Information Processing in the Cloud

Resource Allocation and Security Perspective

by

Jayachander Surbiryala

A dissertation submitted in partial satisfaction of
the requirements for the degree
PHILOSOPHIAE DOCTOR (PhD)

Faculty of Science and Technology
Department of Electrical Engineering and Computer Science
September 2019

University of Stavanger
N-4036 Stavanger
NORWAY
www.uis.no

ISBN 978-82-7644-885-6
ISSN 1890-1387

PhD Thesis UiS no. 487

Summary

Cloud computing has been adopted at faster rates due to its advantages, that are brought to the customers over the Internet. Usage of various services over the Internet is quite high. Users are storing large amounts of data in the Cloud environment, because of On-demand self-service, resource pooling, transparency, pay per-use, and access over the Internet.

As the usage of services over the Internet are increasing, the data generated from various applications and customers is also increasing. Growth in the data creation also leads to the rapid adoption of cloud services. This adoption can either be for storage of the information, processing of the data, or analysis of the information to get actionable intelligence.

Information has become the central part of many organizations and the analysis of such information is required for decision making. Due to existence of Big Data with different volume, velocity, variety, veracity, and value, customers need dynamic solutions to meet their requirements. Large-scale data management and analysis can be easily accomplished dynamically in the Cloud environment without worrying about underlying infrastructure.

As cloud usage is increasing, customers are using more than one cloud service provider depending on their needs and requirements. They have the liberty to compare and choose from the various options from multiple cloud service providers available to store, process, and analyze their data in the cloud environment. As the customer's data is being distributed across multiple cloud service providers, they will not have complete control over their data that is stored and processed on remote servers, where they only have limited access to the actual underlying infrastructure. Data security plays an important role in the Cloud computing environment. Therefore, to protect the customer's data, which gets distributed across the cloud, we need to have proper security mechanisms to protect user data in the cloud environment.

This thesis addresses the problem for resource allocation with cost-effective solution for customers to choose the cloud services from different cloud service providers using nash bargaining principles for distributed resource allocation. Distributed resource allocation helps the customers to reduce the cost of using the cloud services, for the same amount of resources, across various cloud service providers. Further, we have identified the problems associated with usage of the Cloud for the data distribution across the various cloud service providers such as data recovery, security and privacy aspects for customers. We have proposed several methods to protect the users' information in the Cloud while the customers are still using the Cloud services and we introduced some approaches to protect the information in the Cloud after deletion of their data.

Acknowledgements

It was a privilege to work under the supervision of Prof. Chunming Rong. I would like to express my sincere gratitude for the opportunity provided to work under his supervision, which has helped in many aspects during course of my Ph.D study and related research. I would like to thank him for his patience, motivation, immense knowledge and his extended support to complete this thesis through suggestions and encouragement. It was an honor to work under his supervision.

I would like to thank my co-supervisor Assoc. Prof. Chunlei Li, University of Bergen (UiB) for his support during my research work. His guidance helped me in research and writing this thesis.

My sincere thanks also goes to Prof. Weizhong Qiang at Huazhong University of Science and Technology, Dr. Yuri Demchenko and Dr. Zhiming Zhao, senior researchers at University of Amsterdam for providing an opportunity to work with their teams as a visiting research scholar and gave access to the laboratory and research facilities at their respective universities.

I would like to thank Assoc. Prof. Antorweep Chakravorty, University of Stavanger (UiS) and Dr. Bikash Agrawal for your motivation and support throughout this research. In particular, I am grateful to Prof. Christoph Busch, Norwegian University of Science and Technology (NTNU), Prof. Raghavendra Ramachandra, NTNU and Assoc. Prof. Kiran Bylappa Raja, University of South-Eastern Norway (USN) for introducing me to the field of research.

I wish to extend my thanks to COINS, CIPSI and Department of Electrical Engineering and Computer Science, UiS for the financial support provided for my learning and research related activities, academic support by UiS, administrative support by head of the department Tom Ryen, Russel Gene Wolff and Kaja Gjersem Nygaard. I sincerely want to thank my colleagues Faraz Barzideh, Cristina Viorica Heghedus, Rahul Mishra and other researchers at UiS for their help, support and valuable discussions during course of this research.

Special thanks to my parents for supporting me spiritually throughout this research, writing this thesis and my life in general. I would also like to thank my brother for supporting me through the difficult times. Finally, I thank my wife Jaya for her support, patience, understanding and encouragement to finish my PhD. Much of my accomplishments would not have been possible without them.

Jayachander Surbiryala, September 2019

Preface

This thesis is submitted in partial fulfillment of the requirement for the degree of Philosophiae Doctor (PhD) at the University of Stavanger, Norway. The research was carried at University of Stavanger during the period from May 2016 to April 2019, with a research visit to Huazhong University of Science and Technology, Wuhan, China during the period September 2018 to December 2018 followed by an another research visit to University of Amsterdam, Amsterdam, Netherland during the period January 2019 to April 2019. The thesis is written based on the published research articles. Published papers are reformatted to fit with format of the thesis. Content of the original published articles are self-contained.

List of Abbreviations

AWS	Amazon Web Services
CIS	Cloud Information Service
CloudSim	Cloud Simulator
CSA	Cloud Security Alliance
CSP	Cloud Service Provider
CRBM	Conditional Restricted Boltzmann Machines
CCSP	Collection of Cloud Service Providers
FCRBM	Factored Conditional Restricted Boltzmann Machines
GB	Gigabyte
IaaS	Infrastructure as a Service
IEEE	Institute of Electrical and Electronics Engineers
IEEE-SA	IEEE Standards Association
IoT	Internet of Things
IT	Information Technology
ITU	International Telecommunication Union
LSTM	Long Short Term Memory
NASA	National Aeronautics and Space Administration
NIST	National Institute of Standards and Technology
PaaS	Platform as a Service
PCA	Principal Component Analysis
PB	Petabyte
SaaS	Software as a Service
SLA	Service Level Agreement
SOM	Self-Organizing Map
TB	Terabyte
USM	User Shredder Module
VM	Virtual Machine

Contents

List of Figures

List of Papers

The following papers are included in this thesis:

- **Paper I**

  Resource Allocation in Cloud-Based Distributed Cameras
  B. Agrawal, J. Surbiryala, C. Rong
  Published in the proceedings of 2017 IEEE 6th International Congress on Big Data (BigData Congress).

- **Paper II**

  Data Recovery and Security in Cloud
  J. Surbiryala, C. Rong
  Published in the proceedings of 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA).

- **Paper III**

  Secure Customer Data over Cloud Forensic Reconstruction
  J. Surbiryala, C. Rong
  Published in the proceedings of 2018 IEEE International Conference on Consumer Electronics (ICCE).

- **Paper IV**

  Improve Security over Multiple Cloud Service Providers for Resource Allocation
  J. Surbiryala, B. Agrawal, C. Rong
  Published in the proceedings of 2018 1st International Conference on Data Intelligence and Security (ICDIS).

- **Paper V**

  Method to Solve a Privacy and Security Issue in Cloud for Energy Informatics
  J. Surbiryala, C. Rong
  Accepted for published in the proceedings of the 13th World Congress on Engineering Asset Management (WCEAM 2018).

- **Paper VI**

  A Framework for Improving Security in Cloud Computing
  J. Surbiryala, C. Li, C. Rong
  Published in the proceedings of 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA).

Chapter 1

Introduction

This chapter provides an introduction to the research work and it is structured as follows. The first section presents an overview of cloud computing. The second section describes the problem. The third section provides the motivation and justification. The fourth section defines the research questions. The fifth section lists the published research articles and finally, the scope of the thesis and outline of the thesis are presented in six and seventh sections respectively.

1.1 Cloud computing

Cloud computing has shown rapid growth in the development and adoption of its services in the recent years. The role of Cloud computing is to provide computing power or services over the Internet, to a large number of end-users or customers, reliably and efficiently. Delivering the computing power as a utility to end-users is discussed by John McCarthy at MIT in 1961 as "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility. The computer utility could become the basis of a new and important industry." [72]

The "cloud" is narrowly defined as the delivery and usage model of the

computer infrastructure. In the "cloud", people can get the resources they need at the time and in addition these resources are infinitely expandable. The understanding of cloud computing and its broad meaning refers to the delivery and use of services. Over the network, people can obtain the services they need according to their own requirements and with the continuous expansion of network resources. These types of services are diverse, such as, processing power, software applications, Internet-based applications, or other services (which cover other service models). Moreover, Cloud computing has the advantages of large scale, virtualization, high reliability, generality and high scalability [54, 52].

Cloud computing services are delivered over three main service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). These services are provided based on the subscription model that the end-users have chosen with the cloud service provider and they pay for the services they have used. A core concept for cloud computing is to adopt a service model that continuously integrates resources, software and applications [52]. In a cloud computing environment, cloud service providers need to have capabilities for a variety of data, data security mechanisms, simplicity, reliability, etc. [46]. Apart from providing the services, they should be capable of communicating with various services for proper functioning of the cloud. In order to perform some operations they often need to meet the end-user requirements .

1.2 Problem description

Cloud computing has gained a lot of popularity in the recent years, because of many advantages provided by the Cloud computing environment. As a result, there has been a natural increase in the use of cloud services, both among individuals and organizations across the globe. This usage may include storage, processing and analysis. With this adoption of cloud services at all levels across the community, cloud computing has become an indispensable part for many users and organizations, contributing to the proper functioning of their day

to day activities. With the greater use of cloud services for many of the operations on a daily basis, there comes a need for improving the cloud services at various levels. Improvements may refer to either cost-efficiency using the cloud services or to providing the proper security measures for the data that has been stored and processed in the cloud environment [62].

As the cloud services are available to everyone across the globe from the cloud service providers, there should be a proper trade-off for the resources offered by the cloud service providers to end-users based on the location or even usage characteristics (such as storage, processing power and so on) [52]. Considering these parameters, customers should be able to choose the cloud service providers who meet their requirements with the minimum cost. Nonetheless, as the data gets distributed across cloud service providers, it raises more security concerns for the end-users.

Even though there exist many security mechanisms in the cloud environment to protect the data in the Cloud [47, 68, 18, 48, 65], still in many scenarios we need to have proper mechanisms to protect the customer's data. To be able to protect the data, first, we need to identify the problems posed to the customer's data in the cloud environments. Then, we present some of the techniques that can be used to protect the customer data in the cloud environments, using our proposed methods or framework for the Cloud.

1.3 Motivation and justification

Cloud computing plays an essential role in the modern world for many of everyday applications. For instance, storage and usage of the personal data, or processing the information received from various sources to get the actionable intelligence from the Big Data, that is available for more and more organizations. In order to extract valuable knowledge from the Big Data, this data needs to be processed. However, due to the high number of available cloud service providers, end-users may face challenges when choosing the appro-

priate providers, as they need to meet their own requirements by using a single cloud service provider or several cloud service providers to distribute tasks across them. The providers are chosen based on the better deals with lower cost, more CPU, storage and processing power [73, 64, 59, 45, 44, 42, 38, 40].

In the case of using multiple cloud service providers, as the data gets distributed across the various providers, end-users may start to lose control over their data. Thus, how exactly end-users will be able to protect their confidential data in the cloud, becomes a crucial question [27]. On the other hand, once the processing of the data is completed end-users can delete their data in the Cloud. However, simple deletion does not guarantee that their data will be removed entirely from the cloud service provider's servers. End-users may not know if their data is actually deleted or it just appears to them as the information is deleted but in reality it is available on cloud service provider's backup servers [28].

All these aspects raise serious security concerns for the end-users data in the cloud. Data security plays a vital role in using cloud computing, as there are various potential risks for the customer's data. Some of these risks are difficult to identify and access before users have faced such scenarios. However, other risks are well known and they need to be solved in the Cloud computing environment.

Often the reasons for not solving some of the well-known issues are hard to address, or it involves adopting or developing new technologies, which will be an additional burden for the cloud service providers. Thus, introducing third-party applications might also solve the problems to some extent, but it will also add extra players into the scenario. This leads to the trust in the third party playing a crucial role when actually considering the third-party applications and frameworks. Trust in third-party applications can be chosen based on the cloud service provider's evaluation or based on the end-users trust in them. End-user's confidence might be based on previous usage of services from them, or based on the feedback given by the other end-users, or it can be based on the applicability and trustworthiness of the applications.

1.4 Research questions

The thesis aims to address the following research questions: First, it establishes the background for resource allocation among the various cloud service providers. Second, it develops frameworks to protect the data in the cloud and multi-cloud environment, where resources are shared or used from various cloud service providers even after the customer's stop using the cloud services. Third, it presents other frameworks to protect the customer's data while using the cloud services. Adoption of these frameworks would solve many problems in the cloud environment.

Figure 1.1. presents the structural flow of research objectives which are addressed in this thesis:

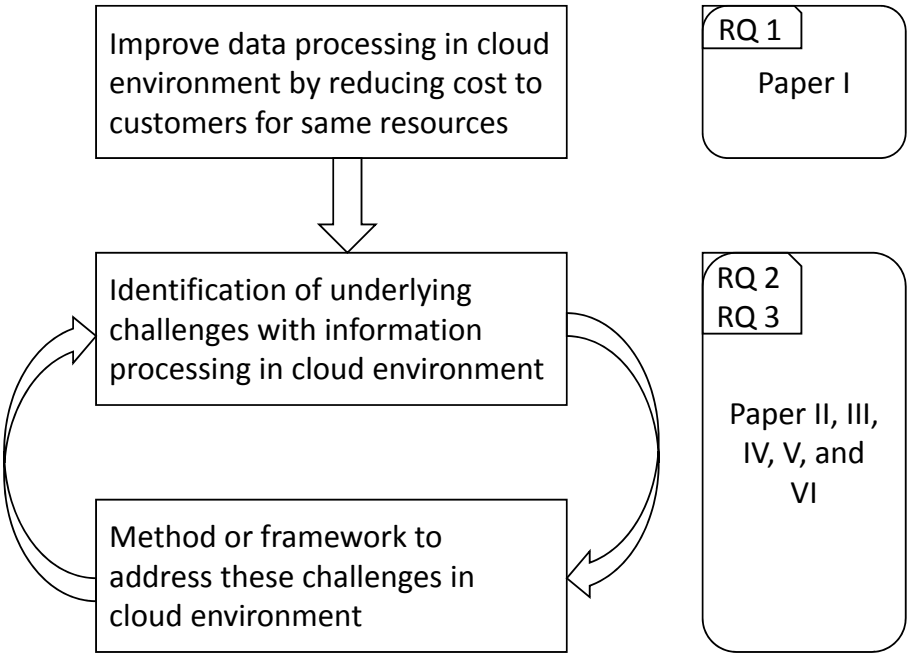

Figure 1.1: Structural diagram of research objectives and publication

The research questions proposed in the study are stated as follows:

(1) Can resource allocation be improved by utilizing the multiple cloud service provider services?

(2) Does the usage of cloud (for resource allocation) raises any other challenges?

(3) Whether identified challenges can be addressed?

1.5 Research publications

A list of research articles published during this research is presented in following subsections, which contains articles that are included in this thesis and additional articles which are not included in this thesis.

Figure 1.2. presents the flow of research objectives and their relationship to the published research papers. Paper I [32] addresses the problem of resource allocation across various cloud service providers, Paper II [28], III [31], and IV [27] discuss about the data security once the customer stops using the cloud services. Paper V [30] and VI [34] considers about the security aspects when the customers are still using the cloud services.

1.5.1 List of included papers

- Paper 0 [25]: "Cloud Computing: History and Overview", is accepted for publication in the 3rd IEEE International Conference on Cloud and Fog Computing Technologies and Applications (IEEE Cloud Summit 2019), IEEE, 2019.

  This paper is edited version of the chapter 2, which covers introduction to Cloud computing and presents an overview of various aspects in the Cloud.

- Paper I [32]: "Resource Allocation in Cloud-Based Distributed Cameras", was published in 2017 IEEE International Congress

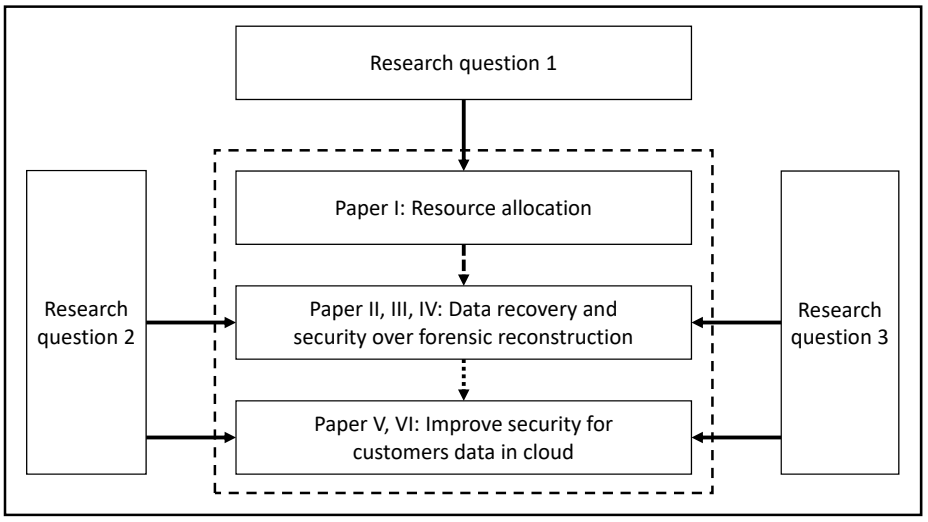

Figure 1.2: Flow of research objectives and their relationship. The dotted square around the papers represents the Cloud environment. Solid lines from research questions to papers represent where these research questions are addressed in all of these papers; dotted lines between the papers represents the flow of papers.

on Big Data (BigData Congress), IEEE, 2017.

In this paper, we proposed a cost-effective and dynamic resource allocation for handling cloud based resource allocation across various cloud service providers to handle large amounts of data in real time.

- Paper II [28]: "Data Recovery and Security in Cloud", was published in 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), IEEE, 2018.

  In this paper, we have demonstrated the possibility of data recovery from the cloud infrastructure once the customers have deleted their data. To address this problem, we have proposed a framework to protect the customer data in the cloud once they are done using their data in a cloud environment.

- Paper III [31]: "Secure Customer Data over Cloud Forensic Reconstruction", was published in 2018 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2018.

In this paper, we have proposed a new framework for Cloud to address the data recovery even after deleting the customers data. The proposed framework uses the principles of forensic applications to protect the customers data from recovery once the data has been deleted.

- Paper IV [27]: "Improve Security Over Multiple Cloud Service Providers for Resource Allocation", was published in 2018 1st International Conference on Data Intelligence and Security (ICDIS), IEEE, 2018.

