues. As we have explained in other requirements, accountability is a requirement that greatly benefits the fulfilment of basic privacy principles.

CSP should allow customers audit their services and provide them with monitoring tools. However, accountability requires collecting data, so, as explained in the literature in section 4.3.2.1.6, users need to be aware which data is collected. Once again, research and development of standards and certifications can ease the finding, comparison and assessment of cloud vendors.

Security safeguards

This requirement affects all phases of data life-cycle, as data should be, according to the description of the requirement, always protected regardless of the phase.

Across the SLR we have seen several technologies and approaches that aim at covering several vulnerabilities and weaknesses that threaten data confidentiality. Some reviewed technological proposals are those described in section 4.3.1 – Virtualization – to protect against virtualization vulnerabilities and ensure isolation, section 4.3.2.1.7 – Unauthorized use of data and access control – like IAM to ensure that data is only accessed by authorized people, section 4.3.3 – Encryption – to protect data privacy and confidentiality through all stages of the data, from generation to transmission, storage and processing, and section 4.3.4 – Storage, retention and destruction of data – to ensure that deleted data is not accessed any more by someone.

Choice, consent and control

This requirement affects the generation of information, use, and archival phases of the data life-cycle. Customers should have the freedom to choose the information they want to publish (generation of information), decide with who they share that information and how is it managed (use), and decide how long this information is available (archival).

In the literature, we have seen in section 4.3.2.1.7 – Unauthorized use of data and access control – that it is rather complex to make sure that data is safe all of the time. A malicious intent from a cloud provider (e.g. a rogue employee) may harm data and remain unnoticed by taking advantage of a possible vulnerability. As proposed by Cavoukian [SLR22], privacy-enhanced technologies coupled with accountability mechanisms can help preventing or minimizing damaging effects.

However, we do not think that these measures are enough. As suggested by Takade et al. [SLR24] in the same section, trust and following best practices are an important factor to guarantee usage of data in an agreed way with the customers. As an additional to boost trustiness, a possible suggestion could be **the development and composition of compensatory schemas in policies and SLAs** (either technologically by for instance solving the issue under a certain amount of time, or economic compensation).

Regarding the archival of information, no solutions or suggestions have been found in the literature for issues related about who enforces retention policies, and how to prevent the concentration of data from multiple customers on a single machine and how to assess the best isolation.

About who enforces retention policies, throughout the SLR we have seen about the importance of defining clear policies and SLAs. Retention policies are not an exception, and as no direct evidence or solution has been found in the literature, we suggest customers to carefully review SLAs and keep well defined requirements.

About data concentration, combination and commingling, Cloud Computing consists on a multi-tenant environment. It is obvious that if your data is stored alongside the data of a big company that contains high valuable information, the machine where that data is stored becomes much more attractive to attackers and hence, the chance your information is affected may rise as well.

Further research needs to be done over this issue. One possible line of research could be **extending data isolation algorithms** like those used in Virtualization technologies (i.e.

machine assignation algorithms) to provide a certain weight to data depending on its importance and define a “maximum weight” that every machine can support (it is assumed that the infrastructure works in a distributed network) in order to divide value concentration. This way, if a machine suffers an attack, the harm per customer would lower. We believe that, in case of privacy breach, risking a fraction of personal data of ten customers is better than exposing a big chunk of personal data from a single customer. Once again, this is just a simple proposal with the objective of pointing a direction for researchers and open discussion.

Purpose

This requirements affects the use, archival and destruction phases of the data life-cycle. According to the description of the requirement, purpose is related to the scope of usage of data (use) and storing (archival), and if data is no longer necessary and retention time has passed, it should be destroyed to prevent unnecessary information breaches.

As we have seen in section 4.3.2.1.7 – Unauthorized use of data and access control – and explained in the *choice, consent and control requirement*, trust, policies, standards and certification development are regarded as the best approaches to be confident that data is used accordingly.

Scope / minimisation

This requirements affects the use, archival and destruction phases of the data life-cycle. Reasons for this assignment are trivial, as this requirement needs to keep consistency with the Purpose requirement. If the usage, sharing and storage of data has to be limited for the purpose it was collected, then the data collected should only be the necessary one to fulfil that purpose. Otherwise extra collected unnecessary data may raise alerts on the customer and affect trust.

Access and accuracy

This requirement affects the storage and transforming and sharing. The reason is that storage requires proper technologies to ensure that data is accessed only by authorized parties.

Through the review we have seen several technologies like the Identity and Access Management (IAM) and User-centric Identity Management (IDM) which allow users to handle their own personal information and determine who has the right to access it.

Even though we have not found direct evidence, accountability could serve as a mean to track how is data managed and who access what data, and encryption algorithms help keeping communication and transmission of credentials safe. More research should be done in this area to confirm this.

Limiting use, disclosure and retention

This requirements affects the transforming and sharing, archival and destruction phases of the data life-cycle.