  In this paper, we have discussed the security issues posed by adoption of cloud services for resource allocation from various cloud service providers. To address these problems, we have proposed to use a shredder module across the various cloud service providers to protect the customer data in the Cloud.

- Paper V [30]: "Method to Solve a Privacy and Security Issue in Cloud for Energy Informatics", was published in Proceedings of the 13th World Congress on Engineering Asset Management (WCEAM 2018), Springer, 2018.

  In this paper, we have considered scenario where customers electricity consumption are stored in the Cloud. To provide security and privacy for the customers, we have proposed to aggregate the data of several customers to leave the patterns of their household usage and providing them with required privacy for the customers even in case someone looks at their energy consumption levels which are stored in the cloud.

- Paper VI [34]: "A Framework for Improving Security in Cloud Computing", wash published in 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), IEEE, 2017.

  In this paper, we have presented various aspects related usage of cloud services and highlighted the ethical and security aspects with customers' data. To address these aspects we have proposed a framework to solve the data security problems, which in turn solves ethical issues associated the customers' data.

1.5.2 List of additional papers

- Paper VII [33]: "PhD Forum: Improving the Security for Storing the Big Data in Cloud Environment", was published in 2017 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, 2017.
- Paper VIII [29]: "Data Recovery in Cloud Using Forensic Tools", was published in 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), IEEE, 2018.
- Paper IX [23]: "Mpchecker: Use-After-Free Vulnerabilities Protection Based on Multi-Level Pointers", was published in IEEE Access 7 (2019).
- Paper X [24]: "Operating Permissioned Blockchain in Clouds: A Performance Study of Hyperledger Sawtooth", was published in 2019 18th International Symposium on Parallel and Distributed Computing (ISPDC), IEEE, 2019.

1.6 Scope of the Thesis

The main scope of the thesis is to identify the issues which are present when adopting cloud services, referring to usage and processing of information in the cloud. Furthermore, it aims to present possible solutions to solve these problems and understand various factors influencing the utilization of cloud services by customers. This will help and encourage end-users to use cloud services by cloud service providers. In addition, it also aims to improve the end-user and provider relationship, by strengthening the cloud services. This can be done either by enhancing the trust of customers in cloud service providers or by providing solutions to some of the essential problems in the Cloud such as security and privacy. The scope of this thesis is limited to explaining the concepts and designs proposed with limited simulation results in private cloud environment.

1.7 Thesis Outline

This thesis consists of two parts: part one presents an overview of basic concepts from Chapter 1 to Chapter 5, followed by part two consisting of a list of research articles used in the thesis.

Chapter 2 presents an introduction to the cloud computing with various details related to history of cloud computing, evolution of cloud computing, characteristics, models and other aspects related to cloud computing. Chapter 3 presents the background for related technologies that are required to understand this thesis. Chapter 4 presents a summary of 6 published research articles used in the thesis. Chapter 5 presents the conclusion of the research work carried in this thesis and future work.

Chapter 2

Cloud Computing

This chapter is organized as follows: Introduction to cloud computing is presented in first section. The second section presents the brief history of cloud computing and the third section talks about the evolution of cloud computing from predecessors utility computing and grid computing. In the fourth section, development of cloud computing is presented along with fundamental characteristics, delivery models, and deployment models. We have discussed cloud applications and security in the fifth section and the sixth section respectively.

Edited version of this chapter has been accepted for publication at IEEE Cloud Summit 2019 [25].

2.1 Introduction

Cloud computing is a model where services are added on demand over the Internet and are dynamically scaled. In the past, the cloud was often used to represent part of the Internet with some infrastructure. Nowadays, cloud is used as a metaphor for the services provided over the Internet. Due to the rapid evolution of cloud services, cloud computing now supports a large number of operations in a fraction of seconds compared to traditional systems where the number of

transactions was limited. This computational power can be used for pre-processing, analysis, and forecasting of future events. To use cloud services users still need to connect with their devices to access and work on these virtual devices with massive processing power around the world [56].

Cloud computing is nothing but the integration of distributed computing, parallel computing, utility computing along with network storage, virtualization, load balance, high available, and various other related technologies. Cloud computing is defined in several ways, but National Institute of Standards and Technology (NIST) defines that "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" [52].

Cloud computing is the delivery of resources or services over the Internet that typically involves dynamic scaling of physical and virtual resources as an add-on service or based on their usage. As per NIST definition of Cloud computing, it should have following features: convenient, usable, on-demand computing services, servers, storage space, software applications, services over the Internet with no or little interaction with the cloud service provider.

With the rapid adoption of cloud computing applications and the Internet of Things (IoT), there is an increasing demand for Big Data storage and processing. Many of the users and small organizations are also looking for significant computing power and high availability.

Cloud computing is often confused with grid computing, utility computing, and autonomic computing. Grid computing is a distributed computing system where a group of computers coupled to form a virtual machine to perform large tasks [70]. Utility computing is a method of packaging several services such as facilities, storage, and so on for billable IT resources [69]. Autonomic computing is a system that has self-management capabilities [71]. Cloud computing capabilities depend on the clusters deployed (grid computing) with various functionalities of utility and autonomic computing.

2.2 Brief History

Evolution of the cloud computing can be mapped back to older systems which have been used in the real-time long before cloud computing has come into existence. In "Cloud Computing", the word "Cloud" means carrier or provider who provides the services over the Internet. "Computing" is the processing or computations or calculations or various resources that are provided by computer. The concept of cloud computing traces back to 1961 by John McCarthy at MIT: "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility. The computer utility could become the basis of a new and important industry." [72]

One of the first companies to start working with the concept of the "cloud computing" is formed by Salesforce in late 1990 [19]. The company started providing their Software as a Service (SaaS), which provided customer relationship management for its users. Salesforce model is one of the typical patterns of cloud computing, in addition to "Platform as a Service" (PaaS), which provided customers with development platform such as Microsoft Azure [12] and Google's Application Engine [7]. The other form is "Infrastructure as a Service" (IaaS) model such as Amazon Elastic Compute Cloud (EC2) started in 2006 [1].

In 2007, many of the US universities started collaborating with Google and IBM and promoted cloud computing programs at their universities. This helped reduce the cost for academic research, sharing the resources between the students, and to build substantial processing power or computing power to access it over the Internet. Many more universities around the globe followed the same trend during the subsequent years [61].

In July 2010, NASA and Rackspace started a joint project called OpenStack with several vendors including AMD, Intel, and Dell. Later on, many other organizations have joined the project. A non-profit organization called OpenStack Foundation is formed in September

2012 to promote OpenStack [20]. Now more than 500 companies are supporting the project [16]. Around 6800 companies are using OpenStack to deploy their cloud services [9].

In October 2011, Trusted Cloud Initiative by Cloud Security Alliance (CSA) published a white paper to help cloud service providers to develop cloud services that meet the requirement for industry standards, secure, access controllable, inter-operable, and manageable [53]. Figure 2.1 shows the history of cloud computing.

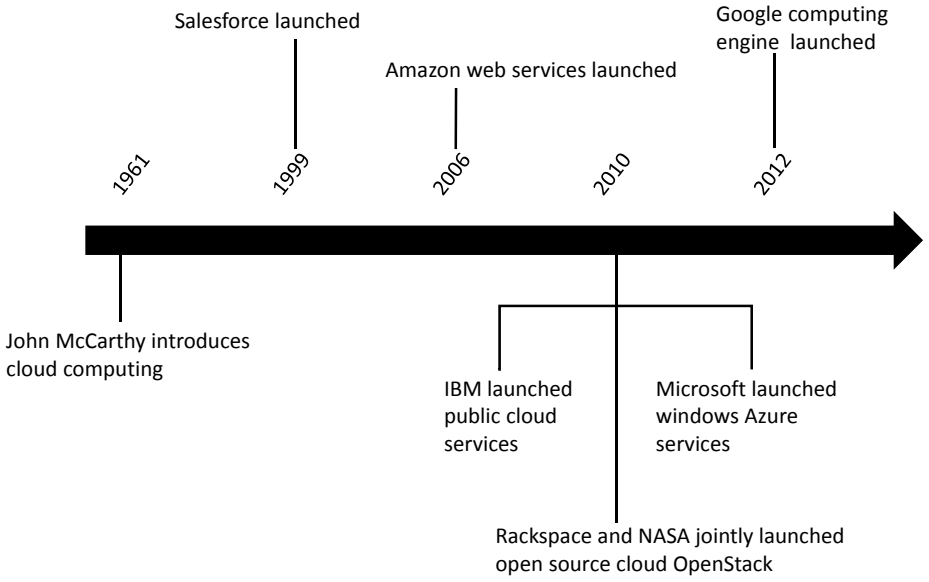

Figure 2.1: History of cloud computing

2.3 Evolution

Cloud computing has evolved over traditional technologies. So, cloud computing can be mapped with older technologies before it has matured to the present level such as utility computing and grid computing.

- **Utility computing :** Around 1960, the processing or computing power prices were high for any purpose, so they came up

with the idea of sharing computing resources. Its goal is to integrate servers, storage systems, and applications distributed around the world to share with multiple users. Sharing would allow the users to use and share the computer resources and customers using the resources can pay for the services used for the period they have used services only [69].

- **Grid computing :** is the process of solving a massive computing problems by breaking it into smaller problems and solve these problems on simple or low-performance machines or computers to get the final result for large problems by distributing the tasks among various machines on the grid [70].
- **Cloud computing :** The concept of cloud computing is very similar to that of utility computing and grid computing. With the evolution of technologies over a couple of decades, cloud computing is possible to reach its goal in the last couple of years. In this decade, cloud computing has matured a lot in terms of technology to meet all its demands.

2.4 Development of Cloud Computing

In 1970, the development of cloud computing was started as a new trend, which has revolutionized the way of working and transformed traditional working environments. Following are important aspects of cloud computing development and adoption:

(1) **Advantages of using Cloud Computing**

Cloud computing simplified software development, business process, and accessing the services over the Internet. The traditional way of accessing services has changed with cloud computing. Adoption of the cloud has reduced costs, made an effective business model, and offers a great scale of flexibility for using the services. Many organizations have adopted cloud services and benefited by moving their services to the cloud. With the adoption of cloud, organizations are improving cross-platform collaboration between the developers, allowing them

to do more innovations on their IT capabilities, which in turn helps the organizations to grow their business and get more revenue [56].

(2) **Hybrid Cloud Computing**

Nowadays some of the organizations started using both private and public cloud services, for various reasons such as cross-platform evaluation, to check applicability in real time scenarios across multiple platforms, and so on. Usage of public cloud services even when they have private cloud services has lead to the development of hybrid cloud computing for compatibility or connectivity between different cloud computing services [57].

(3) **Mobile Cloud Services**

As the usage of mobile devices has increased in the last decade, data generated with individuals has also increased tremendously. To gain more customers, many of the cloud service providers have started supporting mobile devices for using or accessing cloud services. Using these mobile applications or interfaces customers can store their data or access the services provided by cloud service providers [51].

(4) **Cloud Security**

With the evolution of technology and cloud services many of the users are using the cloud services, but still, there is an essential problem in the Cloud which needs to be addressed, that is data security. There are many encryption techniques and security protocols to protect the data, but with the rapid growth in technology and processing power available to attackers, there is scope for new encryption techniques and security protocols for safe and secure future operations of cloud computing

(5) **Cloud Design**

Development of cloud services has rapid adoption of the services even in traditional markets. Cloud-centric markets have advantages over conventional markets because it is convenient for younger generation to use the services. To attract new

customers, cloud based service providers need not have physical presence with various stores in each city [58].

2.4.1 Characteristics

Cloud computing is a distribution of a massive computational power accessible over the Internet, rather than on local machines. Organizations with their private data centers also work on a similar principle. Cloud computing allows organizations to move their resources where they need more processing power for their applications instead of wasting the resources that are not utilized at their full potential.

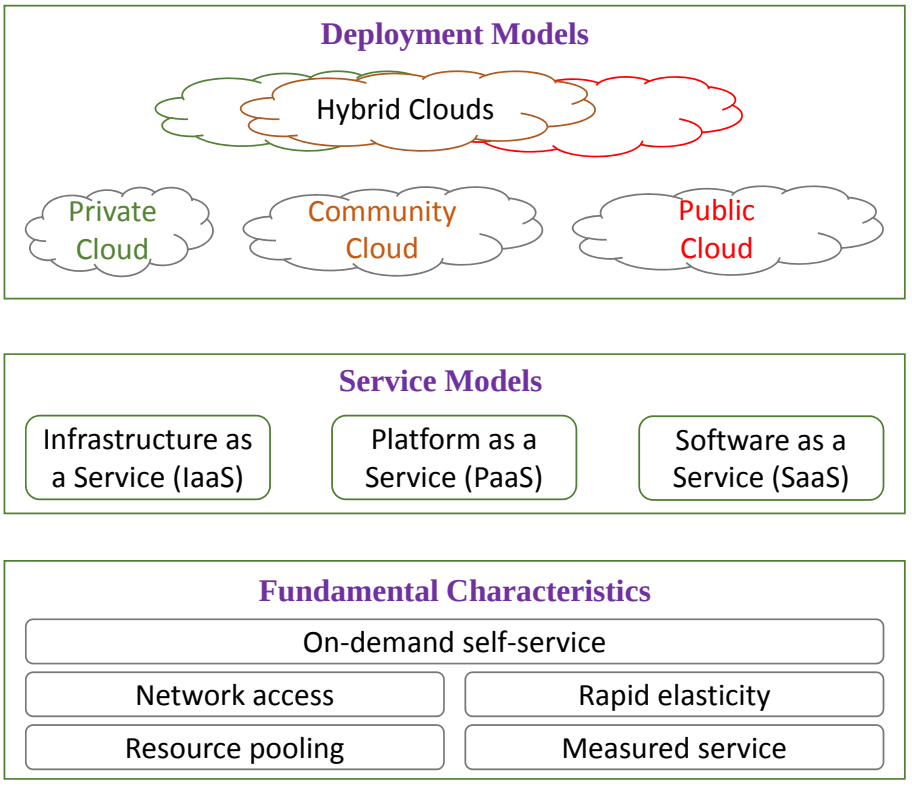

Figure 2.2: Cloud computing fundamental characteristics, service models, and deployment models

They are moving their traditional processing powers to centralized processing of data in their data centers. The shift allows computing or processing power as a commodity which can be traded over the Internet. Common characteristics of cloud computing are large scale, virtualized, low cost, geographically distributed, service oriented, resilient computing, and advanced security for services. Apart from above characteristics, cloud computing should also the following fundamental characteristics [52] as shown in Figure 2.2:

(1) **On-demand self-service :** End-users with the need to use the computing resources at a particular time (e.g., CPU time, network storage, software, and so on) automatically and conveniently, start and stop using them without any human interference.

(2) **Network access :** The computing resources delivered over the Internet, which can be used by various applications from different types of devices such as laptops, desktops, and mobile devices as per the end-user's requirement and availability.

(3) **Resource pooling :** Cloud service providers (groups) all of their computing resources together to serve multiple end-users using multi-tenancy model "with different physical and virtual resources dynamically assigned and reassigned according to consumer demand" [52]. End-user should be able to use resources irrespective of their location to support location independent resource pooling.

(4) **Physical transparency or Rapid elasticity :** End-users can change their resource capabilities automatically to scale up whenever they want to use more resources and release them once they are done using the services to scale down. For end-users, the resources are available for configuring with simple steps to scale up their operations and vice versa. These resources are not limited to end-users; they increase the usage of services to meet their peak requirements at any time.

(5) **Pay per-use or Measured Service :** Even though all the resources are pooled and shared among multiple tenants, the

end-users needs to be charged only for the cloud services they have used. This has to be taken care with a proper mechanism to measure the services used by each customer.

2.4.2 Service Models

Cloud computing can be segregated into the following service levels: Infrastructure as a Service, Platform as a Service, and Software as a Service as shown in Figure 2.2. Figure 2.3 presents the separation between service models with control of cloud service provider and customer of different underlying concepts in each model.

(1) **Infrastructure as a Service (IaaS) :** Customers will get the services for a complete computing infrastructure over the Internet. Example: Amazon EC2 [1] and S3 [2].

(2) **Platform as a Service (PaaS) :** In PaaS, customers will get the platform for the development of software applications. Example: Microsoft Azure [12] and Google AppEngine [7].

(3) **Software as a Service (SaaS) :** Customers will be provided with the Software over the Internet. In this model, users will not get the software; instead, they get the web-based software from the service providers for the intended work. Example: Dropbox [5] and Office365 [11].

2.4.3 Deployment Models

Cloud computing services are provided over the following deployment models [52] as shown in Figure 2.2:

(1) **Public cloud :** provides the cloud services for end-users by allowing them to access the services from the Internet. So, these cloud services are publicly accessible. The Cloud service provider provides the required infrastructure for end-users.

(2) **Private cloud :** model is used within the organizations to meet the cloud requirements across various levels in the organizations.

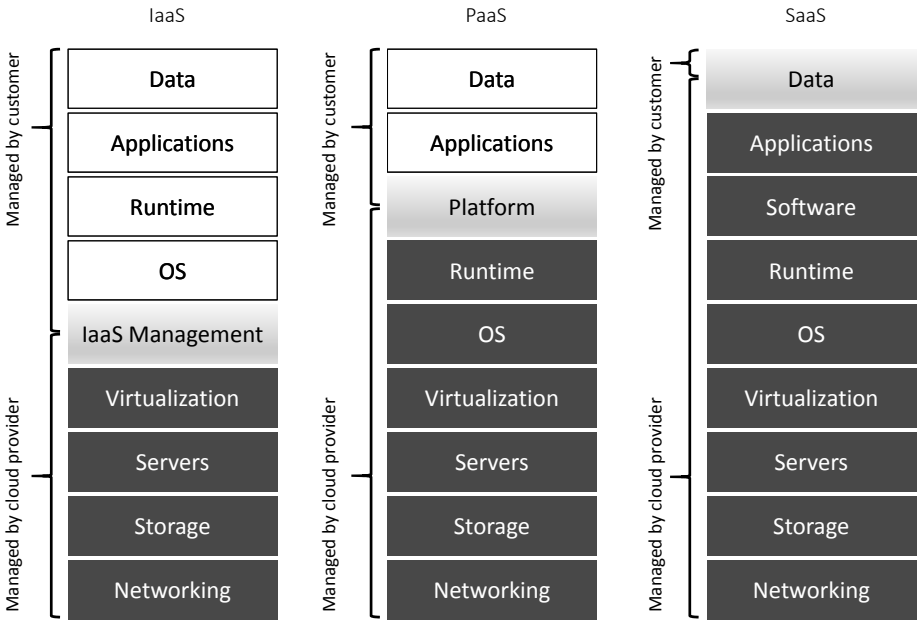

Figure 2.3: Various service models in the Cloud and control of services by cloud providers and customers

They will maintain their infrastructure to set up cloud services. This model will reduce the cost of using the cloud services for the organization in the long run with extra security as these private cloud models are deployed in their private networks behind their firewalls. Private cloud model can be accessed inside their network or by authenticating the user at their firewall.

(3) **Community cloud :** model is used to deploy the cloud infrastructure that can be shared between the several organizations with similar goals. It is comparable to a private cloud but shared among some organizations.

(4) **Hybrid cloud deployment model :** model is used where the customers use more than one model to meet the goals of their organizations or end-users them self.

2.5 Cloud Applications

Cloud computing development is directly or indirectly related to various applications using Cloud Services.

2.5.1 Software Development and Testing

Cloud computing has a lot of influences from software development and testing over the last couple of years or decades.

2.5.1.1 Software development

In the development of a cloud computing environment, software technology and software architecture have a lot of influence. Because of various reasons such as:

(1) The applications or software developed needs to be compatible with the cloud. As the cloud platform works in combination with several aspects such as computing platform/processing power, storage capacity, and architecture used in underlying deployment.

(2) The application should be able to serve a large user-base with huge amounts of data without any problems.

(3) These services must be provided over the Internet

(4) As the services are provided over the Internet, the risk of exposing confidential data is also high. So, cloud services need to have higher security for the application or services, which can stand against attacks, protect private information and data of their users.

(5) These services should be independent of platforms used by customers. i.e., users can use any device to access these services without any issues.

With the cloud computing environment, software development and the working environment changed a lot compared to traditional software development. Many of these changes can be attributed to cloud-based development tools, development platform, development environment, team collaboration, and remote working of various members in the group. Cloud has been used to deploy their own services online and check the services or software and evaluate them for the proper functioning of the services [66].

2.5.1.2 Software testing

With the adoption of a cloud computing environment for software development, software testing has some changes to cope with the new situation.

As discussed in section 2.5.1.1 with the adoption of the cloud computing environment for software development, has some changes in technology and architecture, so to meet these changes software testing also needs to be changed accordingly. Software testing should follow the traditional metrics and also adopt the changes to meet the requirements of a cloud computing environment such as dynamic capabilities, supporting a vast number of users, security, and cross-platform compatibility.

In the cloud computing environment, many of the things for software development changed such as tools, environment, and working patterns to meet the present environment. According to these changes software testing tools, environment, and working patterns should also change to meet the cloud environment. Testing tools need to map over the Cloud environment, unlike traditional methods. Software testing should also support collaboration, knowledge sharing, and test cases reuse in the cloud environment [60].

2.5.2 Cloud Storage

Cloud computing has been added with the extension cloud storage, used for storage of files or data over the network. Cloud combines

the software applications and storage space required for the proper functioning of it's services.

Cloud computing environment has processing capabilities; when this system is equipped to handle or manage large amounts of data by using storage devices, then the cloud computing environment can be treated as cloud storage. So, the cloud storage is the management of cloud computing environment with data storage management system [63].

2.5.3 Cloud Computing and Big Data

Cloud computing and Big data are two paradigms which cannot be separated as their relation is closely connected from a technical point of view. As the term Big Data clearly explain there is a massive amount of data which can not be processed on a single machine; instead, it needs a large system with tremendous processing power. Which can be done using distributed processing, distributed databases, and cloud storage or in other words cloud computing needs to be used. As cloud computing can provide the required amount of resources for processing the Big Data [56].

2.5.4 Gaming

In cloud-based gaming, all the games run on the server side and console from the client side will connect to the server over the Internet and communicate, get the data related to the game in real time. On the client side, there is no need for much processing power or high-end video capable devices except basic units to communicate with the server and receive the data over the Internet. With the adoption of new technologies like 5G mobile networks will make it possible to realize the cloud-based gaming solutions into real gaming solutions. Adoption of cloud-based gaming solution will be cost effective and save a lot of money for users in case they are switching between several games, as some of the architectures might not support some of the games [43].

2.5.5 Internet of Things (IoT)

The term "Internet of Things" (IoT) is coined by Kevin Ashton in 1999 [36]. The IoT is nothing but the things connected to the Internet. The basic block of IoT is the Internet, with an extended network based on the elements attached to it for exchange and communicate of information. K. Rose et al. defined, "The term Internet of Things generally refers to scenarios where network connectivity and computing capability extends to objects, sensors and everyday items not normally considered computers, allowing these devices to generate, exchange and consume data with minimal human intervention." [37].

As the adoption of IoT services increases, the demand for the data storage or processing power (computational power) also increases. These capabilities can be served with a cloud computing environment, as the cloud can handle large amounts of data with the required computational power to support any of the operations to perform on the collected data from IoT devices.

2.6 Cloud security

Cloud security has evolved with the adoption of Cloud Computing. The concept of cloud security is becoming more and more critical with the adoption of cloud services by more users. With many of the users around the world, if the cloud services are not protected adequately, it leaves vast amounts of customers' data vulnerable to attackers from all over the world.

Cloud security can be achieved in several forms, protection against the Network attacks, Software attacks, Intrusion Detection, Access control, analysis of abnormal behavior, analysis of Viruses, analysis of Malware, analysis of Trojans and so on. Security measures, which ensures cloud security are presented below:

(1) **Password :** To secure the cloud services from simple attacks against the access controls, users are encouraged to use a unique password for accessing the cloud-based services. Cus-

tomers should not use simple passwords or reuse the password which has been used on some other services over the Internet. Cloud service providers should make sure that there is no direct relation to user names and passwords stored in their database. In case there is a breach on the cloud service provider, it makes it hard for attackers to match the user names to passwords [47].

(2) **Access Recovery :** Customers should use confidential details or questions, for recovering their access control to the cloud in case they forgot their password. This information can be used to recover access to the Cloud. Users should not use the information which can be gained by using social engineering or just checking some information on their social networking profiles. As most of the personal details are posted on the networking websites. Using such information, attackers can easily gain access to the Cloud without knowing the person.

(3) **Encryption :** Using a good encryption technique by cloud service providers always protects the customer's data, such as Homomorphic Encryption. Usage of a homomorphic encryption technique is still not completely feasible in real time scenarios [34].

(4) **Password Management :** As discussed in the first point, users should not reuse their passwords, and cloud service providers should encourage them to use strong passwords with special characters, symbol, alphabets, and numbers. It is tough for users to remember all of their user names and passwords. They need to have a proper management tool for storing their user names and passwords to protect them from anyone getting access to them [68].

(5) **Multi-factor authentication :** Multi-factor authentication adds an extra layer of security to the traditional approach to access the Cloud services instead of user name and password. To access the cloud services using multi-factor authentication, customers need to have two or more factors to access the cloud services to authenticate them as a genuine user of the Cloud. These factors can be based on anything such as knowledge

(something known to the user, such as an other password), something user has (Biometric features), and something user possesses (RSA key or USB based keys or random text sent to their mobile) [67].

Cloud services providers should support multi-factor authentication methods and encourage the customers to use the Multi-factor authentication instead of using simple authentication using username and password. In this way, it will be easier to defend against unauthorized access to customers data even if someone has customers credentials; they won't be having access to other factors.

(6) **Login Monitor :** Cloud service providers and customers need to monitor recent devices used to access cloud services. Based on that information users can identify if someone has logged in with their credentials and change their passwords in case of a suspicious login from unknown devices or locations. Cloud service providers need to improve the login statistics with proper details for all the devices connected to access the Cloud Services for all the customers [18].

(7) **Personal Devices :** Customers should be careful where they are logging in to cloud to access the services. They should avoid using someone else's device, as they might have key loggers (a program which saves all the keys pressed on a device, while the program is running). In those devices, if they have such applications, attackers will gain user's credential for the Cloud compromising security for customers [48].

(8) **Virus, Malware, and Trojans :** Customers should have good anti-virus and anti-spyware applications on their devices. If they don't have proper protection of their devices, some viruses or malware might store the user credentials for cloud services and gain access to the cloud leaving their personal and confidential details into the hand of attackers. It would be a good habit for users to have good anti-virus and anti-spyware applications to protect their personal devices [65].

Cloud computing at present has matured a lot and solves many of the simple security aspects. Still, there are many open challenges which need to be addressed for more growth in Cloud computing Industry.

Chapter 3

Background

This chapter provides background or summary of the fundamental concepts that are required to understand the various methods presented in the thesis. It also covers the various related technologies or terminologies used in the enclosed papers, which aims to establish the basic knowledge required for this thesis.

3.1 Big Data

With rapid developments, the information technology field over the last two decades has increased the data in different areas at a faster rate. The exponential increase of the data with various organizations is leading them to use of the Cloud services to store the Big Data. One might think, What is Big Data? Big Data is a term which is usually used to call the large data sets which are collected from various sources while implementing, developing or using the services. It gets challenging to work with such data, as it can be in different forms.

Data can be treated as Big Data when it has the following characteristics:

(1) **Volume:** The data which is generated is very largescale.

(2) **Velocity:** The data can be coming from its sources at different speeds.

(3) **Variety:** The data can be a combination of different types of data.

(4) **Variability:** The data can be inconsistent and it is not that easy to process it.

(5) **Veracity:** The data can vary and might affect the actual analysis.

Adoption of cloud services for Big Data is increasing at a faster rate due to the rapid evolution of the Cloud services. We need to consider all the possible ways to store and secure the Big Data in the Cloud environment. These Big Data sets are mainly used for the analysis to find new correlations to help their own organizations. Many of the organizations are interested in utilizing the Big Data to improve their systems and analyzing the data from the users at real-time. Real-time data from the customers interaction with the services of their organization. If we look into some of the organizations working with large amounts of data such as eBay, Google and Facebook, each day these organizations will be collecting and analyzing petabyte (PB) of data to improve their services [41].

The increase in adoption of the mobile devices is in turn leading to the adoption of the many services over the cloud. There are billions of mobile devices around the world connected to the Internet and generating a large amount of data. All of these service providers are collecting extensive data from these users. It is estimated that the amount of information created in the last five years is more than the amount of data generated before [39]. Many of the organizations are using these data to analyze and understand their customers to improve their services and create new features or products based on the collective data analysis of their customers.

3.2 Data Science

Data Science is defined as "extraction of actionable knowledge directly from data through a process of discovery, or hypothesis formulation and hypothesis testing" by National Institute of Standards and Technology (NIST) [26]. Data science is a multidisciplinary area that expands the knowledge or perception of data from various sources using data extraction, data mining, statistics, predictive analysis, machine learning and deep leaning. Rapid growth in data creation, data storage, data processing, computing and cloud technologies in various domains retail, banking, health, industrial, private and public sectors makes the data science a crucial component in these fields.

3.3 Distributed Processing

Data analysis has become an integral part of the data science, which lead to adoption of distributed processing for faster and better results. In traditional systems, data used to be processed on a single machine. Whereas in distributed processing systems, data is managed across multiple devices across the Cloud environment. Distributed processing supports processing of large amounts of data with the processing power located on different networked devices at reduced cost for customers in real-time [55]. Analyzing the different types of data in distributed cloud environment fetches faster results. Distributed environments helps in storing, processing and analyzing the vast amounts of data at faster rates.

3.4 Cloud Simulator (CloudSim)

"Cloud simulator (CloudSim) provides a generalized and extensible simulation framework that enables modeling, simulation and experimentation of cloud computing infrastructure and application services". CloudSim was developed in the CLOUDS Laboratory at the com-

puter science and software engineering department of the University of Melbourne in Australia. It is an open source tool [50].

CloudSim framework includes several modules in its architecture, which helps a lot in the development of cloud computing algorithms. Figure 3.1 shows the interaction between the various modules in CloudSim. It supports the research and development of cloud computing and provides the following features [50]:

(1) It provides the characteristics of cloud computing.

(2) Supports the modeling and simulation of large-scale cloud computing infrastructure.

(3) It is a platform which supports the data centers, service agents, scheduling, resource management, resource monitoring and resource allocation strategies.

(4) It provides a virtualization engine to help establish and manage multiple virtualization services on data center nodes. It helps in virtualization based on data center and offers host-to-virtual machine mapping capabilities.

(5) It allows time sharing and space sharing in virtualized allocation services.

(6) Virtualized server hosts, with customizable policies for provisioning host resources to virtual machines.

(7) Energy-aware computational resources.

(8) Data center network topologies and message-passing applications.

(9) Supports for dynamic insertion of simulation elements, stop and resume simulations.

(10) Support for user-defined policies for allocation of hosts to virtual machines and policies for allocation of host resources to virtual machines.

Resources of a host machine in a data center can be mapped to several virtual machines, so there is a relationship between virtual machines

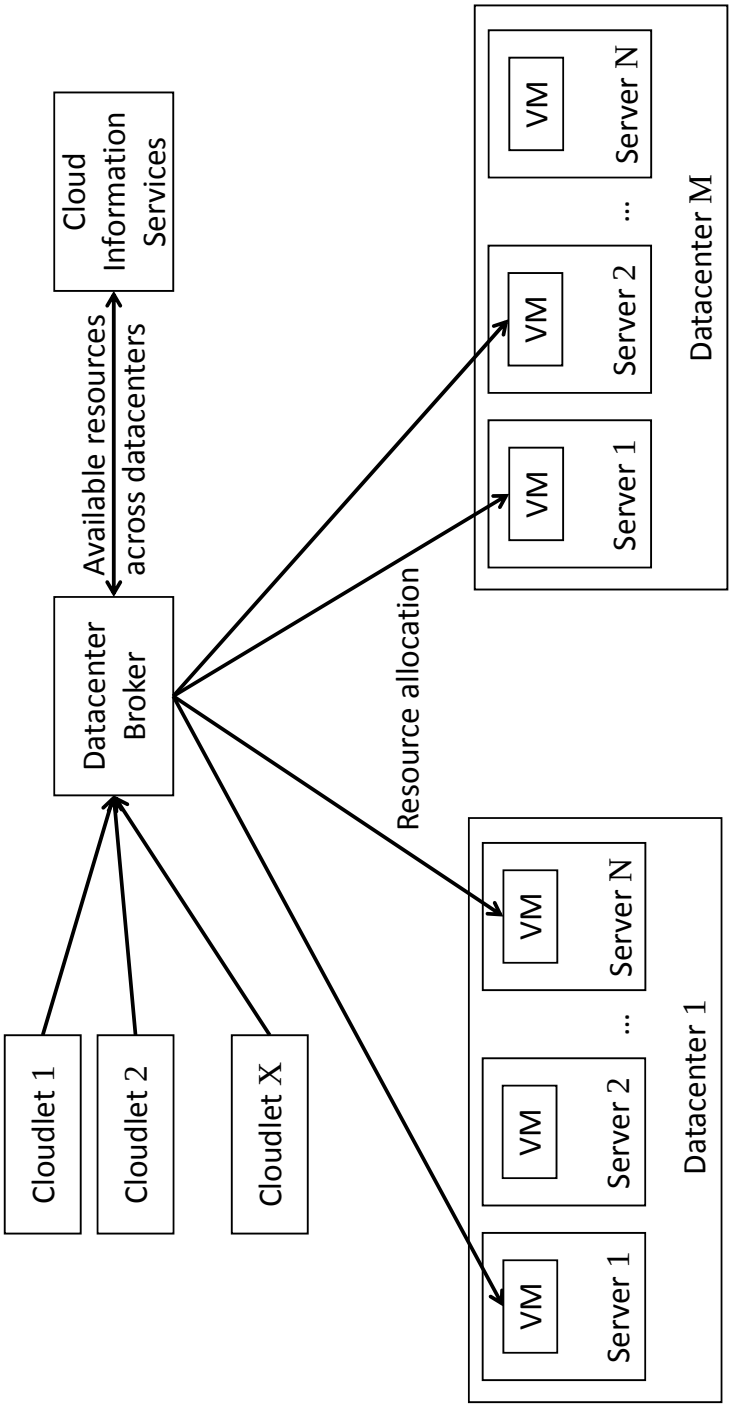

Figure 3.1: Interaction of various cloud simulator modules between different services [50]

and host resources. CloudSim's CIS (Cloud Information Service) and Data Center Broker realize resource discovery and information interaction, which is the core of scheduling simulation. The user-developed scheduling algorithm can be implemented in the Data Center Broker method, thus realizing the simulation of the scheduling algorithm.

Many organizations are using CloudSim for cloud resource provisioning, energy-efficient management of data center resources, optimization of cloud computing and research activities. Many of the experiments are not feasible in real cloud environments, so they are adopting cloud simulation tools [50, 49].

3.5 OpenStack

OpenStack is an open source Cloud computing management platform project developed by the National Aeronautics and Space Administration (NASA) and Rackspace. It combines several significant features to support Cloud computing [20]. The primary goal of the project is to provide a cloud computing management platform which is simple to implement, scalable, rich in features (several features) and standardized. OpenStack offers Infrastructure as a Service (IaaS) solutions through a variety of services with the help of APIs for integration. The OpenStack cloud computing platform helps the users to implement IaaS similar to Amazon EC2 and S3. OpenStack allows anyone to build and deliver Cloud computing services on their own or helps them to establish private cloud within the organizations or departments within the enterprise without exposing them to outside their firewall.

The initial design of OpenStack has two main modules: Nova and Swift. Nova is the virtual server deployment and business computing module developed by NASA. Swift is the distributed cloud storage module developed by Rackspace [20].