As explained in accountability, and access and accuracy requirements, accountability and Identity Management protocols can be a way forward ensuring that data usage is limited for use to specific parties defined by customers.

Regarding retention, no direct evidence in the literature has been found to ensure that data is kept no longer than requested. However, once again, approaches like accountability, and reviewing privacy policies to ensure that they are aligned with requirements could prove good solutions.

At this point, we have analysed and summarized the results of the Systematic Literature Review performed in Chapter 4. After this discussion, we are in shape to provide the answers to the review questions we posed for our SLR in Chapter 3.

5.2. The review questions: answers

This section will propose an answer to the questions posed for the Systematic Literature Review (see section 3.1.1)

Q1. What is the impact of privacy and confidentiality requirements in Cloud Computing architectures?

As we have seen through the SLR, privacy and confidentiality have a relevant impact on the design of Cloud Computing architectures and models, the definition of stakeholders' roles and the necessary technologies to maintain the required levels of confidentiality.

Through analysing privacy and confidentiality issues and challenges we have learned that data in the Cloud has to follow a specific life-cycle to ensure that this information is properly protected and managed. Most proposed solutions and suggestions revolve around protecting this data throughout all phases and fulfilling a set of privacy requirements.

From the SLR, privacy and confidentiality impacts Cloud Computing architectures by forcing stakeholders to alter cloud service models taking data protection into account and adding extra technologies to support these changes. The impact could be resumed in the following points:

- Cloud vendors need to provide a strong, robust and reliable technological base in their cloud implementations to provide a good quality service while protecting data with efficient encryption, isolation and access control algorithms.

- Cloud vendors need to properly manage the data life-cycle and define privacy models to enable organizations have more control and management over their data. The implementation of standards, following best practices frameworks and providing monitoring and accountability capabilities to customers help into achieving this goal.

- Cloud vendors need to provide consistent, clear and compliant policies to customers and offer contractual negotiation at a small fee (like Privacy-as-a-Service) to suit customer requirements.

- Cloud vendors should properly delete data and that it is not accessible in any way, as storage resources are constantly reused in a Cloud Computing environment.

- Cloud vendors should provide customers with information about the location of their data and store it on servers located in low legal risk jurisdictions.

- Cloud Customers require mechanisms to have a clear understanding of their own privacy and structural requirements, define their business specification to choose an appropriate cloud vendor whose policies fit with their requirements and assess the risks associated with moving to the cloud.

To sum up, and as Andreas Weiss pointed out in our interview (refer to Appendix C) Cloud Providers must offer security and confidence on customers' data through a solid and reliable business model. This way trust can be built up among customers leading to an increase of Cloud adopters.

Q2. Which are the currently identified issues and challenges regarding privacy and confidentiality in Cloud Computing platforms? What are some of the solutions proposed to solve these issues?

Through Chapter 4 we have identified and classified several issues and challenges found in the literature and established relationships among works in order to find more non-explicit issues and gaps. The most mentioned issues in the literature regard to virtualization, data outsourcing (specially issues regarding fitting customers' privacy requirements, audit and monitoring, unauthorized uses and access controls), legal issues and encryption. It is no surprise that these challenges are the most relevant given that virtualization is the core technology that enables Cloud Computing (see section 2.5 – Underlying technologies) and that moving sensitive data out of the boundaries of company's control and sharing responsibility with unknown parties has inherent risks and compliance issues with laws.

Some reviewed issues or challenges may not have a specified solution in the literature. However, sometimes due to the nature of some issues, they may be related with solutions provided by other authors. In the analysis performed in section 5.1 and throughout chapter 4 we discussed and proposed several alternatives, directions and linking of works that may help solving issues that in the literature did not have a direct and specific solution.

However, given the commercial nature of Cloud Computing, it is surprising that there had not been much about the cost expenses of ensuring privacy, being discussions about ethics and trustiness the most predominant in the gathered literature. Probably, given that confidentiality is a subset of privacy which just focuses on protecting information leakage, that could hinder our searching in the literature, excluding lots of useful results which could bring a more extended and accurate answer to this question. In chapter 6 we propose future work with the aim of improving this systematic review.

CHAPTER 6

FINAL CONCLUSIONS AND FUTURE WORK

Chapter 6: Final Conclusions and Future work

6.1. Conclusion

In this Master Thesis we have provided an State-of-the-Art about privacy and confidentiality issues in Cloud Computing, focused in the area of privacy and confidentiality and providing the reader with a portion of the current research background in this field. The main objective was giving readers an insight of what is Cloud Computing, and provide an overview of several research initiatives regarding this topic.