OpenStack covers all aspects of networking, virtualization, operating systems, servers and more. It is a cloud computing platform project

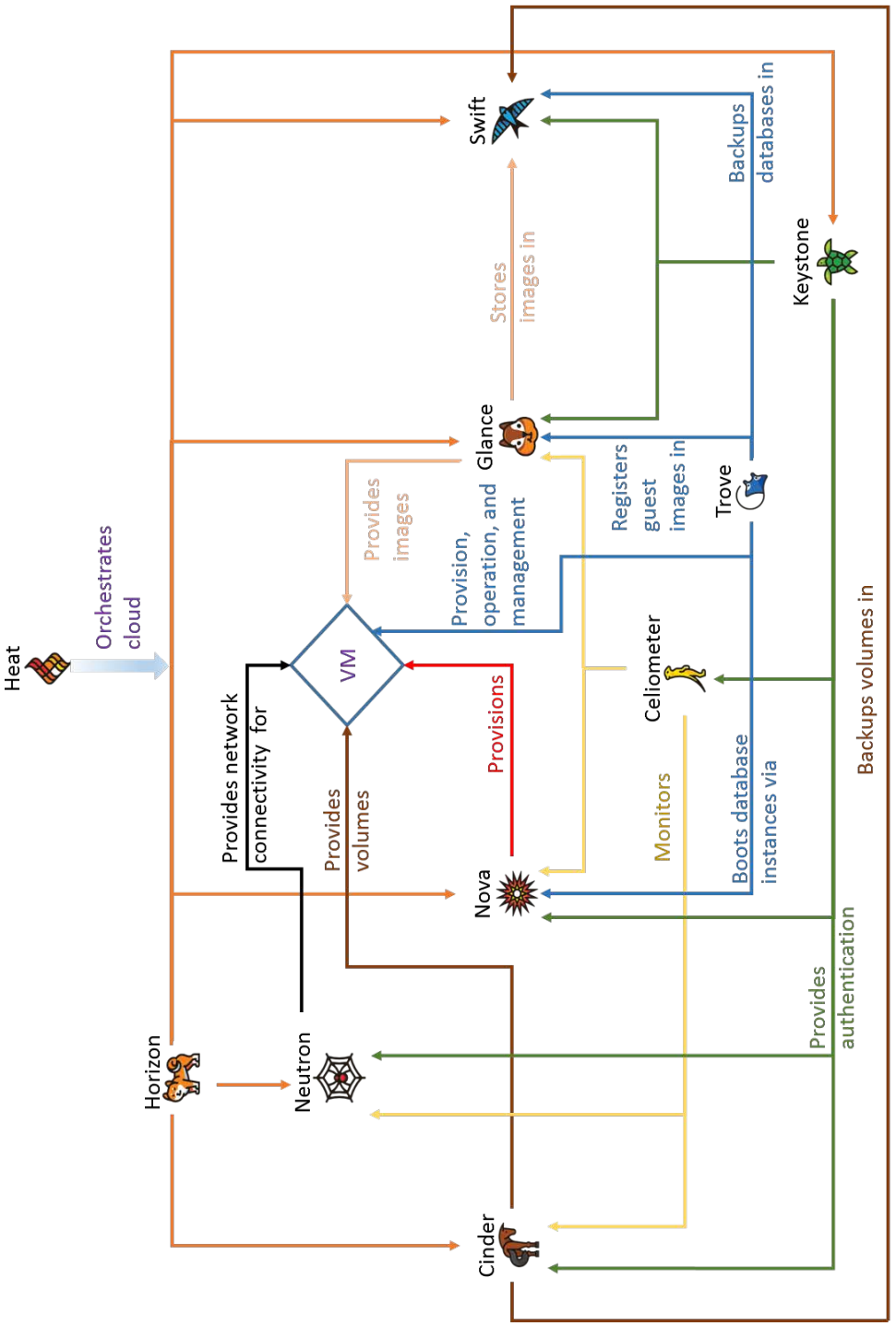

Figure 3.2: Conceptual architecture of OpenStack and interaction between different services [35] [15]

under development that is broken down into several small modules or projects. Not all modules are directly included or integrated until the module has matured to meet the important requirements they are planned for. Figure 3.2 shows the interaction between the various modules in OpenStack. Some of the main modules of the OpenStack are:

- **Compute Service (Nova):** A set of controllers that manage the entire lifecycle of a virtual machine (VM) instance for providing virtual services based on cloud deployer requirements. Nova is responsible for VM creation, boot, pause, adjust, migrate, restart, shutdown, suspend, destroy to CPU, memory and specification requirements [14].
- **Object Storage Service (Swift):** Swift project is designed for: A system for object storage with built-in redundancy and high fault tolerance in large-scale scalable systems, allowing storage or retrieval of files [21].
- **Image Service (Glance):** Image services are supported by Glance, which provides a set of virtual machine images for search and retrieval system, that supports several virtual machine image formats and supports the basic functionalities uploading new images, editing of existing images and deleting images [6].
- **Identity Service (Keystone):** Keystone supports the Identity services in OpenStack environment. It provides authentication, service rules and service tokens for other OpenStack services, managing Domains, Projects, Users, Groups and Roles [10].
- **Network Management Service (Neutron):** Neutron provides network connectivity services with virtualization in for other OpenStack services. It allows defining Network, Subnet and Router. It also enables to configure DHCP, DNS, load balancing and L3 services. It supports various type of networks depending on the underlying hardware [13].
- **Block Storage Service (Cinder):** Cinder provides a stable block storage service for running instances. Its plug-in-driven

architecture facilitates the creation and management of block devices, such as creating volumes, deleting volumes and mounting and unmounting volumes on instances [4].

- **Dashboard Service (Horizon):** Horizon provides a web management portal for various services in OpenStack to simplify the multiple operations for users in using the services with complete access controls in the user interface [8].
- **Metering Service (Ceilometer):** Ceilometer is service, which collects all the operations happening inside the OpenStack and provides the data support for monitoring and billing services [3].
- **Orchestration Service (Heat):** Heat provides a collaborative deployment for automatic deployment of cloud infrastructure (computing, storage and network resources) [17].
- **Database Service (Trove):** Trove provides scalable and reliable relational and non-relational database engine services for users in OpenStack environments [22].

Chapter 4

Contributions

The objective of this chapter is to provide contributions made with various research articles published with a summary of 6 published research articles included in this thesis. The summary of each article contains the motivation for the work, method or technique used, results and conclusions.

4.1 Overview

This thesis provides background for information processing in the Cloud for resource allocation, data recovery and data protection. Thesis work improves the security for the Information stored and processed in the Cloud computing. Three research questions are addressed in six peer-reviewed articles in various conferences.

In this dissertation, we addressed the need for new frameworks to solve various problems in the Cloud computing environment. Contributions of this thesis can be understood from the motivations discussed for each paper. Research carried out in this thesis can be divided into three main areas: resource allocation, data recovery and data protection. Data protection can further be divided in two sub categories: data protection while using cloud service and data protection post usage of cloud services. Paper I [32], addresses the

resource management and allocation across multiple cloud service providers. Papers II [28], III [31], IV [27], V [30] and VI [34] deals with data protection in the Cloud environment. In particular, papers II [28], III [31] and IV [27] tackles the data protection post usage of cloud services. Papers V [30] and VI [34] deal with data protection while using cloud services. Summary of the paper's contribution are listed below.

- Paper I [32] proposes a cost-effective solution for end-users to choose cloud service providers by taking into considerations various factors such as: memory, storage and processing power by using Nash Bargaining principles. This approach reduces the cost for customers by dynamically choosing the cloud service providers based on the requirements with better price from Cloud Service Providers.
- Paper II [28] introduces a framework to reconstruct the deleted data in the Cloud using the forensic tool using the memory analysis of the free or unallocated memory of underlying infrastructure in the Cloud.
- Papers II [28], III [31] and IV [27] demonstrate techniques which can protect the customer information in the Cloud after deletion to fight against data reconstruction techniques using forensic tools.
- Papers V [30] and VI [34] introduces approaches to protect the information in the Cloud while the customers are still using the Cloud services.

In the following sections of this chapter, we provide the motivation for carrying out the research for each paper, method used to solve the problem and conclusion based on the methods.

4.2 Paper I

Resource Allocation in Cloud-Based Distributed Cameras

This paper was published in IEEE International Congress on Big Data (BigData Congress), IEEE, 2017.

Motivation: Big Data from various sources of information needs to be stored, retrieved and analyzed. Such systems require a large amount of processing power and resources to analyze the data to get actionable intelligence. Cloud computing on the other side provides the storage required to save and share the data, computational power to process the data and deals with various services which can processes the big data. As processing, these huge amounts of data need to have large processing powers, which comes at the cost. To reduce the cost of analysis on such Big Data, we have considered one of the scenarios in cloud-based distributed cameras, where data collected from millions of public cameras needs to be stored and analyzed.

Method: We aim to provide a reliable collaborating experience for customers while using cloud computing resources for storage and processing of large amounts of data by bargaining theory to maximize the resources and minimize the cost for the end-user. We have used a resource allocation algorithm for analysis of large data from more than 70,000 public cameras. The resource manager finds cost-effective cloud service providers with cloud instances, which can be scaled as per the requirement.

The optimization problem discussed is one of the fundamental issues in bargaining theory. Cloud instances with memory, storage and processing power are its features, based on which bargaining has to be carried out. By using the Nash bargaining game theory among various cloud service providers, we try to maximize the total utility gains for the customers. To check the applicability of the Nash bargaining theory, we have used CAM2 server video data stream analysis.

Conclusion: This paper presents the results based on the simulation of cloud resource pricing and resource allocation between the various

cloud service providers using CloudSim [50]. To fit the real-time scenario, data centers across various locations, different processing powers and operating systems were considered for this environment. Results for resource handling were presented in terms of resource cost and execution rate. By applying the Nash Bargaining principle, data centers of cloud service providers handled more request compared to Greedy and Static approaches. Several experiments verified the efficiency of the proposed model using the Nash Bargaining principle. This approach can reduce the cost for customers by dynamically switching between the cloud service providers when the price from other cloud service provider is better than the present cloud service provider.

Comment: As the data gets distributed across various cloud service providers for data processing and analysis, customers start to lose control over their data. This raises security concerns for their data. These problems are addressed rest of the papers [28, 31, 27, 30, 34].

4.3 Paper II

Data Recovery and Security in Cloud

This paper was published in 9th International Conference on Information, Intelligence, Systems and Applications (IISA), IEEE, 2018.

Motivation: Evolution of the cloud is noticeable in the last couple of years with increasing popularity and rapid adoption of cloud services by customers and organizations. One of the main reason for this rapid growth is convenient to use cloud services from mobile devices over the Internet from anywhere for storage, backup and data analysis. Data recovery is one of the essential concepts while dealing with storage devices which are the backbone for the cloud infrastructure. Can someone with access to cloud infrastructure gain access to confidential customer data even after they have deleted their data from the Cloud? If yes, that leads to security and privacy issues for end-users. What can be done to solve these problems?

Method: In this paper [28], to check if it is possible to reconstruct the deleted data in the Cloud, we have used one of the forensic tools which works based on the memory analysis of underlying hardware for unallocated or free space to reconstruct the deleted information. Based on this analysis, we came to the point that, it is still possible to reconstruct the deleted data using forensic tools. To protect customers data against the data reconstruction techniques, we have proposed a method to solve data recovery problem. In this proposed technique, we would be renaming all of the existing file formats to some other format. Without knowing the new format of the data, it becomes hard for someone who tries to reconstruct the deleted data. Even if they recover the deleted information which is present in the new format, they will not get any actionable information without knowing the actual format.

Conclusion: By using one of the forensic tools based on the memory analysis, we were able to reconstruct the data deleted from the Cloud. So, it is possible to restore the data using data recovery tools. These data recovery techniques and tools raise the security concerns for the end-users even after deletion of the data from the Cloud, especially when someone who has access to cloud infrastructure. To fix this particular security issue in the cloud, we have proposed a framework by renaming the data type of the customer data to some other data types for protecting the data from memory reconstruction after deletion of the data.

Comment: Attackers will have all the time required to if they gain access to infrastructure by trying to recover all types of data. Once they realize that data format has been modified it will be easy for them to try get the information out of the deleted data. To further address this problem, we have tried different methods in papers III [31] and IV [27].

4.4 Paper III

Secure Customer Data over Cloud Forensic Reconstruction

This paper was published in IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2018.

Motivation: Adoption of cloud computing services has been increasing with an increase in the use of Internet of Things (IoT), which has been producing a large amount of data. This Big Data generated from the IoT devices are stored, retrieved and processed at a later point in time. Cloud computing can cope up with the increased demand for the Big Data processing, but does all the cloud service providers protect the customer's data with various security issues in the cloud? Once the processing of these Big Data sets is done by the end-user or organizations, can they be sure that this information has been completely deleted from the cloud service providers servers? The security concern for end-users raises a trust issue with cloud service providers. What can be done to secure the customer data, once they are done using it in the cloud?

Method: As the cloud service providers have access to the cloud infrastructure, they can use some forensic tools to reconstruct the data even after deleting the data from servers. So, how can we make sure that cloud service providers will not be able to rebuild the deleted data files in the Cloud? For this purpose, we have proposed a framework using the rewriting of the memory content of the end-users data, once they are done using the cloud services or data content. This method can be used for deleting the data while using the Cloud services or for removing the data once they are done with the cloud services. Rewrite methodology makes sure that random information is placed in all these memory locations rewriting the actual end-users data.

Conclusion: This paper presents a method to secure the end-users data in the Cloud, once they are done using their data. The proposed framework works similar to the forensic applications, which operates based on the memory reconstruction of the deleted files. We have intended to rewrite the complete data in the Cloud which needs to

be protected on all the memory locations of data stored on cloud service providers servers. The proposed method first identifies the files which need to be deleted in the Cloud. Then it will locate all the content on the Cloud servers. Once it has all the locations of the information, the rewrite module starts rewriting the data on all of those locations in the Cloud.

Comment: In this paper, we have addressed the problem associated with single cloud service provider. As discussed in Paper I [32] if the end-user has distributed their data across multiple cloud service providers. This problem is addressed in the paper IV [27].

4.5 Paper IV

Improve Security Over Multiple Cloud Service Providers for Resource Allocation

This paper was published in International Conference on Data Intelligence and Security (ICDIS), IEEE, 2018.

Motivation: Data processing has changed from traditional ways to the Cloud, where data from various devices or sources is stored, retrieved and analyzed. Processing of such large amounts of data requires huge processing powers, which is available at various cloud service providers for end-users. To reduce the cost of operations or distribute the tasks or segregate the operations on the data, end-users are distributing various tasks across the different cloud services providers. These servers are located across the globe and locations of these servers is not known to end-users in some of the cases. Most of the cloud service providers are providing high availability for their cloud services; they might have several backup servers to achieve high availability in case of a failure. What will happen to customers data once they have stopped using Cloud services? Customers data is completely deleted from all Cloud servers? Even if the data is deleted from all of their servers, can end-users be sure that, someone with access to the servers will not be able to reconstruct the deleted data using forensic applications?

Method: To solve these problems, we have proposed a framework to solve the problem of reconstructing the data from the deleted servers in case of resource allocation or resource sharing over multiple cloud service providers using forensic applications. The proposed framework has a new module known as User Shredder Module (USM). The primary role of USM is to interact with Collection of Cloud Service Providers (CCSP) to get the information related to the data which needs to be deleted and protect the data by rewriting those memory locations.

Conclusion: Resource allocation over multiple cloud service providers better solution to reduce the cost and distribute the task across the various cloud service providers. Once the end-users start distributing their data across multiple cloud service providers, security issues from different platforms come into the picture. Once they are done using these services, end-users might not have control over the data which has been stored and processed in multiple cloud service providers. The proposed framework uses the user shredder module to protect customer data across cloud service providers.

Comment: We have discussed enough for protecting the data once it has been deleted from the Cloud environment. In paper V [30] and VI [34], we have proposed frameworks to address the security challenges while the data is still stored in the Cloud.

4.6 Paper V

Method to Solve a Privacy and Security Issue in Cloud for Energy Informatics

This paper was published in Proceedings of the 13th World Congress on Engineering Asset Management (WCEAM 2018), Springer, 2018.

Motivation: Increased use of electronic devices, systems and appliances, usage of the electricity by end-users has been growing at rapid speed. To meet with the increased need for the power and meet the present requirements of end-users, electricity suppliers and producers

started analyzing the usage patterns of their customers to predict future demands. To achieve this, they are storing the end-users energy usage patterns. With the increased use of cloud services and rapid adoption for storage and processing, energy usage patterns of the end-users are stored in the Cloud. As storage and processing of the information is happening in the Cloud, it leads to the security and privacy issue for the end-users if someone gains access to the energy usage data. Energy consumption data can reveal when their household members are at home and when they are not present at home. Based on this information someone can target specific home to do something illegal or target them for personal reasons. End-users will lose privacy in their life and it raises security concern for them.

Method: To solve the security and privacy issues associated with the energy data stored in the cloud environment, we have proposed a method for storing the data in the Cloud. As energy consumption of all end-users is stored in the Cloud separately. Instead of saving the energy usage of each end-users separately, we have proposed aggregate the usage of electricity for several households or communities to form single data entry for all of them. Even if someone accesses this aggregated data stored in the Cloud, it will be hard for them to draw any conclusion of the usage patterns of a particular household preserving privacy for them. Doing so, it will not raise any security issues for the household members.

Conclusion: This paper presents a method to provide security and privacy for the end-users energy data stored in the cloud for various reasons by energy supply companies. The proposed method aggregates the energy consumption of multiple households. Aggregation of several household energy consumption patterns removes the individual household energy consumption patterns. But overall consumption patterns of these households will not be affected by this aggregation, which allows the energy supplying companies to effectively use the data for future predictions of these households effectively. The aggregated household consumption data can be applied with their predictions models without worrying about compromise in security or privacy for their customers.

Comment: This paper addresses specific scenario where data can be group. To address this problem all types of data, we have discussed the scenario to tackle the data protection in paper VI [34].

4.7 Paper VI

A Framework for Improving Security in Cloud Computing

This paper was published in IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), IEEE, 2017.

Motivation: Cloud computing has been evolving over a couple of years with the increased use of cloud-based services. Cloud services are developed using several older technologies to support and create new services. With new services in the Cloud have lead to widespread use in real time. With the increased processing power and efficiency, usage of cloud services has become cheaper and has reached many of the end-users. Adoption of cloud services has been raising several security issues for customers' data and also leads to ethical questions regarding how to protect the customer data and provide security without compromising their privacy. It is inevitable for small organizations to use cloud services at the cost of less secured cloud services. There are many open challenges in the cloud environment, which needs to be addressed. To solve these problems, we proposed a framework to solve ethical and security aspects related to cloud computing.

Method: In this paper [34], we have proposed a framework using the homomorphic encryption techniques, which allows the end-users to perform operations on the encrypted data (ciphertext) in the cloud without decrypting them into plain data. The operations performed on encrypted data leads to encrypted results. When these encrypted results are decrypted, the results will match the operations performed on plain data.

Conclusion: As cloud computing provides many advantages for end-users and organizations to support their goals, it also creates

security issues by storing confidential information in it. To solve these issues, cloud service providers need to improve their security measures to address the security issues in the Cloud. Even if the cloud service providers, provide better security for their services, but trusting the cloud service provider (CSP) depends on end-user. To improve the trust in CSP, they should adopt the proposed framework using the homomorphic encryption for solving data security problems in cloud environments. Adoption of the framework will resolve many security issues in the Cloud environment and encourage end-users to trust in CSP to use their services.

4.8 Research Questions

In this section, we present an analysis of all the research question, with an explanation for each question. Figure 4.1 presents the flow of research carried out in this thesis and their relationship to the published research papers (same Figure 1.2).