First, in chapter 2 we reviewed general aspects of Cloud Computing which enables newcomers to this paradigm to understand what is it and which are its general aspects. Afterwards, in chapter 3 we documented our literature gathering strategy for the topic of privacy and confidentiality, then exposed the results of this search in chapter 4 in the form of a Systematic Literature Review. Finally, in chapter 5 we discuss the results obtained in chapter 4 and extracted some conclusions and defined research paths by establishing relationships among requirements, phases and found issues, which gave us an idea of the impact of privacy and confidentiality in cloud architectures and requirements.

Our work provides scientific value to the community by reviewing, structuring, ordering all the found evidence in a categorized way, allowing researchers and practitioners to quickly identify their areas of interest and which proposals exist to tackle existing issues and challenges on privacy and confidentiality under Cloud Computing paradigm.

Cloud computing is a challenge specially for organizations which operate around the world, facing sometimes conflicting privacy rules and regulations. Organizations need to adopt a systematic approach to addressing privacy in the cloud and integrate privacy risk management frameworks in their business models. Suggestions like the ones offered by NIST Guidelines [SLR20] and ENISA [SLR19] can prove a useful starting point. However, given the complexity of existing global legislation, we recommend to look for legal advice depending on the importance of data stored in the cloud.

As a last personal conclusion, I could learn to stay in touch with research methods, and learned to interpret and understand on-going research and immature proposals. I also learned to perform Systematic Reviews and the importance of following the so called “scientific method”, documenting in detail searches and results to produce reproducible, verifiable and reliable output. A well performed Systematic Literature Review enables to easily keep an updated state-of-the-art over a specific topic by re-applying its research method from time to time as long as sources, searches and exclusion criteria are clearly defined.

We are aware that this Master Thesis has some weaknesses in the searches, basically due to the search expression we used, as documented in several parts through the review. In the next section, we describe and propose future work to improve the quality of the work we offer in this document.

6.2. Future work

The first thing we noted while working on the Systematic Literature Review is that filtering by confidentiality eliminated a lot of results which may have uncovered more issues related to privacy. We found that confidentiality is a subset, an attribute of privacy (dedicated to establish ways to protect from data leakages) alongside privacy-preservability (which deals with legislation and ethical matters), accountability (tracking of data) and integrity (which deals with the accuracy of data).

This means that we think it is somewhat impractical to pretend doing an unbiased and exhaustive search over privacy and confidentiality issues at the same time without harming and biasing privacy. We should either have looked for privacy in general, or focus on confidentiality and data protection (see section 3.1.2.2 – Keyword finding).

As future work or proposal for researchers interested in the topic, we visualise two ways of improving this Master Thesis and increasing the quality of the search expression by using the same keywords we found:

1) *Focusing on confidentiality:* as we have explained, confidentiality regards to the protection of data and prevention of leakages. The most relevant keywords that match this description are confidentiality, attack detection, recovery (of data loss), protection, data, access control. Mixing these keywords in a proper search expression may give relevant results which may provide more insight on issues we reviewed like virtualization, access control and encryption, approaches that aim at providing data protection in Cloud Computing.

2) *Focusing on privacy preservability*: by focusing on this attribute, we could gather more and deeper evidence on ethical issues regarding privacy on Cloud Computing, like discussions over legislation, interaction and relationship between Cloud Customer and Cloud Provider, privacy risk assessment and frameworks for requirement negotiation with CSPs.

Finally, we could provide an experimental basis to all the reviewed issues to gather more evidence, have a surface for testing and comparison and detect more technical gaps and challenges. We would do this by providing a proof-of-concept to privacy in Cloud Computing, comparing and study in-detail the data security mechanisms of several open-source cloud vendors like Eucalyptus [32], OpenNebula [31] and Nimbus [33] in order to find which approaches do they implement and compare them to popular and more advanced proprietary clouds like Amazon EC2 or Microsoft Azure.

REFERENCES

References

[1] Expert Group Report, K. Jeffery [ERCIM], B. Neidecker-Lutz [SAP Research]: *The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond 2010* (2010)

[2] J. Timmermans, V. Ikonen: *The Ethics of Cloud Computing, A Conceptual Review* (2010 2nd IEEE International Conference on Cloud Computing Technology and Science)

[3] D. Xu, H. Liu: *Reviewing some Cloud Computing Platforms* (April. 2010, pp. 161-16)

[4] R.PalsonKennedy, T.V.Gopal: *Assessing the Risks and Opportunities of Cloud Computing – Defining Identity Management Systems and Maturity Models*

[5] I. Ruiz-Agundez, Y. K. Penya and P. G. Bringas: *A Flexible Accounting Model for Cloud Computing* (2011 Annual SRII Global Conference)

[6] L. Zhao, A. Liu, J. Keung: *Evaluating Cloud Platform Architecture with the CARE Framework* (2010 Asia Pacific Software Engineering Conference)

[7] B. Sodhi and T.V Prabhakar Department of Computer Science and Engineering: *Application Architecture Considerations for Cloud Platforms*