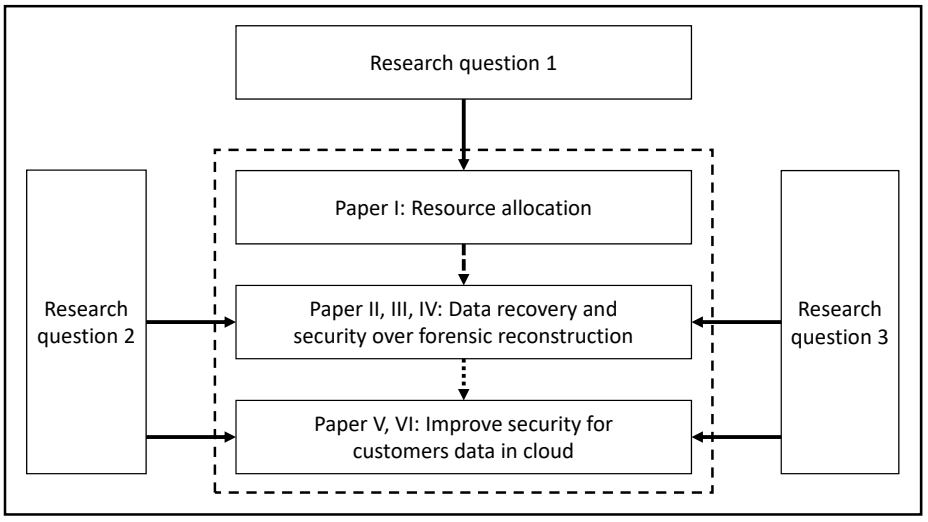

Figure 4.1: Flow of research objectives and their relationship (same Figure 1.2)

(1) Can resource allocation be improved by utilizing the multiple cloud service provider services?

Based on the results presented in Section 4.2 and Paper I [32], it is clearly visible that resource allocation can be improved by taking in to the consideration the optimization problem, in regards to the resources required (memory, storage and processing) and cost associated with them. Resource distribution across various cloud service providers will reduce the cost for the customers by dynamically switching between the cloud service providers when there is a change in quote from cloud service providers. Thus, resource allocation can be improved by using multiple cloud service provider services.

(2) Does the usage of cloud (for resource allocation) raises any other challenges?

To investigate this, we explored data recovery options in the cloud environment as discussed in Section 4.3 and Paper II [28]. Based on the results presented in Paper II, it is possible to reconstruct the information, even after the data has been deleted, using memory reconstruction tools. Reconstruction of the data leads to the security issue for the customer's data. It is evident that usage of the resource allocation will raise new challenges for customers. All the problems addressed in Papers II [28], III [31], IV [27], V [30] and VI [34] are also applicable in case of resource allocation and information processing in the Cloud environment.

(3) Whether identified challenges can be addressed?

In order to address these challenges, we have proposed several techniques in Papers II [28], III [31], IV [27], V [30] and VI [34]. Various aspects associated with these techniques are discussed in sections 4.3, 4.4, 4.5, 4.6 and 4.7. The challenges faced as part of the information processing in the cloud environment can be solved with adoption of proposed techniques. However, adoption of these new techniques will have extra costs for either customer or cloud service providers.

Chapter 5

Conclusion and Future Work

This chapter outlines the conclusion for research on information processing in the Cloud for resource management, data recovery and data protection. The following sections describe the conclusion and future work for this research.

5.1 Conclusion

This thesis proposes an extension to research work to improve the security for the Information stored and processed in the Cloud computing. Two parts of the thesis addressed with three research questions in 6 peer-reviewed articles in international conferences.

It provides a cost-effective and dynamic resource allocation method to handle huge amounts of data coming from several cameras distributed across various places, simulating the large datasets which are being processed in the Cloud Environments. We utilized the Nash Bargaining principles to effectively distribute the workload by considering cost-effective resource utilization over time. The cost-effective resource utilization is achieved by using the resource manager to determine the value based on workload, average waiting time and

user request are presented in Paper I [32]. As the information gets distributed across various cloud service providers, security for the information processed on the cloud might be at the risk.

In Paper II [28], we have looked in the possibility of reconstructing the data in the Cloud Environment using the forensic tools. We have presented the results showing that it is possible to reconstruct the data. In Paper II [28], III [31] and IV [27], we presented various methods to secure the data in the Cloud, once the processing of the information is completed. Method described in Paper II [28] can be applied in case of single cloud service provider or multiple cloud service providers. Paper III [31] talks more about securing the information in the Cloud and Paper IV presents the applicability of the data security across the multiple cloud service providers taking the scenario presented in Paper I [32]. Paper II [28], III [31] and IV [27] provide data security in the Cloud over forensic reconstruction. As the customers using the Cloud services, it is still possible to get the information while they still working on their data. We need new measures to protect the while processing the data in the Cloud Environment.

As the data stored in the Cloud can directly compromise the security and privacy of the individual. In Paper V [30], we analyzed the issue with electricity consumption patterns which are stored in the Cloud environment for load forecasting by electricity service providers. We have proposed a method using aggregation of data of several customers into single data entry, without losing the load data patterns of all customers. Still this method will not solve the complete problem of data security in Cloud. In Paper VI [34], we have proposed a framework using the Homomorphic Encryption to solve the security problems associated with storing data in the Cloud Environment.

5.2 Future Work

A possible extension of this work can be developing a common framework to secure the customer data in the Cloud over Forensic Reconstruction which can be applicable to multiple cloud service providers. This will be direct extension to the work presented in Paper II [28], III [31] and IV [27].

In future work, we would like to investigate the data governance models for Cloud computing Environments. The reason is that data governance provides more reliability and security for the data. This can be achieved with the help of reliable framework for data governance. Data governance allows to set proper controls to ensure the data is available to intended users, data will be consistent, will be accessible for actual users and will not allow access to unknown users.

We would like to develop Blockchain based system for information control in the Cloud environment, which could support sharing of information, control over their data, could trace the information, who has access to the data and who has accessed the information. With these controls Cloud users will be aware when their information has been leaked.

References

[1] **Amazon**. *Amazon Elastic Computing Cloud*. http://aws.amazon.com/ec2. [Online; accessed 31-Jan-2019].

[2] **Amazon**. *Amazon Web Services*. http://s3.amazonaws.com. [Online; accessed 31-Jan-2019].

[3] **Ceilometer**. *Welcome to Ceilometer's documentation!* https://docs.openstack.org/ceilometer/stein/index.html. [Online; accessed 31-July-2019].

[4] **Cinder**. *OpenStack Block Storage (Cinder) documentation*. https://docs.openstack.org/cinder/stein/index.html. [Online; accessed 31-July-2019].

[5] **INC Dropbox**. "Dropbox." In: *http://www.dropbox.com* (). [Online; accessed 31-Jan-2019].

[6] **Glance**. *Welcome to Glance's documentation!* https://docs.openstack.org/glance/stein/index.html. [Online; accessed 31-July-2019].

[7] **Google**. *Google App Engine*. http://code.google.com/appengine. [Online; accessed 31-Jan-2019].

[8] **Horizon**. *Horizon: The OpenStack Dashboard Project*. https://docs.openstack.org/horizon/stein/index.html. [Online; accessed 31-July-2019].

[9] **iDatalabs**. *Companies using OpenStack*. https://idatalabs.com/tech/products/openstack. [Online; accessed 31-Jan-2019].

[10] **Keystone**. *Keystone, the OpenStack Identity Service*. https://docs.openstack.org/keystone/stein/index.html. [Online; accessed 31-July-2019].

[11] **Microsoft**. "Office365: Documents and Outlook." In: (). [Online; accessed 31-Jan-2019].

[12] **Microsoft**. *Windows Azure*. http://www.microsoft.com/azure. [Online; accessed 31-Jan-2019].

[13] **Neutron**. *Welcome to Neutron's documentation!* `https://docs.openstack.org/neutron/stein/index.html`. [Online; accessed 31-July-2019].

[14] **Nova**. *OpenStack Compute (nova).* `https://docs.openstack.org/nova/stein/index.html`. [Online; accessed 31-July-2019].

[15] **OpenStack**. *Conceptual architecture.* `https://docs.openstack.org/ocata/install-guide-ubuntu/common/get-started-conceptual-architecture.html`. [Online; accessed 31-July-2019].

[16] **Openstack**. *Companies Supporting The OpenStack Foundation.* `https://www.openstack.org/foundation/companies/`. [Online; accessed 31-Jan-2019].

[17] **Openstack-Heat**. *Welcome to the Heat documentation!* `xxx`. [Online; accessed 31-July-2019].

[18] **ManageEngine ADAudit Plus**. *Real-Time Monitoring of User Logon Actions.* `https://www.manageengine.com/products/active-directory-audit/monitor-user-logon-actions.html`. [Online; accessed 31-Jan-2019].

[19] **Salesforce**. *Salesforce CRM.* `http://www.salesforce.com/platform`. [Online; accessed 31-Jan-2019].

[20] **Lauren Sell**. *OpenStack Launches as Independent Foundation, Begins Work Protecting, Empowering and Promoting OpenStack.* `https://www.businesswire.com/news/home/20120919005997/en/OpenStack-Launches-Independent-Foundation-Begins-Work-Protecting`. [Online; accessed 31-Jan-2019].

[21] **Swift**. *Welcome to Swift's documentation!* `https://docs.openstack.org/swift/stein/index.html`. [Online; accessed 31-July-2019].

[22] **Trove**. *Welcome to Trove's documentation!* `https://docs.openstack.org/trove/stein/index.html`. [Online; accessed 31-July-2019].

[23] **Weizhong Qiang, Weifeng Li, Hai Jin, and Jayachander Surbiryala**. "Mpchecker: Use-After-Free Vulnerabilities Protection Based on Multi-Level Pointers." In: *IEEE Access* 7 (2019), pp. 45961–45977.

[24] **Zeshun Shi, Huan Zhou, Yang Hu, Jayachander Surbiryala, Cees de Laat, and Zhiming Zhao**. "Operating Permissioned Blockchain in Clouds: A Performance Study of Hyperledger Sawtooth." In: *2019 18th International Symposium on Parallel and Distributed Computing (ISPDC)*. IEEE. 2019, pp. 50–57.

[25] **Jayachander Surbiryala and Chunming Rong**. "Cloud Computing: History and Overview." In: *The 3rd IEEE International Conference on Cloud and Fog Computing Technologies and Applications (IEEE Cloud Summit 2019)*. IEEE. 2019.

[26] **Wo L Chang, David Boyd, et al.** *NIST Big Data Interoperability Framework: Volume 6, Big Data Reference Architecture*. Tech. rep. 2018.

[27] **Jayachander Surbiryala, Bikash Agrawal, and Chunming Rong**. "Improve security over multiple cloud service providers for resource allocation." In: *2018 1st International Conference on Data Intelligence and Security (ICDIS)*. IEEE. 2018, pp. 145–148.

[28] **Jayachander Surbiryala and Chunming Rong**. "Data Recovery and Security in Cloud." In: *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE. 2018, pp. 1–5.

[29] **Jayachander Surbiryala and Chunming Rong**. "Data recovery in cloud using forensic tools." In: *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. IEEE. 2018, pp. 309–314.

[30] **Jayachander Surbiryala and Chunming Rong**. "Method to Solve a Privacy and Security Issue in Cloud for Energy Informatics." In: *in Proceedings of the 13th World Congress on Engineering Asset Management (WCEAM 2018)*. Springer, 2018.

[31] **Jayachander Surbiryala and Chunming Rong**. "Secure customer data over cloud forensic reconstruction." In: *2018 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE. 2018, pp. 1–4.

[32] **Bikash Agrawal, Jayachander Surbiryala, and Chunming Rong**. "Resource allocation in cloud-based distributed cameras." In: *2017 IEEE International Congress on Big Data (BigData Congress)*. IEEE. 2017, pp. 153–160.

[33] **Jayachander Surbiryala**. "PhD Forum: Improving the Security for Storing the Big Data in Cloud Environment." In: *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE. 2017, pp. 1–2.

[34] **Jayachander Surbiryala, Chunlei Li, and Chunming Rong**. "A framework for improving security in cloud computing." In: *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. IEEE. 2017, pp. 260–264.

[35] **M Ali Babar, David Silver, and Ben Ramsey**. "An Investigation into the Use of OpenStack for Submarine Mission." In: (2015).

[36] **Somayya Madakam, R Ramaswamy, and Siddharth Tripathi**. "Internet of Things (IoT): A literature review." In: *Journal of Computer and Communications* 3.05 (2015), p. 164.

[37] **Karen Rose, Scott Eldridge, and Lyman Chapin**. "The internet of things: An overview." In: *The Internet Society (ISOC)* (2015), pp. 1–50.

[38] **AT Saraswathi, YRA Kalaashri, and S Padmavathi**. "Dynamic resource allocation scheme in cloud computing." In: *Procedia Computer Science* 47 (2015), pp. 30–36.

[39] **Karan Singh**. *Learning Ceph*. Packt Publishing Ltd, 2015.

[40] **Rafael Xavier, Hendrik Moens, Bruno Volckaert, and Filip De Turck**. "Design and evaluation of elastic media resource allocation algorithms using CloudSim extensions." In: *Network and Service Management (CNSM), 2015 11th International Conference on*. IEEE. 2015, pp. 318–326.

[41] **Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani**. "Big data: survey, technologies, opportunities, and challenges." In: *The Scientific World Journal* 2014 (2014).

[42] **Yi Zhu, Yan Liang, Qiong Zhang, Xi Wang, Paparao Palacharla, and Motoyoshi Sekiya**. "Reliable resource allocation for optically interconnected distributed clouds." In: *Communications (ICC), 2014 IEEE International Conference on*. IEEE. 2014, pp. 3301–3306.

[43] **Ryan Shea, Jiangchuan Liu, Edith C-H Ngai, and Yong Cui**. "Cloud gaming: architecture and performance." In: *IEEE network* 27.4 (2013), pp. 16–21.

[44] **Zhen Xiao, Weijia Song, and Qi Chen**. "Dynamic resource allocation using virtual machines for cloud computing environment." In: *IEEE transactions on parallel and distributed systems* 24.6 (2013), pp. 1107–1117.

[45] **Mansoor Alicherry and TV Lakshman**. "Network aware resource allocation in distributed clouds." In: *INFOCOM, 2012 Proceedings IEEE*. IEEE. 2012, pp. 963–971.

[46] **Deyan Chen and Hong Zhao**. "Data security and privacy protection issues in cloud computing." In: *2012 International Conference on Computer Science and Electronics Engineering*. Vol. 1. IEEE. 2012, pp. 647–651.

[47] **Alexa Huth, Michael Orlando, and Linda Pesante**. "Password security, protection, and management." In: *United States Computer Emergency Readiness Team* (2012).

[48] **Keith W Miller, Jeffrey Voas, and George F Hurlburt**. "BYOD: Security and privacy considerations." In: *It Professional* 14.5 (2012), pp. 53–55.

[49] **Xiaoying Bai, Muyang Li, Bin Chen, Wei-Tek Tsai, and Jerry Gao**. "Cloud testing tools." In: *Proceedings of 2011 IEEE 6th International Symposium on Service Oriented System (SOSE)*. IEEE. 2011, pp. 1–12.

[50] **Rodrigo N Calheiros, Rajiv Ranjan, Anton Beloglazov, César AF De Rose, and Rajkumar Buyya**. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." In: *Software: Practice and experience* 41.1 (2011), pp. 23–50.

[51] **Dijiang Huang et al.** "Mobile cloud computing." In: *IEEE COMSOC Multimedia Communications Technical Committee (MMTC) E-Letter* 6.10 (2011), pp. 27–31.

[52] **Peter Mell, Tim Grance, et al.** "The NIST definition of cloud computing." In: (2011).

[53] **J Orea et al.** "Quick guide to the reference architecture: Trusted Cloud Initiative." In: *Cloud Security Alliance* (2011).

[54] **Subashini Subashini and Veeraruna Kavitha**. "A survey on security issues in service delivery models of cloud computing." In: *Journal of network and computer applications* 34.1 (2011), pp. 1–11.

[55] **Paul Zikopoulos, Chris Eaton, et al.** *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.

[56] **Borivoje Furht and Armando Escalante**. *Handbook of cloud computing*. Vol. 3. Springer, 2010.

[57] **Anthony D JoSEP, RAnDy KAtz, AnDy KonWinSKi, LEE Gunho, DAViD PAttERSon, and ARiEL RABKin**. "A view of cloud computing." In: *Communications of the ACM* 53.4 (2010).

[58] **Ali Khajeh-Hosseini, David Greenwood, and Ian Sommerville**. "Cloud migration: A case study of migrating an enterprise it system to iaas." In: *2010 IEEE 3rd International Conference on cloud computing*. IEEE. 2010, pp. 450–457.

[59] **Marian Mihailescu and Yong Meng Teo**. "Dynamic resource pricing on federated clouds." In: *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*. IEEE Computer Society. 2010, pp. 513–517.

[60] **Leah Muthoni Riungu, Ossi Taipale, and Kari Smolander**. "Research issues for software testing in the cloud." In: *2010 IEEE Second International Conference on Cloud Computing Technology and Science*. IEEE. 2010, pp. 557–564.

[61] **Nabil Sultan**. "Cloud computing for education: A new dawn?" In: *International Journal of Information Management* 30.2 (2010), pp. 109–116.

[62] **Anthony T Velte, Toby J Velte, Robert C Elsenpeter, and Robert C Elsenpeter**. *Cloud computing: a practical approach*. McGraw-Hill New York, 2010.

[63] **Jiyi Wu, Lingdi Ping, Xiaoping Ge, Ya Wang, and Jianqing Fu**. "Cloud storage as the infrastructure of cloud computing." In: *2010 International Conference on Intelligent Computing and Cognitive Informatics*. IEEE. 2010, pp. 380–383.

[64] **Sivadon Chaisiri, Bu-Sung Lee, and Dusit Niyato**. "Optimal virtual machine placement across multiple cloud providers." In: *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific*. IEEE. 2009, pp. 103–110.

[65] **Igor Muttik and Chris Barton**. "Cloud security technologies." In: *Information security technical report* 14.1 (2009), pp. 1–6.

[66] **Pavan Yara, Ramaseshan Ramachandran, Gayathri Balasubramanian, Karthik Muthuswamy, and Divya Chandrasekar**. "Global software development with cloud platforms." In: *International Conference on Software Engineering Approaches for Offshore and Outsourced Development*. Springer. 2009, pp. 81–95.

[67] **William N Owen and Eric Shoemaker**. *Multi-factor authentication system*. US Patent 7,373,515. 513 2008.

[68] **Shirley Gaw and Edward W Felten**. "Password management strategies for online accounts." In: *Proceedings of the second symposium on Usable privacy and security*. ACM. 2006, pp. 44–55.

[69] **Matthew Smith, Michael Engel, Thomas Friese, Bernd Freisleben, Gregory A Koenig, and William Yurcik**. "Security issues in on-demand grid and cluster computing." In: *Sixth IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06)*. Vol. 2. IEEE. 2006, 14–pp.

[70] **Miguel L Bote-Lorenzo, Yannis A Dimitriadis, and Eduardo Gómez-Sánchez**. "Grid characteristics and uses: a grid definition." In: *Grid Computing*. Springer. 2004, pp. 291–298.