[8] B. Prasad Rimal, A. Jukan, D. Katsaros, Y. Goeleven: *Architectural Requirements for Cloud Computing Systems: An Enterprise Cloud Approach* (J Grid Computing (2011) 9:3–26)

[9] K. Lee, D. Hughes: *System Architecture Directions for Tangible Cloud Computing* (2010 First ACIS International Symposium on Cryptography, and Network Security, Data Mining and Knowledge Discovery, E-Commerce and Its Applications, and Embedded Systems)

[10] R. B. Bohn, J. Messina, F. Liu, J. Tong, J. Mao: *NIST Cloud Computing Reference Architecture* (2011 IEEE World Congress on Services)

[11] I. Foster, Y. Zhao, I. Raicu, S. Lu: *Cloud Computing and Grid Computing 360-Degree Compared*

[12] B.Yamini, D.Vetri Selvi: *Cloud Virtualization: A Potential Way to Reduce Global Warming*

[13] Z. Shu-Qing, X. Jie-Bin: *The Improvement of PaaS Platform* (2010 First International Conference on Networking and Distributed Computing)

[14] R. Clarke: *User Requirements for Cloud Computing Architecture* (2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing)

[15] L. Zhang, Q. Zhou: CCOA: *Cloud Computing Open Architecture* (2009 IEEE International

Conference on Web Services)

[16] S. S. Yadav, Z. Hua: *CLOUD: A Computing Infrastructure on Demand*

[17] B. Gowrigolla, S. Sivaji, M.Roberts Masillamani: *Design and Auditing of Cloud Computing Security*

[18] A. Bedra: *Getting Started with Google App Engine and Clojure* (Published by the IEEE Computer Society, July/August 2010)

[19] I. Brandic, S. Dustdar, T. Anstett, D. Schumm, F. Leymann, R. Konrad: *Compliant Cloud Computing (C3): Architecture and Language Support for User-driven Compliance Management in Clouds* (Proceedings of the 3rd International Conference on Cloud Computing (IEEE Cloud 2010), 5-10 July, Miami, USA)

[20] X. Wang, B. Wang, J. Huang: *Cloud computing and its key techniques*

[21] B. Kitchenham, S. Charters: *Guidelines for performing Systematic Literature Reviews in Software Engineering* (9 July, 2007)

[22] ENISA, European Network and Information Security Agency: *Cloud Computing: Benefits , Risks and Recommendations for information security* (November 2009)

[23] B. Reddy Kandukuri, R. Paturi V, Dr. Atanu Rakshit: *Cloud Security Issues* (2009 IEEE International Conference on Services Computing)

[24] H. Takabi and James B.D. Joshi, G.J. Ahn: *Security and Privacy Challenges in Cloud Computing Environments* (IEEE Computer and reliability societies, November/December 2010)

[25] M. Okuhara, T. Shiozaki, T. Suzuki: *Security Architectures for Cloud Computing*

[26] B. Grobauer, T. Walloschek, E. Stöcker: *Understanding Cloud Computing Vulnerabilities* (IEEE Computer and reliability societies, March/April 2011)

[27] ExpertON Group: *Cloud Vendor Benchmark 2011*

[28] M. Mahjoub, A. Mdhaffar, R. Ben Halima, M. Jmaiel: *A comparative study of the current Cloud Computing technologies and offers* (2011 First International Symposium on Network Cloud Computing and Applications)

[29] GreenPeace: *How dirty is your data? A Look at the Energy Choices That Power Cloud Computing*

[30] J. Weinman*: The 10 laws of cloudonomics.* (http://tinyurl.com/5wv9d7).

[31] OpenNebula: http://www.opennebula.org

[32] Eucalyptus: http://www.eucalyptus.com/

[33] Nimbus: http://www.nimbusproject.org/

[34] R. Ananthanarayanan, K. Gupta: *Cloud analytics: do we really need to reinvent the storage stack?* (2009)

[35] Q. Zhang, L. Cheng, R. Boutaba: *Cloud computing: state-of-the-art and research challenges* (J Internet Serv Appl (2010) 1: 7–18)

[36] I. Iankoulova, M. Daneva: *Cloud Computing Security Requirements: a Systematic Review* (2012 Sixth International Conference on Research Challenges in Information Science, p 7 pp., 2012)

[37] Zimory Cloud Computing: http://www.zimory.de/

[38] RightScale Cloud Management: http://www..rightscale.com

[39] J. Peng, X. Zhang, Z. Lei, B. Zhang, W. Zhang, Q. Li: *Comparison of Several Cloud Computing Platforms* (2009 Second International Symposium in Information Science and Engineering)

[40] Asia-Pacific Economic Cooperation (APEC): http://www.apec.org

[41] ISO 27001: http://en.wikipedia.org/wiki/ISO/IEC_27001:2005

[42] Bowers KD, Juels A, Oprea A: *Proofs of retrievability: Theory and implementation* (Proc. of the 2009 ACM Workshop on Cloud Computing Security, CCSW 2009, Co-Located with the 16th ACM Computer and Communications Security Conf., CCS 2009. New York: Association for Computing Machinery, 2009)