[71] **Jeffrey O Kephart and David M Chess**. "The vision of autonomic computing." In: *Computer* 1 (2003), pp. 41–50.

[72] **Simson Garfinkel**. *Architects of the information society: 35 years of the Laboratory for Computer Science at MIT*. MIT press, 1999.

[73] **Christos H Papadimitriou**. "On the complexity of the parity argument and other inefficient proofs of existence." In: *Journal of Computer and system Sciences* 48.3 (1994), pp. 498–532.

Paper I: Resource Allocation in Cloud-Based Distributed Cameras

Resource Allocation in Cloud-Based Distributed Cameras

B. Agrawal[1], J. Surbiryala[2], C. Rong[2]

[1] Analytic Innovation Center (DNVGL)
Oslo, Norway
[2] Department of Electrical Engineering and Computer Science,
University of Stavanger, Stavanger, Norway

Abstract:
The data collected from millions of public cameras needs to be retrieved, stored, and analyzed. A system is required which needs to allocate significant amounts of resources to analyze large-scale visual data. Cloud computing provides shared storage, computation, and various services to handle such a tremendous amount of data collected from distributed cameras. In order to reduce the overall cost of analysis, this paper presents a resource allocation algorithm that provides cost-effective resources with the degree of demand; scaling automatically in proportion to demand fluctuation. In the cloud, the users prefer reliable resources at minimum cost whereas service providers prefer efficient resources utilization with maximum profit. Hence, it is necessary to have resource bargaining that assures and satisfies both cloud users and service providers. We propose a bargaining scheme using a game theoretic approach to managing cost and resource utilization in cloud-based distributed cameras. Our experiments show that our approach can lead to 10-15% reduction in cost by dynamically utilizing the resources and switching the service provider when it gets a better deal from other service providers.

1 Introduction

Millions of public cameras capturing events around the world, there is a need for a system to stored, analyzed and retrieve such massive amount of data. A simple analysis of the image (100 KB), every one minute from 70,000 cameras will generate 9TB of visual data per day. However, with the advent of the cloud computing paradigm, many individuals and organizations are moving towards cloud storage systems; this move is motivated by benefits such as shared storage, computation, and services transparently among a massive number of users.

Cloud service provider offers many instance types with different CPU, memory, storage, and network capabilities. Choosing the proper instance type for the analysis will be cost-effective. This paper addresses these issues to provide a cost-effective solution and maximum resource utilization for analysis of data from public cameras. The pay-per-use pricing model in the cloud is very appealing for both service providers and consumers, but conflicting objectives between the two parties hinder its effective application [11]. One of the major challenges in the cloud is the development of efficient resource management policies discussed in [2]. Resource demand in clouds occurs autonomously and unpredictably. Game theory is used in various application to solve diverse problems such as internet pricing, and congestion control [29]. One of the most widely used concepts in game theory is the Nash Bargaining theory [33].

In the past few years, many economic-based policies and resource allocation models have also been widely studied, including Resource Auction [7] [16], Proportional-Share [24], and First Price [30]. In many distributed and large systems, economic-based models have proven to be effective for resource allocation. However, it also raises other problems that cannot be ignored [25] [20] [14]. Firstly, most of the resource allocation based on economic models bring about extra communicational and computational overhead [28]; Secondly, most of the high-end applications which require a co-allocation of multiple resources and thus a process of price negotiation often results in low efficiency [21].

Currently, many existing cloud systems are obtainable on a fixed price based which has some advantages with low maintain cost and an easy implementation [21]. However, fixed pricing mechanism have some disadvantages such as negative effects on system performance with scale out, such as low resource utilization [18], load unbalancing [10], and a low QoS satisfaction level [15]. In this paper, we address the above issues and present a cloud resource allocation with pricing model using bargaining theory to overcome the shortcomings of existing resource allocation models regarding efficiency, fairness, and pricing. In our model, Virtual Machine (VM) resource configuration and provision are defined as cooperative gaming model to optimize the resource and price.

This paper presents resource manager for CAM2 [1] a case study to minimize the cost and maximize resource for analyzing the image and video streams from the network cameras. Our method allocates cloud instances based on the resource requirements of analysis and allocates the most cost effective cloud instances between various cloud service providers. It can also suggest switch between various cloud service providers based on their price scheme and geo-located datacenters.

Our experiments use two different analysis programs that represent different workloads in terms of Storage, CPU, and memory: (1) object count, and (2) human detection. Based on workloads our experiment is conducted on different cloud instances from various cloud service provider to provide a cost-optimized solution.

We propose efficient resource allocation using bargaining theory to optimize cost, and reduce average waiting time. The main contributions of the paper include: **(1)** cloud resource manager aiming to reduce analysis cost by allocating cost-effective cloud instances; **(2)** the resource manager is implemented on different cloud services together with a workload similar with CAM2 for image analysis; **(3)** recommendation of multiple of cloud service offers from different service providers, and **(4)** demonstration of Nash equilibrium in existence in the resource allocation game.

[1] https://cam2.ecn.purdue.edu/

The rest of the paper is organized as follows: Section 2 presents the related works. Section 3 presents details on our approach towards various problems. Section 4 gives an overview of the requirements of resource allocation, design and approach of our analysis. Section 5 evaluates our approach and presents the results & finally, Section 6 draws the conclusion.

2 Related Work

A survey paper on different networking games based on game theory can be found in [26]. The use of game theory for the resource allocation problem in cloud computing is discussed in many papers. Game-theoretic methods [5] [19] have been used to analyze cooperation in peer-to-peer networks and telecommunications. Nishida et. al. [22] attempt to make peers maintain a state by achieving a Nash equilibrium. The optimal virtual machine placement across multiple cloud providers [23] minimizes the cost of hosting VMs in a multiple service providers environment. However, this work is limited to static cloud scenarios.

In game theory, bargaining concept aims at bridging the gap between the users and service providers. Cooperative games are less popular in the cloud environment, but they have the power to allow users to make a decision from a set of decisions; it is possible to reach an equilibrium in polynomial time [31]. Game theory has been applied to study different issues in utility computing (grid and cloud computing).

Most of the resource allocation work on an allocation of resources within a single data center [2] [13]. Some cloud resources management approaches do focus on multi-cloud deployments [6] [9] [4] [21]. These approaches, however, do not focus on how VMs hosted in these distributed clouds can be connected to end users, which is critical for distributed cameras connected in different geo-locations. Most of dynamic resource allocation approach provides solution for slow run of systems [3] [8]. Most of this system tends to analyze the

characteristics of the job and prioritised jobs based on their execution priority, thus make other job slow. The problem with this approach does not scale with large amount of jobs with same proprity level.

However, none of these works considered the resource management for reducing the overall cost for analyzing streaming data. This paper proposes a resource manager for reducing the overall analysis cost in a cloud environment by allowing users to switch whenever they get a better deal from another cloud service provider.

3 Approach

In this paper, we aim to provide reliable collaboration experiences by the using bargaining theory to maximize the resource and minimize the cost. It presents a resource allocation algorithm for executing analysis of data from more than 70,000 public cameras. The resource manager allocates cost-effective cloud instances as presented in this section, and automatically scales the cloud resources.

Cloud service providers offer many cloud instance types with different capabilities regarding CPU, memory, network bandwidth, storage, and geographical location. This paper address some questions that arise, e.g.,

- How many resources are required for analysis?
- Can single cloud instance analyze multiple data streams without affecting another analysis running on cloud instance?
- What is the most cost-effective cloud instance and cloud service provider to use for analysis?
- How to reduce latency for collecting streaming data from different geo-located cameras?

The analytic program running in CAM2 can be as simple as object count or can be complicated as face detection. It is hard to have a prior estimate of the resource requirements of different analysis programs running in CAM2. The overview architecture of CAM2 is

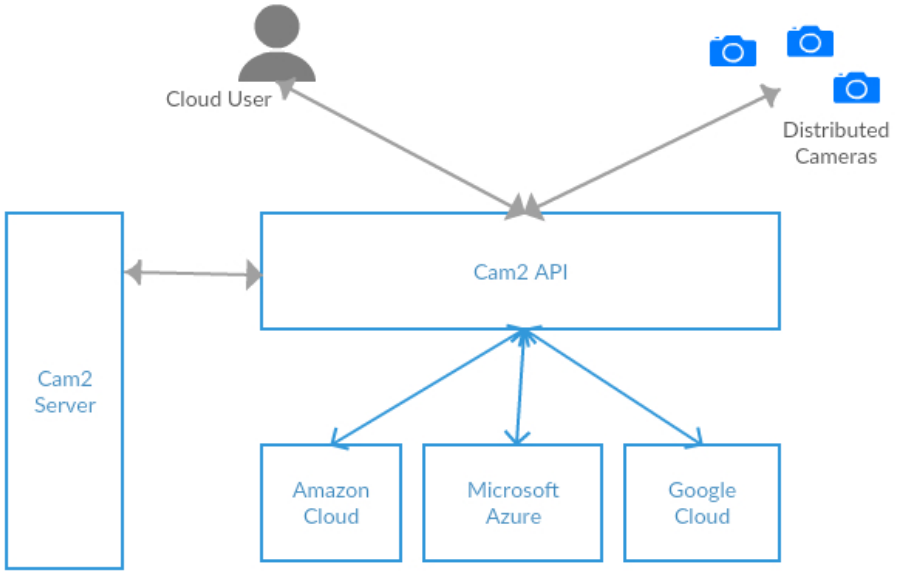

Figure 1: CAM2 interface with other cloud service for fair resource allocation.

presented in figure 1, to show simple user interaction with CAM2 API for performing analysis in the cloud.

To answer resource allocation problem, we need to estimate the resource requirement for running analysis program. Lets us consider, i as cloud instance and a as analysis program for analysis data stream. The CPU utilization per camera is denoted by $CPU_{i,a}$ and memory utilization is denoted by $Mem_{i,a}$. We consider constant frame rate f for all the cameras in our analysis for this research. In real-time, different cameras produce different frame rate for actual analysis. From many experiments, it shows that maintaining 90% threshold range for CPU and memory utilization leads to meet the performance requirements without comprising any latency and QoS [1] [17]. If the threshold value of CPU and memory utilization is denoted by CPU_H and Mem_H respectively, then maximum number of streams a cloud instance i can handle is estimated as:

$$N_{i,a} = min(\frac{CPU_H}{CPU_{i,a}}, \frac{Mem_H}{Mem_{i,a}}) \tag{5.1}$$

Now, for calculating an effective cost for running analysis program in different cloud instances. The most effective cloud instance is one which can minimize the cost by running multiple analysis programs.

$$i^e = \underset{i}{\operatorname{argmin}} \frac{C_i}{N_{i,a}} \tag{5.2}$$

where, C_i is the hourly cost of an instance type i and i^e is the cost-effective cloud instance. The total cloud instance needed for analysing the data from N cameras is $\frac{N}{N_{i^e,a}}$, and the total cost for overall analysis would be $[\frac{N}{N_{i^e,a}}]C_i$. Our proposed resource allocation scheme will select the cost-effective cloud instances for analyzing the data from cameras from different cloud service providers using bargaining approach.

4 Resource Allocation Model

The paper focuses on minimizing the cloud instances cost by bargaining with different cloud service providers as well as allocating resource to run multiple analysis programs without exceeding 90% threshold level. In this section, we provide resource allocation algorithm to achieve a cost-effective solution.

4.1 Bargaining Problem

In cooperative games, players are assumed to try reaching an agreement that gives mutual advantage. There are n users. Each user i has their own utility function $(U_i(x_i))$ for the allocated resource x_i and it has a minimum desired utility $(U_i(R_{0i}))$, called the disagreement point. The disagreement point is the point when both the users do not come to the agreement. Assume $\mathbf{S} = (U_1(x_1), U_2(x_2), ..., U_n(X_n)) \subset \mathbb{R}^n$ is a joint utility set(or a feasible utility set) that is nonempty, convex, closed, and bounded and let $\mathbf{D} = (d_1, d_2, ..., d_n) = (U_1(R_{01}), U_2(R_{02}), ..., U_n(R_{0n})) \in \mathbb{R}^n$be the disagreement point. The pair **(S,D)** defines

the bargaining problem. The Nash Bargaining theory gives a unique and fair Pareto optimal solutions.

4.2 Problem Definition

Objective: Minimizing the cost and latency and at the same time maximizing the resource utilization.

Table 5.1: Notations

Notation	Description
G	the number of geographically distributed data centers
P	the set of physical machines
VM_j	the set of VMs on the physical machine j
U_i	the user request application i at time t
a_i	the application by user i at time t
n	the total number of user/application request
T	the time slots of equal length
c_{ir}	the CPU processing capability on VM v_r for P_i
m_{ir}	the memory requirement on VM v_r for P_i
s_{ir}	the storage requirement on VM v_r for P_i
R_{max}	the total resource on a physical machine
$\lambda_i(t)$	indicate whether VM is active at time t ($\lambda_i(t) = 1$) or not ($\lambda_i(t) = 0$)
$\overline{C}$	total cost over time T

In this paper, we consider the system consists of n heterogeneous physical machines, represented as $P(t) = \{P_1, P_2, ..., P_n\}$ and $VM(t) = \{v_1, v_2, ..., v_r\}$ be set of VMs at time t. Let the minimum resource requirement for VMs on P_i: minimum CPU requirement on P_i be c_{ir}, memory requirement be m_{ir} and storage requirement be s_{ir}. Let $J(t) = \{j_1, j_2, ..., j_r\}$ be set of assigned jobs at time t. Key notations used in our research are listed in Table 5.1.

Problem 1: Total computation cost: Every task has a budget $b(j_r)$ for executing the job j_r within time duration t. Let $\alpha_{ir}, \beta_{ir}, \gamma_{ir}$ be unit cost per CPU, memory and storage respectively for virtual machine v_r. Now, for any VM v_r, the maximum amount of any resource (CPU, memory, and storage resource) it may receive in P_i at any time t is given as:

$$\begin{aligned} cpu : R_{max}^{c_{ir}}(t) &= \frac{b(j_r)}{\alpha_{ir}} \\ memory : R_{max}^{m_{ir}}(t) &= \frac{b(j_r)}{\beta_{ir}} \\ storage : R_{max}^{s_{ir}}(t) &= \frac{b(j_r)}{\gamma_{ir}} \end{aligned} \tag{5.3}$$

The total cost can be computed by summing all the running systems at time T. $\lambda_i(t)$ indicates whether VM is active at time t ($\lambda_i(t) = 1$) or not ($\lambda_i(t) = 0$).

$$\begin{aligned} &\text{minimize} && \overline{C} = \sum_{i=1}^{n}\sum_{t=0}^{T}\frac{\lambda_i(t)}{t} \\ &\text{subject to} \\ &&& \sum_{i=1}^{n} V_{ir}(t)R_{max}^{c_{ir}}(t) \leqslant CPU_i \\ &&& \sum_{i=1}^{n} V_{ir}(t)R_{max}^{m_{ir}}(t) \leqslant M_i \\ &&& \sum_{i=1}^{n} V_{ir}(t)R_{max}^{s_{ir}}(t) \leqslant S_i, \quad \forall_{i,t} \leqslant T(J_r) \end{aligned} \tag{5.4}$$

where, $V_{ir}(t) = 1$ means VM v_r is active and has been allocated on P_i at time t. The cost $\overline{C}$ is sum of the times taken for all jobs. If the resources are under-utilized, the cost $\overline{C}$ will be increased for a cloud-based datacenter. It is very important to maximize the resources utilization in the servers, so that overall cost can be minimized. So now, the optimization goal is to minimize the cost.

Problem 2: The main objective is to maximize the resource utilization with respect to CPU, memory and storage.

$$\begin{aligned} \text{maximize} \quad R = \sum_{i=1}^{n}\sum_{r=1}^{r}\sum_{t=0}^{T} V_{ir}(t) * [&\frac{R_{max}^{c_{ir}}(t) * w_i^C(t)}{CPU_i} + \\ &\frac{R_{max}^{m_{ir}}(t) * w_i^M(t)}{M_i} + \frac{R_{max}^{s_{ir}}(t) * w_i^S(t)}{S_i}] \\ &\text{subject to } V_{ir}(t), t \leqslant T(J_r) \end{aligned} \tag{5.5}$$

where, R is the estimated resource utilization at server i in time t. w_i^C,w_i^M and w_i^S are the weights given to resources and total weight of resources is always 1; i.e., $w_i^C(t) + w_i^M(t) + w_i^S(t) = 1$. Now our goal is to find out the best placement strategy of each VM, so that resource utilization is maximized.

Let $H_i(t)$ be the utility function for server i at time t. Each server will seek to maximize the Nash product $\prod R_i(t)$ at time t, where

$R_i(t) = H_i(q) - H_i(d)$ to attain the Pareto-optimal solution for the bargaining situation. Utility function of server i is given as:

$H_i(q) = H_i(d) + e^{\sum_{i=1}^{n} \sum_{r=1}^{r} \sum_{t=0}^{T} V_{ir}(t) * E_{ir}(t)}$

The utility gain of the server i is:

$$R_i(t) = |H_i(q) - H_i(d)| = e^{\sum_{i=1}^{n} \sum_{r=1}^{r} \sum_{t=0}^{T} V_{ir}(t) * E_{ir}(t)} \tag{5.6}$$

Now each server i will try to maximize the Nash product $R_i(t)$ which can be interpreted as follows:

$$max \prod_{i=1,t=0}^{n,T} R_i(t) \tag{5.7}$$

From equation 5.6 and 5.7, we get:

$$max \sum_{i=1}^{n} \sum_{r=1}^{r} \sum_{t=0}^{T} V_{ir}(t) * E_{ir}(t) \tag{5.8}$$

Now, for each server i we can define anticipated resource $E_{ir}(t)$ as:

$$E_{ir}(t) = \frac{R_{max}^{c_{ir}}(t) * w_i^C(t)}{CPU_i} + \frac{R_{max}^{m_{ir}}(t) * w_i^M(t)}{M_i} + \frac{R_{max}^{s_{ir}}(t) * w_i^S(t)}{S_i} \tag{5.9}$$

Now, from equation 5.8 and 5.9, we get:

$$max_{t<T} \sum_{i=1}^{n} \sum_{r=1}^{r} \sum_{t=0}^{T} V_{ir}(t) * [\frac{R_{max}^{c_{ir}}(t) * w_i^C(t)}{CPU_i} + \frac{R_{max}^{m_{ir}}(t) * w_i^M(t)}{M_i} + \frac{R_{max}^{s_{ir}}(t) * w_i^S(t)}{S_i}] \tag{5.10}$$

The optimization variable $V_{ir}(t)$ is an indicator of the strategy adopted by each player i in the bargaining game. The estimated resource utilization ratio of server i can be treated as the profit gains of player i by adopting strategy $V_{ir}(t)$.