[43] Definition of Encryption: http://en.wikipedia.org/wiki/Encryption

[44] Y. Zhang, A. Juels, M. K. Reiter, T. Ristenpart: *Cross-VM Side Channels and their use to extract Private Keys* (CCS '12 Proceedings of the 2012 ACM conference on Computer and communications security, Pages 305-316)

[45] Definition of Policy: http://en.wikipedia.org/wiki/Policy

[46] Definition of Privacy Policy: http://en.wikipedia.org/wiki/Privacy_policy

[47] *What is a policy?*: http://partcfood.msvu.ca/section4/

[58] Definition of Terms of Service: http://en.wikipedia.org/wiki/Terms_of_service

[59] *Google declares that Gmail users “should not expect privacy”:* http://www.siliconnews.es/2013/08/14/google-usuarios-gmail-privacidad/ (14 August 2013 - SiliconNews)

[60] *Sweden forbids the use of Google Apps in the public sector:* http://computerhoy.com/noticias/apps/suecia-prohibe-uso-google-apps-sector-publico-4285 (14th June 2013 – Computer Hoy)

[61] *NSA controls Internet through optical fiber private firms:* http://actualidad.rt.com/actualidad/view/99517-NSA-fibra-optica-eeuu-vigilancia (9th July 2013 – RT)

[62] *Google and Facebook DID allow NSA access to data and were in talks to set up 'spying rooms' despite denials by Zuckerberg and Page over PRISM project*: http://www.dailymail.co.uk/news/article-2337863/PRISM-Google-Facebook-DID-allow-NSA-access-data-talks-set-spying-rooms-despite-denials-Zuckerberg-Page-controversial-project.html (8th June 2013 – MailOnline)

[63] *Why Facebook Home bothers me: It destroys any notion of privacy:* http://gigaom.com/2013/04/04/why-facebook-home-bothers-me-it-destroys-any-notion-of-privacy/ (4th April 2013 – GigaOm)

[64] Amazon S3: http://aws.amazon.com/es/s3/

[65] Windows Azure: http://www.windowsazure.com

[66] Terremark: http://www.terremark.com/

[67] Rackspace: http://www.rackspace.com/

[68] Google App Engine: https://appengine.google.com/

[69] Force.com: http://www.force.com/

[70] Google Docs: http://docs.google.com/

[71] iCloud: https://www.icloud.com/

[72] Zoho: http://www.zoho.com

[73] Salesforce CRM: http://www.salesforce.com/crm/

[74] eBay: http://www.ebay.com

[75] Amazon EC2: http://aws.amazon.com/ec2/

[76] Juniper: http://www.juniper.net/

[77] Vmware: http://www.vmware.com/

Systematic Literature Review References

[SLR1] J. C. Roberts, W. Al-Hamdani: *Who can you trust in the Cloud? A review of security issues within Cloud Computing* (InfoSecCD '11: Proceedings of the 2011 Information Security Curriculum Development Conference)

[SLR2] D. Agrawal, A. El Abbadi, S. Wang: *Secure Data Management in the Cloud* (DNIS'11: Proceedings of the 7th international conference on Databases in Networked Information Systems)

[SLR3] D. Lin, A. Squicciarini: *Data Protection Models for Service Provisioning in the Cloud* (SACMAT '10: Proceeding of the 15th ACM symposium on Access control models and technologies)

[SLR4] M. Mowbray, S. Pearson, Y. Shen: *Enhancing privacy in cloud computing via policy-based obfuscation* (The Journal of Supercomputing , Volume 61 Issue 2)

[SLR5] R. Mishra, S. K. Dash: *A Privacy Preserving Repository for Securing Data across the Cloud* (2011 3rd International Conference on Electronics Computer Technology (ICECT))

[SLR6] W. Itani, A. Kayssi, A. Chehab: *Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures* (DASC '09. Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009)

[SLR7] S. A. Almulla, C.Y.Yeun: *Cloud Computing Security Management* (2010 Second International Conference on Engineering Systems Management and Its Applications (ICESMA))

[SLR8] S. Sengupta, V. Kaulgud, V.S. Sharma: *Cloud Computing Security - Trends and Research Directions* (2011 IEEE World Congress on Services (SERVICES))

[SLR9] O. Levina, J. Oetting, Y. Chou: *Enforcing Confidentiality in a SaaS Cloud Environment* (2011 19th Telecommunications Forum (TELFOR))

[SLR10] M. Zhou, R. Zhang, W. Xie, W. Qian, A. Zhou: *Security and Privacy in Cloud Computing: A Survey* (2010 Sixth International Conference on Semantics, Knowledge and Grids)

[SLR11] Z. Xiao, Y. Xiao: *Security and Privacy in Cloud Computing* (IEEE Communications Surveys & Tutorials, Volume PP, Issue 99)

[SLR12] X. Ma: *Security Concerns in Cloud Computing* (2012 Fourth International Conference on Computational and Information Sciences)

[SLR13] J. Kong: *A practical approach to improve the data privacy of virtual machines* (2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010))

[SLR14] M. Mowbray, S. Pearson, Y. Shen: *A Privacy Manager for Cloud Computing* (First International Conference, CloudCom 2009, Beijing, China, December 1-4, 2009. Proceedings)

[SLR15] Y. Tian, B. Song, J. Park, and E. Huh: *Inter-cloud Data Integration System Considering Privacy and Cost* (Second International Conference, ICCCI 2010, Kaohsiung, Taiwan, November 10-12, 2010. Proceedings, Part I)

[SLR16] T. Mather, S. Kumaraswamy, S. Latif: *Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance* (available for free at Google Scholar)

[SLR17] S. Pearson: *Taking Account of Privacy when Designing Cloud Computing Services* (CLOUD '09 Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing)

[SLR18] R. Gellman: *Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing* (available online at https://observatorio.iti.upv.es/resources/report/186)

[SLR19] ENISA: *Cloud Computing: Benefits, risks and recommendations for information security* (available online at http://www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-risk-assessment/at_download/fullReport)

[SLR20] NIST: *Guidelines on Security and Privacy in Public Cloud Computing* (available online at http://csrc.nist.gov/publications/nistpubs/800-144/SP800-1[Other]44.pdf)

[SLR21] S. Subashini and V. Kavitha: *A survey on security issues in service delivery models of cloud computing* (Journal of Network and Computer Applications, Vol. 32, No. 1, pp 1-11, 2010)

[SLR22] A. Cavoukian: *Privacy in the clouds* (Identity in the Information Society Volume 1, Issue 1 , pp 89-108)

[SLR23] S. Pearson, A. Charlesworth: *Accountability as a Way Forward for Privacy Protection in the Cloud* (CloudCom '09 Proceedings of the 1st International Conference on Cloud Computing, Pages 131 – 144)

[SLR24] D. Chen, H. Zhao: *Data security and privacy protection issues cloud computing* (2012 International Conference on Computer Science and Electronics Engineering, Hangzhou, China, March 2012)

[SLR25] M. Mowbray, S. Pearson: *A Client-Based Privacy Manager for Cloud Computing* (COMSWARE '09 Proceedings of the Fourth International ICST Conference on COMmunication System softWAre and middlewaRE)

[SLR26] P. T. Jaeger, J. Lin, J. M. Grimes: *Cloud Computing and Information Policy: Computing in a Policy Cloud?* (Journal of Information Technology & Politics, 5:3, 269-283, 2008, article downloadable at http://dx.doi.org/10.1080/19331680802425479)

[SLR27] D. Chen, H. Zhao: *Data Security and Privacy Protection Issues in Cloud Computing* (2012 International Conference on Computer Science and Electronics Engineering)

APPENDIXES

Appendix A: Glossary

Table A-1 shows a list of relevant terms and abbreviations that appear across this document with their definitions.

Term	Definition
Accountability	The capability of identifying with undeniable evidence who triggered what event to the data.
Audit and monitoring	Deals with how can organizations monitor and control the activities of their Cloud Service Providers over their data and watch what happened in the Cloud system, with the objective to assure that privacy requirements, SLA and compliance with laws are enforced when their personal information is in the cloud
API	An *Application Programming Interface* is a series of software routines and development tools that comprise an interface between a computer application and lower-level services and functions (such as the operating system, device drivers, and other software applications). APIs serve as building blocks for programmers putting together software applications. In the context of cloud computing, APIs are sets of web services methods for accessing/manipulating cloud resources.
Asymmetric encryption	The use of two different keys, first for encryption (public key) and then for decryption (private key) of data.
Authentication	The act of confirming the identity of an individual or system.
Authorization	The act of specifying access rights to resources or functionality.
Confidentiality	Is a property of privacy that ensures that no information is made available to unauthorized individuals, entities or processes.
CSP	A *Cloud Service Provider* provides and manages services of cloud computing platform.
Elasticity	The capability of a system to increase or decrease its computing resources as needed.
Hybrid cloud	An environment consists in a combination of private and public clouds where an organization may run non-core applications in a public cloud, while maintaining core applications and sensitive data in-house in a private cloud.
Hypervisor	A software/hardware platform in a virtualization system that manages multiple Virtual Machines and allows several operating systems to run on a host computer concurrently.
IaaS	*Infrastructure as a Service* is the delivery of computer infrastructure as