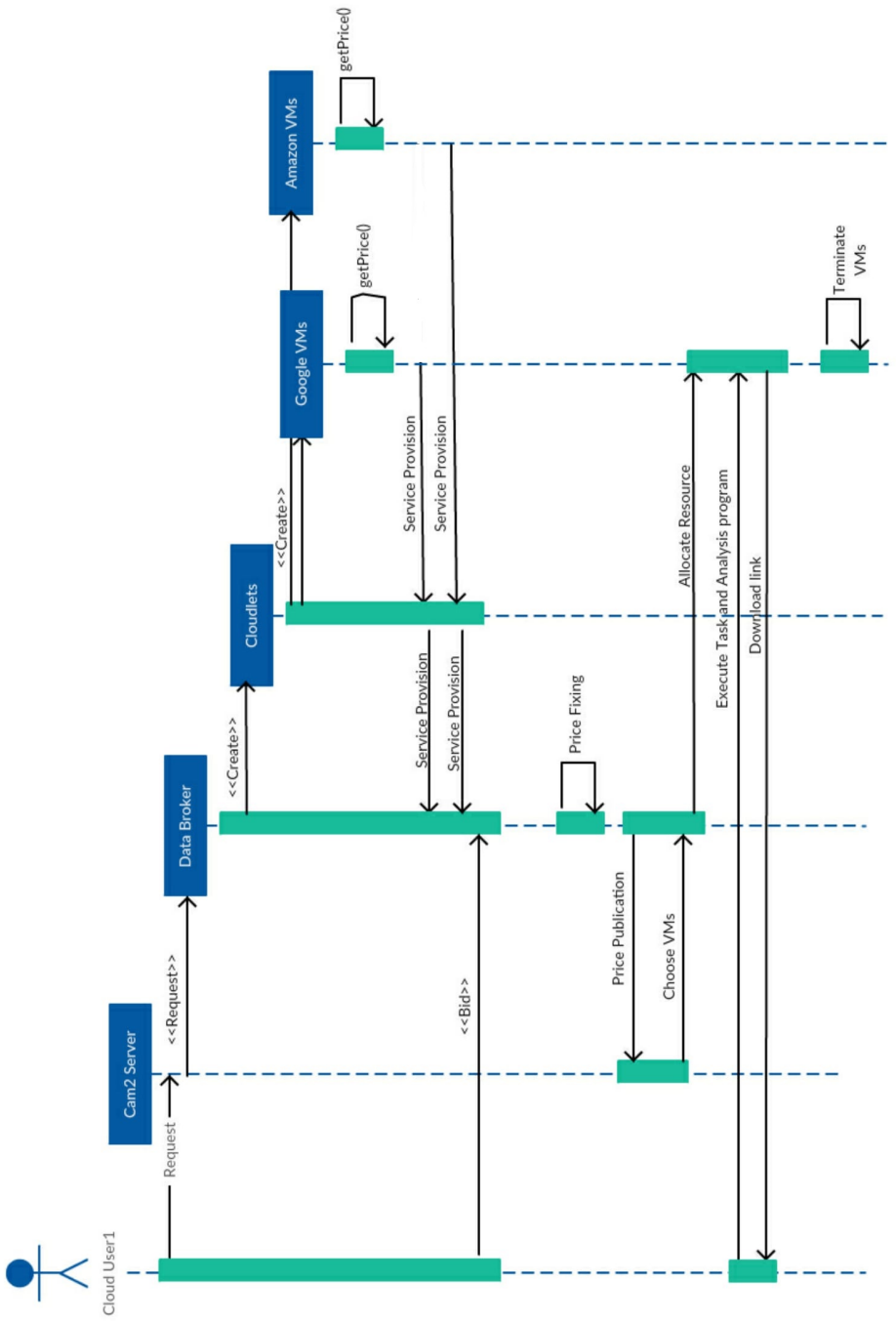

Figure 2: Sequence diagram for program flow in cloudsim.

4.3 Joint optimal solution

We consider to optimize the cost and resource utilization jointly. To expose the tradeoff between two conflicting objectives, we define our objective function as the simple weighted sum of cost $\overline{C}$ and the resource utilization R.

$$\begin{aligned} &\text{minimize} \quad \overline{C} + \alpha.(R_{max} - R) \\ &\text{subject to} \quad V_{ir}(t), t \leqslant T(j_r) \end{aligned} \tag{5.11}$$

By varying the value of α in the interval $[0, \infty)$ and normalizing $\overline{C}$ and R to the interval $[0, 1]$, we can plot a typical Pareto curve for the joint optimization problem referred to equation 5.11. To solve the joint optimization problem, we adopt Nash bargaining solution [32], which ensure optimality and fairness simultaneously. In bargaining theory, we consider two independent players with conflicting objectives who are trying to cooperate with each other to achieve an optimal point. We can find optimal point of $\overline{C} + R$ on the Pareto curve. Using Nash bargaining solution, we can express equation 5.11 as:

$$\begin{aligned} &\text{maximize} \quad |\overline{C} - C_0|.|R_0 - R)| \\ &\text{subject to} \quad V_{ir}(t), t \leqslant T(j_r) \end{aligned} \tag{5.12}$$

where, (C_0, R_0) is the disagreement point that represents the starting point of the bargaining. C_0 denote as the minimum total cost, and R_0 as maximum resource utilization. Applying optimization decomposition theory [27], we can derive an efficient solution. The equation 5.12 can be written as:

$$\text{maximize} \quad log(\overline{C} - C_0) + log(R_0 - R) \tag{5.13}$$

4.4 Proposed resource allocation algorithm

The optimization problem we listed above is a bargaining problem. Each VM is represented by a tuple containing its dimensions (i.e.

CPU, memory, and storage). The Nash bargaining game theory discusses how two or more players distribute commodities among them so that their total utility gains are maximized. In our scenarios, each cloud instances will seek to maximize the Nash product $\prod G_k(t)$ at time t.

In figure 2, cloud user requests for resources for analysis video data stream to CAM2 server. CAM2 starts the resource bargaining by calculating price from different service providers and also stating the price it is expected to pay, expected starting time, required resource (R), and a total number of jobs to be submitted. The service provider initially bargains with for execution time of the job. Now, it is up to cloud user who needs to decide how fast the job needs to be executed. Cloud user sends its bid to the service provider, and finally, CAM2 server decides how much resource needs to be allocated and which cloud instance the program needs to run. After analysis of required resources, a download link is sent to the user via CAM2 server.

5 Result

5.1 Data Collection:

We present the simulation results based on cloud resource pricing and resource allocation using CloudSim [12]. To model real scenario of cloud datacenter, we model CPU with different MIPS, memory with different size, and storage with various data plans (in Gigabyte - GB). At the same time variation of different CPU, operating system, and data center location were considered when we initialize heterogeneous computing environment. To model different cloud users, different kind of analysis applications were considered in CAM2 server. We have chosen 200 users to simulate the actual load in cloudsim and to evaluate different set of tasks. All the users bid accordingly to their budget and deadline constraints.

The VM costs were based on the Amazon EC2, Google compute platform, and Microsoft Azure images with an hour price to provide

Table 5.2: The cost of cloud instances with different geo-located datacenter.

Type	Storage	CPU	Memory	Cost (US)/hr	Cost(EU)/hr	Cost (Asia)/hr
Amazon Ec2						
m3.large	32 SSD	2	7.5	$0.133	$0.146	$0.196
m3.xlarge	80 SSD	4	15	$0.266	$0.293	$0.392
m3.2xlarge	160 SSD	8	30	$0.532	$0.585	$0.784
c3.4xlarge	2 x 160 SSD	16	30	$0.84	$0.956	$1.058
c3.8xlarge	2 x 320 SSD	32	60	$1.68	$1.912	$02.117
Azure						
D2	100 SSD	2	7	$0.188	$0.168	$0.225
D3	200 SSD	4	14	$0.376	$0.336	$0.451
D4	400 SSD	8	28	$0.752	$0.672	$0.902
D11	100 SSD	2	14	$0.238	$0.224	$0.242
D12	200 SSD	4	28	$0.476	$0.449	$0.483
Google						
n1-standard-2	375 GB	2	7.5	$0.070	$0.078	$0.078
n1-standard-4	375 GB	4	15	$0.140	$0.156	$0.156
n1-standard-8	375 GB	8	30	$0.280	$0.312	$0.312
n1-standard-16	375 GB	16	60	$0.560	$0.624	$0.624
n1-standard-32	375 GB	32	120	$1.120	$1.248	$1.248

CPU, memory, and storage as listed in table 5.2. We used different cloud services for storage and computation. For the experiment, we used all these cloud platforms for simulation based on pricing model and geo-location of the data center. The simulation parameter of different VMs instances pricing and resource is configured according to the pricing scheme listed in table 5.2. For our experiment to generate real scenario, we extract pricing model from different cloud services. To generate the workload of CAM2, we consider the user requests for analysis in CAM2 follows a Poisson distribution with arrival rate, θ. The deadline for each task is fixed and set to 1 hour. Consequently, 1000 VMs containing tasks can be requested randomly at any time t. All the VMs consist of same amount of resources: 1000 MIPS CPU cycles., 160 GB storage space, 1000 Mbps bandwith and 2000 MB/S disk I/O speed. The maximum requirement id determined on the basis of level of budget that users have randomly set for each task.

5.2 Data Analysis:

5.2.1 Resource Cost

Our primary objective is to minimize the running cost of cloud instances by maximizing the utilization of resources. Also, our approach to select cloud instances based on the price of different geo-located data centers as analysis is performed on distributed cameras. We present the average resource cost for various task arrived at CAM2 server. Figure 3 shows that the proposed Nash Bargaining algorithm provides a cost-effective solution for allocating cloud instances compare with Greedy and static allocation algorithm. The Greedy algorithm works well when the request has less resource requirement, as the user requirement increases the cost of allocating resource also increases. Whereas, our approach with Nash method can handle a large number of requests efficiently by minimizing the cost. The main reason for this behaviour is the utilization of bargaining concept for handling more number of resources. Our approach makes a dynamic request to the cloud providers for a price based on demand,

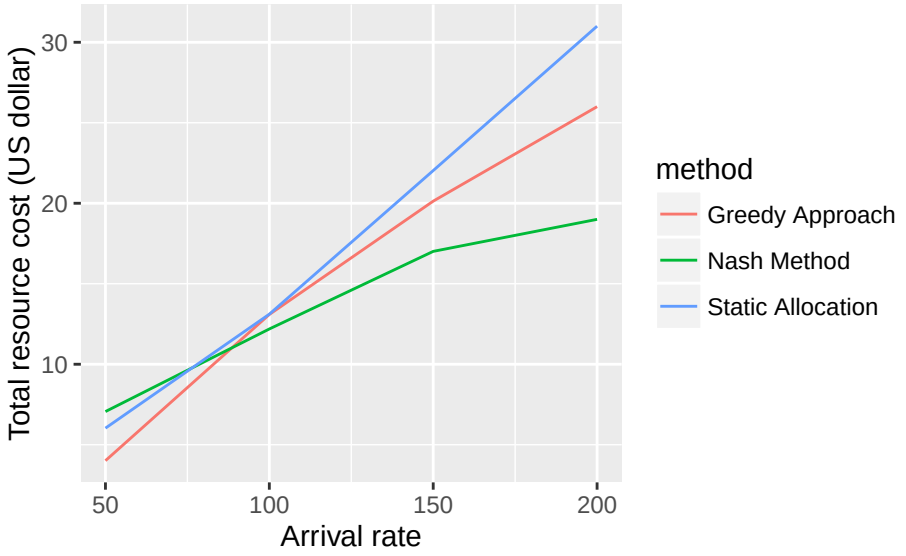

Figure 3: Simulation results showing resource cost for different approach.

geo-location, and distribution of task. In static allocation approach, VM that causes overload is migrated and is transferred to the under-utilized cloud instances. The Greedy approach selects destination instances greedily.

5.2.2 Execution Rate

The execution rate denotes, the task that finished within task deadline. Higher the execution rate means less waiting time for the task to accomplished. As shown in Figure 4, it is clear that our approach with Nash method has less waiting time compare with other algorithms. As Nash approach utilize maximum resources in given instances whereas, other approaches are underutilized. The arrival rate is Poisson distribution of workload generated from simulation for tasks. Nash approach provide fair resource allocation by reducing average waiting time for task and by utilizing maximum resources.

Figure 5 shows that by applying Nash Bargaining allocation approach, the cloud data center can handle more requests as compared to the

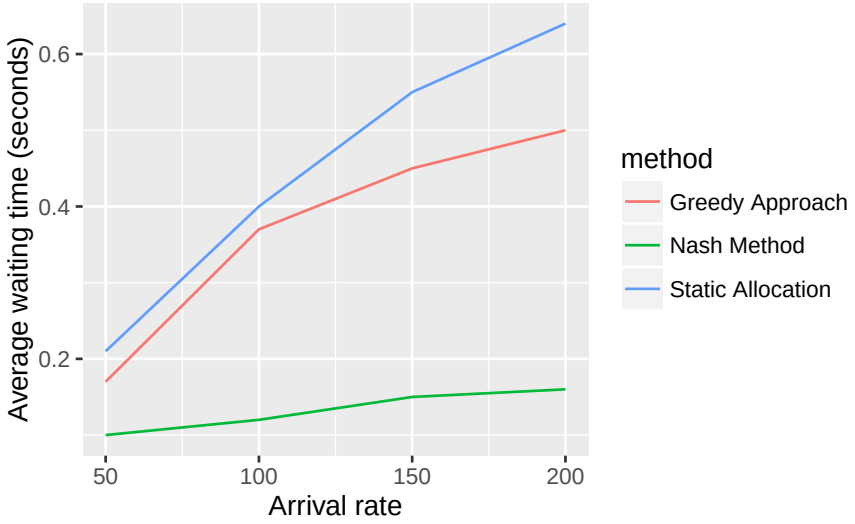

Figure 4: Simulation results for average waiting time for task from public cameras.

Greedy approach and Static approach. In Greedy and static approach most of the instances were underutilized that is why the amount of

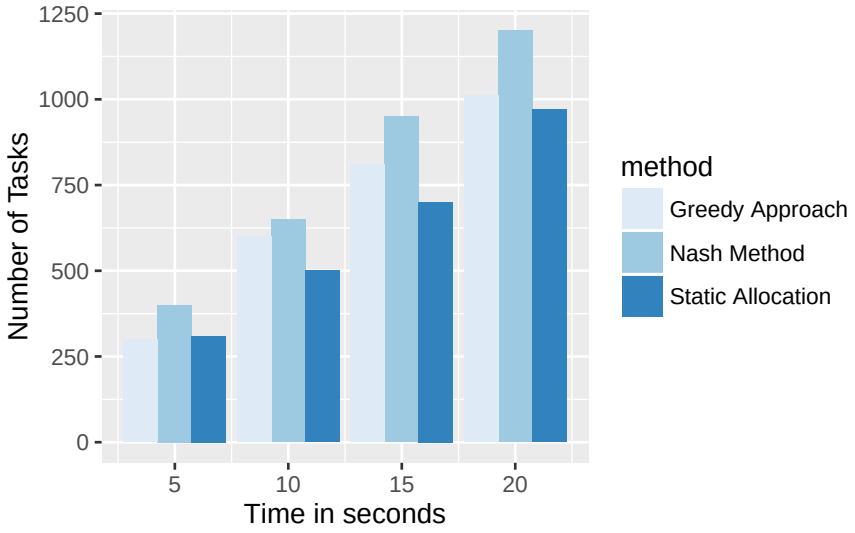

Figure 5: Number of request handled by cloud instances.

request handled by both approaches are poor. In contrast, the Nash solution utilized most of the resources present in the instances. It shows that by applying proposed Nash-based allocation algorithm, the data center can handle more requests as compare with other allocation methods. The bargaining power provides more fair utilization of resources.

6 Conclusion

In this paper, we proposed a cost-effective and dynamic resource allocation method to handle large-scale data streams from distributed cameras. We utilize the Nash Bargaining approach as the resource manager in the part of the CAM2 project. The efficiency of the proposed model was verified by several experiments in terms of a number of requests handled, average waiting time, and total resource cost. The resource manager determines its bid value based on workload, average waiting time, and user request. Experimental results clearly show that the proposed algorithm can improve efficiency and provide cost-effective resource utilization over time. In the future, Nash algorithm can be deployed to check it's effectiveness in real scenarios with CAM2 server.

References

[1] **Andreas Wolke, Martin Bichler, and Thomas Setzer**. "Planning vs. dynamic control: Resource allocation in corporate clouds." In: *IEEE Transactions on Cloud Computing* 4.3 (2016), pp. 322–335.

[2] **Brendan Jennings and Rolf Stadler**. "Resource management in clouds: Survey and research challenges." In: *Journal of Network and Systems Management* 23.3 (2015), pp. 567–619.

[3] **AT Saraswathi, YRA Kalaashri, and S Padmavathi**. "Dynamic resource allocation scheme in cloud computing." In: *Procedia Computer Science* 47 (2015), pp. 30–36.

[4] **Rafael Xavier, Hendrik Moens, Bruno Volckaert, and Filip De Turck**. "Design and evaluation of elastic media resource allocation algorithms using CloudSim extensions." In: *Network and Service Management (CNSM), 2015 11th International Conference on*. IEEE. 2015, pp. 318–326.

[5] **Lingyang Song, Dusit Niyato, Zhu Han, and Ekram Hossain**. "Game-theoretic resource allocation methods for device-to-device communication." In: *IEEE Wireless Communications* 21.3 (2014), pp. 136–144.

[6] **Yi Zhu, Yan Liang, Qiong Zhang, Xi Wang, Paparao Palacharla, and Motoyoshi Sekiya**. "Reliable resource allocation for optically interconnected distributed clouds." In: *Communications (ICC), 2014 IEEE International Conference on*. IEEE. 2014, pp. 3301–3306.

[7] **Hans J Peters**. *Axiomatic bargaining game theory*. Vol. 9. Springer Science & Business Media, 2013.

[8] **Zhen Xiao, Weijia Song, and Qi Chen**. "Dynamic resource allocation using virtual machines for cloud computing environment." In: *IEEE transactions on parallel and distributed systems* 24.6 (2013), pp. 1107–1117.

[9] **Mansoor Alicherry and TV Lakshman**. "Network aware resource allocation in distributed clouds." In: *INFOCOM, 2012 Proceedings IEEE*. IEEE. 2012, pp. 963–971.

[10] **Danilo Ardagna, Sara Casolari, Michele Colajanni, and Barbara Panicucci**. "Dual time-scale distributed capacity allocation and load redirect algorithms for cloud systems." In: *Journal of Parallel and Distributed Computing* 72.6 (2012), pp. 796–808.