	a service where OS and applications can be deployed.
IDM, IAM	*Identity management* and *Identity and Access Management* is the management of the identity life cycle.
ISP	An *Internet Service Provider* is a company that offers its customers access to the Internet.
ITIL	The *Information Technology Infrastructure Library* is a set of concepts, policies and best practices for managing IT infrastructure, development, and operations.
Key management	Provisions made in a cryptography system design that are related to the generation, exchange, storage, safeguarding, use, vetting, and replacement of keys.
Multitenancy	Multitenancy is the capability to allow multiple users share and use resources at the same time.
NIST	The National Institute of Standards and Technology is a standards organization and measurement standards laboratory.
OAuth	An open authorization protocol standard that lets users give third-party websites limited access to their data without giving away their passwords. The OAuth protocol enables websites or applications (consumers) to access protected resources from web services (service providers) via an API, without requiring users to disclose their service provider credentials to those consumers.
OpenID	An open, decentralized, free framework for a user-centric digital identity. OpenID eliminates the need for multiple usernames across different websites, simplifying your online experience.
Outsourcing	Refers to the export and delegation of several business processes to third party companies.
PaaS	*Platform as a Service* is a delivery model whereby cloud vendors offer a development studio solution through the cloud.
Privacy	Privacy relates to the collection, use, disclosure and destruction of personal data.
Private cloud	A deployment model that emulates public cloud computing, but on a private network.
Public cloud	A cloud service that is hosted, operated, and managed by a third-party vendor from one or multiple data centres, and offered to multiple customers.
SaaS	*Software as a Service* is a model of software deployment whereby a provider licenses an application to customers for use as a service through the cloud.
SAML	*Security Assertion Markup Language* is an XML-based standard for exchanging authentication and authorization data between security domains—that is, between an identity provider (a producer of

	assertions) and a service provider (a consumer of assertions)
Scalability	The capability of a system to scale its capacity without loosing performance.
SLA	A service-level agreement is a part of a service contract where the level of service is formally defined.
Symmetric encryption	Use of a single secret key for both the encryption and decryption of data.
Virtualization	The creation of a virtual (rather than actual) version of something, such as an operating system, a storage device, an application, or network resources.
VPN	A virtual private network is a computer network in which some of the links between nodes are carried by open connections or virtual circuits in some larger networks (such as the Internet), as opposed to running across a single private network.
XACML	eXtensible Access Control Markup Language is a declarative access control policy language implemented in XML and a processing model, describing how to interpret the policies.

Table A-1: Terms and definitions related to Cloud Computing

Appendix B: Popular Cloud Vendors

This section will list and describe some of the most popular Cloud Computing providers, with the aim of giving readers awareness about the Cloud market.

Amazon EC2

Amazon Elastic Compute Cloud (Amazon EC2) offers IaaS which allows customers to launch and manage server instances which run under the Xen virtualization technology. After creating and starting an instance, users can fully control all the software stack and upload applications and make changes to it, and finally bundle an image machine, launching an identical copy of that image whenever and wherever necessary, as EC2 also provides the ability to place instances in multiples locations, allowing for an inexpensive and low latency network connectivity, EC2 machine images are stored in and retrieved from Amazon Simple Storage Service (Amazon S3), which stores data as *objects* that are grouped in *buckets*. Each object contains from 1 byte to 5 gigabytes of data. Buckets must be explicitly created before they can be used.

EC2 can be monitored with Amazon CloudWatch, a management tool to collect and process information to obtain metrics like CPU usage, network input/output, disk read/write operations, etc.

Microsoft Windows Azure platform

Microsoft Windows Azure platform offers PaaS and provides tools for developers that ease the deployment and development of Windows-based applications. It consists in three components, each of them providing a specific set of services to customers:

- Windows Azure, provides a Windows-based environment for running applications and storing data on servers in data centers.

- SQL Azure provides data services in the cloud based on SQL Server, providing a database management system (DBMS) in the cloud.

- .NET Services offer distributed infrastructure services to cloud-based and local applications, facilitating the creation of distributed applications. Windows Azure supports applications built with the .NET Framework and other languages like C#, Visual Basic and C++, and developers can create web applications using technologies such as ASP.NET and Windows Communication Foundation (WCF). .NET Services also offers the Access Control component, which provides a cloud-based implementation of single identity verification.

Windows Azure platform offers monitoring services with software called the Fabric Controller. With each application in the cloud, the users upload a configuration file that provides an XML-based description of what the application needs. This file allows the Fabric Controller to decide where new applications should run, choosing physical servers to optimize hardware utilization.

Google App Engine

Google App Engine is a PaaS used to deploy web applications in Google data centers. The currently supported programming languages are Python and Java, and web frameworks like Django, CherryPy, Pylons, and web2py, as well as a custom Google-written web application framework similar to JSP or ASP.NET. Google handles deploying code to a cluster, monitoring, failover, and launching application instances as necessary.

AbiCloud

Abicloud is a cloud computing platform is an open-source cloud platform used to build, integrate and manage public, private and hybrid cloud in the homogeneous environments. Using Abicloud, user can easily and automatically deploy and manage the server, storage system, network, virtual devices and applications and so on. Abicloud supports virtual machine platforms like VirtualBox, VMWare and Xen. The main difference between Abicloud and other cloud computing platforms is its powerful web-based management function and its core encapsulation manner.

Abicloud can be used to deploy and implement private cloud as well as hybrid cloud according to the cloud providers' request and configuration. It can also manage EC2 according to the rules of protocol.

Eucalyptus

Eucalyptus (which stands for Elastic Utility Computing Architecture for Linking Your Programs To

Useful Systems) is an open-source implementation of Amazon EC2 and compatible with business interfaces. It implements an elastic computing structure which uses clusters or workstations with a standard based on service level protocol that permit users lease network for computing capability. Currently, Eucalyptus is compatible with EC2 from Amazon, and may support more other kinds of clients with minimum modification and extension, allowing the connection of customers' applications.

Nimbus

Nimbus is a cloud computing solution which provides IaaS that was used on scientific research on its first stages. It allows users to build the computing environment through the deployment of virtual machines. It includes context agent module, web service resource framework module, EC2 WSDL module and remote interface module used to manage all kinds of physical resources on the cloud computing platform.

OpenNebula

OpenNebula is a research project in virtualization infrastructure and cloud computing of European Union. Like Nimbus, OpenNebula is also an open source cloud service framework, and allows users deploy and manage virtual machines on physical resources, set up flexible virtual infrastructure that can automatically adapt to the change of the service load and deploy any types of clouds. Though OpenNebula is mainly used to manage private and hybrid infrastructure, it also supports public cloud platform by providing interfaces and functions to virtual machines, storage and network management, and so on.

OpenNebula cloud computing platform has many advantages: It can dynamically adjust the scale of the infrastructure of the cloud platform by increasing the number of hosts and partition clusters to meet different requirements, and can manage all the virtually and physically distributed infrastructures and can create infrastructure with the heterogeneous resources at data center. This can guarantee use the resources more efficiently and can much reduce the number of the physical resources through the integration of servers which further reduce the cost.

Appendix C: Interview with Andreas Weiss

Andreas Weiss is the Director of EuroCloud. EuroCloud is an independent non-profit organisation aimed at creating awareness of Cloud Computing throughout the society and take an active role in the design of cloud industry processes and standards. This organisation provides the EuroCloud Certificate to good Cloud Service Providers.

Andreas held a conference on 23rd of March at the Swiss Cloud Conference 2012 in Switzerland. Cristoph Fischer, a student of Business Information Technology from the Zurich University of Applied Sciences who lives in Switzerland could attend to that conference, and we thought it would be interesting to ask Andreas Weiss some questions about Cloud Computing, so we prepared some questions together and formulated them to Andreas in a personal interview.

In this section we will cover the interview Cristoph Fischer (C) and Andreas Weiss (A) held in the Swiss Cloud Conference 2012.

C: Among all non-functional requirements, is really security the most important one for customers who have to store their data in a multi-tenant environment?

A: Yes, indeed. Data is one of the most important and valuable resource any company has, so it's obvious its their main concern when they partially lose control over it and put it into the Web.

C: If that's so, in which ways do cloud providers offer security and confidence to cloud customers?

A: They should have a good business model. It should not happen that the provider can be taken over or go bankruptcy. For me it's very important to know that the provider has been doing this for some time!

C: Which are the main "keypoints" that define a cloud provider as a "good one"? Are there any existing guidelines?

A: Consistency, reliability and motivation - no guidelines so far.

C: Which are nowadays the most popular "Free clouds" besides Google App Engine?

A: When talking about Software as a platform, Facebook/Twitter/Google Analytics are the most popular free clouds. Then Amazon Cloudfront (simpel CDN) too is pretty popular.

C: Besides big enterprises, do you think most SMEs will slowly move to cloud services in a near future?

A: Yes, SMEs are the most profitable enterprises when we look at them in the long term (long tail - most beneficial).

C: What is the latest greatest "hit" or finding regarding Cloud Computing topic?

A: If you mean by "hit" just what people found out recently than it is for sure the fact, that latency (for the usability) is a big problem for Cloud Providers and their clients. Speed sells and if you can not be on the fast track, than you have a problem. Check https://cloudsleuth.net/global-provider-view for more information.

C: Do you think that every business who wants to achieve a cost reduction in their processes should move his data and operations to a cloud? Or there are situations where cloud is just not worth?

A: No, not every business does benefit from this situation. What needs to be done in the first place is that somebody writes a good business case. No business case equals to no benefit from the cloud. I don't have any detailed information about other situations but actually think if you can handle the loss of your soft factors, all businesses can achieve cost reduction.

C: Well, these have been all questions I had for you, Andreas. Thank you for your time and your helpful answers.

A: Thanks to you.