[11] **Luis Rodero-Merino, Eddy Caron, Adrian Muresan, and Frédéric Desprez**. "Using clouds to scale grid resources: An economic model." In: *Future Generation Computer Systems* 28.4 (2012), pp. 633–646.

[12] **Rodrigo N Calheiros, Rajiv Ranjan, Anton Beloglazov, César AF De Rose, and Rajkumar Buyya**. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." In: *Software: Practice and experience* 41.1 (2011), pp. 23–50.

[13] **Patricia Takako Endo, Andre Vitor de Almeida Palhares, Nadilma Nunes Pereira, Glauco Estacio Goncalves, Djamel Sadok, Judith Kelner, Bob Melander, and Jan-Erik Mångs**. "Resource allocation for distributed cloud: concepts and research challenges." In: *Network, IEEE* 25.4 (2011), pp. 42–46.

[14] **Aminul Haque, Saadat M Alhashmi, and Rajendran Parthiban**. "A survey of economic models in grid computing." In: *Future Generation Computer Systems* 27.8 (2011), pp. 1056–1069.

[15] **Dusit Niyato, Athanasios V Vasilakos, and Zhu Kun**. "Resource and revenue sharing with coalition formation of cloud providers: Game theoretic approach." In: *Proceedings of the 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*. IEEE Computer Society. 2011, pp. 215–224.

[16] **Radu Prodan, Marek Wieczorek, and Hamid Mohammadi Fard**. "Double auction-based scheduling of scientific applications in distributed grid and cloud environments." In: *Journal of Grid Computing* 9.4 (2011), pp. 531–548.

[17] **Zhiming Shen, Sethuraman Subbiah, Xiaohui Gu, and John Wilkes**. "Cloudscale: elastic resource scaling for multi-tenant cloud systems." In: *Proceedings of the 2nd ACM Symposium on Cloud Computing*. ACM. 2011, p. 5.

[18] **Linlin Wu, Saurabh Kumar Garg, and Rajkumar Buyya**. "SLA-based resource allocation for software as a service provider (SaaS) in cloud computing environments." In: *Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium on*. IEEE. 2011, pp. 195–204.

[19] **Dimitris E Charilas and Athanasios D Panagopoulos**. "A survey on game theory applications in wireless networks." In: *Computer Networks* 54.18 (2010), pp. 3421–3430.

[20] **Mario Macıas, Omer Rana, Garry Smith, Jordi Guitart, and Jordi Torres**. "Maximizing revenue in Grid markets using an economically enhanced resource manager." In: *Concurrency and Computation: Practice and Experience* 22.14 (2010), pp. 1990–2011.

[21] **Marian Mihailescu and Yong Meng Teo**. "Dynamic resource pricing on federated clouds." In: *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*. IEEE Computer Society. 2010, pp. 513–517.

[22] **Hiroshi Nishida and Thinh Nguyen**. "A global contribution approach to maintain fairness in p2p networks." In: *Parallel and Distributed Systems, IEEE Transactions on* 21.6 (2010), pp. 812–826.

[23] **Sivadon Chaisiri, Bu-Sung Lee, and Dusit Niyato**. "Optimal virtual machine placement across multiple cloud providers." In: *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific*. IEEE. 2009, pp. 103–110.

[24] **Michal Feldman, Kevin Lai, and Li Zhang**. "The proportional-share allocation market for computational resources." In: *Parallel and Distributed Systems, IEEE Transactions on* 20.8 (2009), pp. 1075–1088.

[25] **Daniel J Veit and Wolfgang Gentzsch**. "Grid economics and business models." In: *Journal of Grid Computing* 6.3 (2008), pp. 215–217.

[26] **Eitan Altman, Thomas Boulogne, Rachid El-Azouzi, Tania Jiménez, and Laura Wynter**. "A survey on networking games in telecommunications." In: *Computers & Operations Research* 33.2 (2006), pp. 286–311.

[27] **Daniel P Palomar and Mung Chiang**. "A tutorial on decomposition methods for network utility maximization." In: *Selected Areas in Communications, IEEE Journal on* 24.8 (2006), pp. 1439–1451.

[28] **Rajkumar Buyya, David Abramson, and Srikumar Venugopal**. "The grid economy." In: *Proceedings of the IEEE* 93.3 (2005), pp. 698–714.

[29] **Peter Marbach**. "Analysis of a static pricing scheme for priority services." In: *Networking, IEEE/ACM Transactions on* 12.2 (2004), pp. 312–325.

[30] **Brent N Chun and David E Culler**. "User-centric performance analysis of market-based cluster batch schedulers." In: *Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on*. IEEE. 2002, pp. 30–30.

[31] **Christos H Papadimitriou**. "On the complexity of the parity argument and other inefficient proofs of existence." In: *Journal of Computer and system Sciences* 48.3 (1994), pp. 498–532.

[32] **Ken Binmore, Ariel Rubinstein, and Asher Wolinsky**. "The Nash bargaining solution in economic modelling." In: *The RAND Journal of Economics* (1986), pp. 176–188.

[33] **John F Nash Jr**. "The bargaining problem." In: *Econometrica: Journal of the Econometric Society* (1950), pp. 155–162.

Paper II:
Data Recovery and Security in Cloud

Data Recovery and Security in Cloud

J. Surbiryala[1], C. Rong[1]

[1] Department of Electrical Engineering and Computer Science,
University of Stavanger, Stavanger, Norway

Abstract:
Cloud computing has been evolving with an increasing popularity, which leads to the rapid adaption of cloud services for various reasons among the individuals and organizations. The main reason for this shift is because of the numerous benefits provided by cloud services such as low costs, computational power, and storage services over the Internet. Data recovery is one of the important concepts while dealing with storage devices which are basically the backbone of the cloud infrastructure. Someone with access to these servers or devices can use data recovery techniques to reconstruct the confidential data of customers once the customers have deleted their confidential or private data from the cloud. Reconstruction of such data leads to a security problem and privacy concerns for users. Even after some gains access to their data users are not aware that someone else has access to their data even though it has already deleted in their point of view from the cloud. In this paper, we look into the security problem which can arise based on the usage of data recovery tools on cloud infrastructure, once the users have deleted their data. To address this problem, we have proposed a simple method using Rename.

1 Introduction

Development in computer science is happening at a rapid pace with various developments in the processing of big data, access controls, cloud computing, Internet of Things (IoT) and so on. With the rapid changes in the ecosystem is leading to various developments in cloud computing, data analytics, and other related fields. Cloud provides various opportunities to support different services for managing them effectively. It provides many advantages for the customers such as pay per use, access various services over the Internet. Cloud services can be used by anyone from public cloud service providers or users/organizations can set up their own private cloud to meet their requirements [9].

Cloud provides various advantages to its customers that is one of the main reason for the rapid adaption of the cloud services. Cloud has been developed or adopted on top of traditional technologies which includes hardware and some of the software applications. With those technologies, the cloud is getting developed on top without any of the new developments related to infrastructure, platform, and software services apart from the new requirements which needs to be developed for the cloud. This one of the reason for the faster evolution of the cloud to serve customers [10].

The rapid evolution of cloud services and cloud computing, many of the organizations started providing cloud services. With this evolution, it is a bit challenging for customers to choose the appropriate cloud service providers to meet their goals. As it raises many security concerns for the customer to choose services (cloud service providers) from outside their organization. With new kind of threats every day, customers will be willing to choose the service providers who provide better security for their data in Cloud [8]. Some of the small organizations or individual started setting up their own private clouds to meet their needs for processing data, which might be expensive for individual or small organizations to run these cloud infrastructure and protect from the external threats, but these private clouds are most probably not exposed to real networks, they might be running on them behind their firewalls.

In the cloud one of the main aspects which need to be addressed is security. Security problems in the cloud need to be addressed at each step to promote the services for new customers and encourage the old customers to continue using the services. Security issues can be better addressed by cloud service providers rather by their customers, but some of the issues can be solved with the help of third-party applications as well [7]. In [3], we have presented various problems in the cloud and how these problems can be solved using the homomorphic encryption. Data recovery is an important concept which needs to be understood well, how this can be applicable to the cloud environments, and what needs to be done to protect the data in the cloud after the usage.

The focus of this paper is to check the data recovery in a cloud environment. As the data recovery can raise security concern for the customer data once they have deleted their data from the cloud, provided someone with access to the cloud infrastructure it might be possible to get access to the customer data. So, in this paper we are looking at the following problems:

(1) Is data recovery applicable in a cloud environment?

(2) Does these data recovery techniques are going to raise any security challenges in the cloud?

(3) What can be done to fix these security issues?

Our main contribution in this work is to highlight that data recovery is possible in the cloud environment. Applicability of the data recovery in cloud leads to a security issue which needs to be addressed to protect the customer's data once they have deleted. To solve this problem, we have proposed a framework using the simple rename module to protect the data in the cloud before data is deleted. With the adoption of the proposed framework, customer data cannot be reconstructed without knowing the actual format of the data.

Rest of the paper is organized as follows: Introduction to the Cloud computing and the problems we are going to address in this paper are presented in Section 2. In Section 3, we have discussed the data recovery in the cloud with one of the application as a reference to

data recovery tools. A proposed framework to address the problem has been presented in Section 4. In Section 5, we have discussed some of the problems addressed in this paper and finally, in Section 6 we draw the conclusion.

2 Cloud computing

As per National Institute of Standards and Technology (NIST) "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" [6]. The cloud models need to have five fundamental characteristics, three service models, and four deployment models.

2.1 Characteristics

Five characteristics which needs to be available in cloud services are [6]:

(1) **On-demand self-service:** customer can provision computing capabilities they want to use without any human interaction from the service provider, such as network storage and server time.

(2) **Broad Network access:** all the services are available over the network and can be accessed through standard devices, such as desktops, laptops, and mobiles.

(3) **Resource pooling:** Service provider should be able to provide resources for multiple consumers based on their requirements at dynamic phase irrespective of their location.

(4) **Rapid elasticity:** customers can easily provision and release the cloud resources as per their requirement. Capabilities of

cloud service provisioning for the customer should not be limited.

(5) **Measured Service or Pay per use:** cloud systems should optimize resource utilization and customers have to pay only for the resources used.

2.2 Service models

In the cloud, there are three types of service models [6]:

(1) **Cloud Infrastructure as a Service (IaaS):** The service provider provides the customer to choose the basic infrastructure such as storage, processing power, network, and other computing resources. The customer can use this infrastructure to develop their own platform and applications on top. The customer doesn't have to worry about the infrastructure related issues, and they will have complete control over the platform and applications deployed in the cloud.

(2) **Cloud Platform as a Service (PaaS):** The services provided to customers are deployed over IaaS. They need not worry about the infrastructure and platform. They can deploy their own applications on top of these services and they will have control over deployed applications or service, but they will not have complete control of underlying cloud infrastructure and platform.

(3) **Cloud Software as a Service (SaaS):** Service provider provides the customer with capabilities to use the applications running on their cloud environment. These applications can be accessible from through client interfaces such as browsers (e.g., dropbox, email) using personal devices (e.g., desktop. Laptop, mobile). The customer doesn't have to worry about the underlying cloud infrastructure, platform, and other services or applications running.

2.3 Deployment Models

There are four deployment models in cloud [6]:

(1) **Private cloud:** organization will have their own cloud infrastructure managed by the organization or third party.

(2) **Community cloud:** several organizations with same objectives will have shared cloud infrastructure managed by one of the organizations or third party.

(3) **Public cloud:** cloud services are available to general public or to groups and managed by an organization selling cloud services.

(4) **Hybrid cloud:** cloud infrastructure is a combination of two or more cloud deployment models, can be used as per their requirement.

Cloud has been developed on top of the various technologies which is dependent on them directly or indirectly for its main functionalities. The various problems associated with these technologies would be still applicable for cloud computing. Some of the security issues are addressed, still there are many open challenges in the Cloud which need to be addressed [4]. Apart from existing challenges cloud getting exposed to the new challenges with its evolution.

2.4 Problem

As discussed in the Section 1, data recovery creates trust issues between the cloud service providers and users. How customers can trust the cloud service providers with their data usage in Cloud. As the cloud service providers or someone with access to the cloud infrastructure can recover the deleted data using the data recovery techniques. What needs to be done to increase the user trust in cloud service providers. For that purpose, we will address into several problems in this paper.

(1) Is it possible to recover the data once it has been deleted in a cloud environment?

(2) Does these data recovery techniques are going to raise any security challenges in the cloud?

(3) How to fix the data recovery problem in a cloud environment?

3 Data Recovery in Cloud

In this section, we will discuss the data recovery options which are available and can be used effectively in the cloud to recover the data. Data recovery tools work based on either related memory analysis or carving techniques. There are many open source applications and companies which provide these services to recover the data [1]. We will not go much into the details of all techniques, but we will present one of the applications which can be used to recover deleted files in a cloud environment.

3.1 PhotoRec

PhotoRec is one of the applications which is freely available for the users to recover the data after deletion from memory devices. The application works by reading the memory blocks and reconstructing the data present in those memory locations. PhotoRec is an open source application which works with the different file system. This application can recover around 480 types of data.

Most of the present file formats stores the files in blocks. These blocks of related data are usually stored in contiguous memory blocks. When these files are deleted, usually metadata related to the files will be deleted and the memory related to that data will be marked as free space and available for allocation.

Until new data is stored in these memory locations deleted data will be present, but will not be visible in our systems. It is possible to recover the deleted data using data recovery tools such as PhotoRec.

PhotoRec uses the boot record of these machines to find the block size, then based on this information it will start recovery of the data by matching the content with correct format of the data [5].

3.2 Yelp Photo Dataset

For data reconstruction, we have used the Yelp Photo Dataset 11 [2]. This yelp photo dataset includes around 200,000 jpeg images of various food items. The total size of the dataset is 7.50 gigabytes (GB). We mainly used these images to check how much content can be recovered using the memory analysis tools, once the images are deleted. The main reason for using this dataset is, it does not include any confidential information and to check how much data can be recovered.

3.3 Results

Yelp Photo Dataset 11 consisting 200,000 jpeg images are deleted. Then, we have used data recovery tool (PhotoRec) to recover the data. We were able to recover around 7.28 GB of data leading to a loss of just 2.93 %.

3.4 Security problem

It is evident from the above results, it is possible to recover the data in the cloud using memory analysis tools once data has been deleted in the cloud. If these recovery tools are applied in the cloud who has access to the actual infrastructure after the deletion of data without the knowledge of the customer, they will be able to recover the confidential user's data which is deleted from the Cloud. Reconstruction of data without users knowledge leads to a privacy and security concerns for the users.

4 Proposed Framework

With the rapid increase in the adoption of cloud services to achieve the personal or organizational goals usage of the cloud services are increasing at the faster rate. Once the usage of cloud services has been done, the customer will collect their analysis or data anything which is present in the cloud and move out. But using the data recovery techniques similar to the one presented in the previous section can be used to recover the data, by compromising the confidential information.

To address the security problems associated with the data recovery and protect the customer data in the cloud after deletion, we propose a simple framework in the cloud to solve data recovery problem and making hard for unintended recovery of confidential customer data after deletion.

4.1 Components

For a better understanding of the proposed framework, let us first introduce the various components of the framework as shown in the Fig. 1.

- **Users** or customers are the persons or organizations who are using the cloud services to achieve their goals by using the Cloud services provided the cloud service providers. They can use the cloud services for any purpose.
- **Cloud** here in this context is the services provided by service providers for the users to use them over the network. Cloud services can be a storage area, processing power, or some other services. But all the services are provided over the network.
- **Server** is the actual hardware or infrastructure used by the users to store or process their data. In real time, a single server might be shared between the several users.
- **Rename** is a proposed new module which needs to be developed

or can be a simple program which can be accessible to users for protecting their data. The main functionality of this rename module would be, it would take the location or path of the files which needs to deleted from the cloud. It will modify extension of all those files with some other file types to make it hard data recovery.

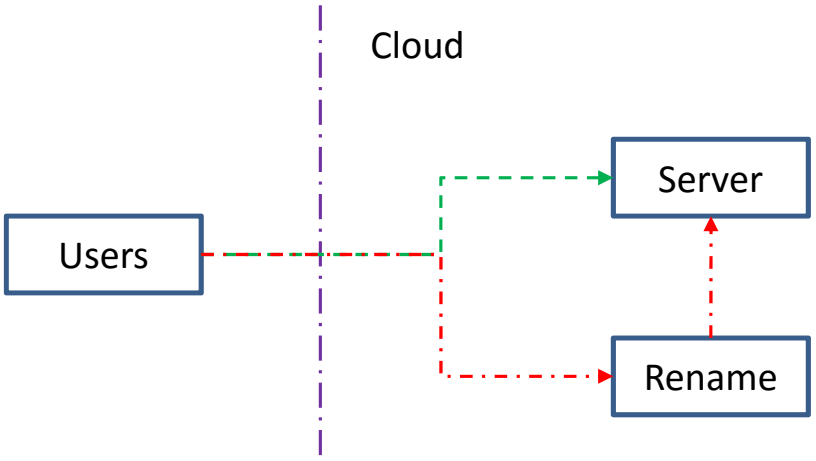

Figure 1: Overview of proposed framework

4.2 Approach

In this proposed new approach, to protect the customer data in the Cloud we are proposing to use a new module (rename). Rename module will be renaming all the files and changes the file format of all the files. The main purpose of this changing the format of the files is to protect the data. As it will hard for the someone else to know what type of file the original data had and it makes impossible to understand the data in another format. The proposed system includes the following steps:

- Users need to know the path to files, which they want to protect and pass it to rename module.

- Then enforce the renaming of data with some different type for all the files in Cloud.

By following this approach the content is not actually deleted, but the information present in the data will not be understood by anyone without knowing the actual type of the data. Then, users can go ahead and delete the files or data from the Cloud. This way even if some gains access to the infrastructure they might be able to reconstruct some files using data recovery tools presented in section 3, but they will not be able to understand the content.

Our proposed framework consists of a new module (rename) for changing the extensions of the files compared to general cloud architecture. As shown in the Fig. 1, when the users want to protect the data from reconstruction, they will send a request to Rename module. Then, the module will start renaming the files with some other type of extensions. Once the rename operation has been done, the customer can see the new files with modifications and some random types of data. Then customers can delete the data in Cloud without worrying about data recovery. Sequence diagram for Renaming the data is shown in the Fig. 2 .

4.3 Implementation requirements

Implementation of renaming module needs some proper understanding of what types of data exist in order to protect the data effectively without any compromises on different types of data formats. The first module needs to find the type of data and their format. Then, the module needs to find another format to which the data need to modified or changed so that no one will be able to reconstruct the data into the human-readable or understandable format. Once the renaming has been done, users will be free to delete the data without worrying about data recovery.

The proposed framework has several steps involved in protecting users data in the Cloud. These steps can be divided into the following steps based on the proposed framework:

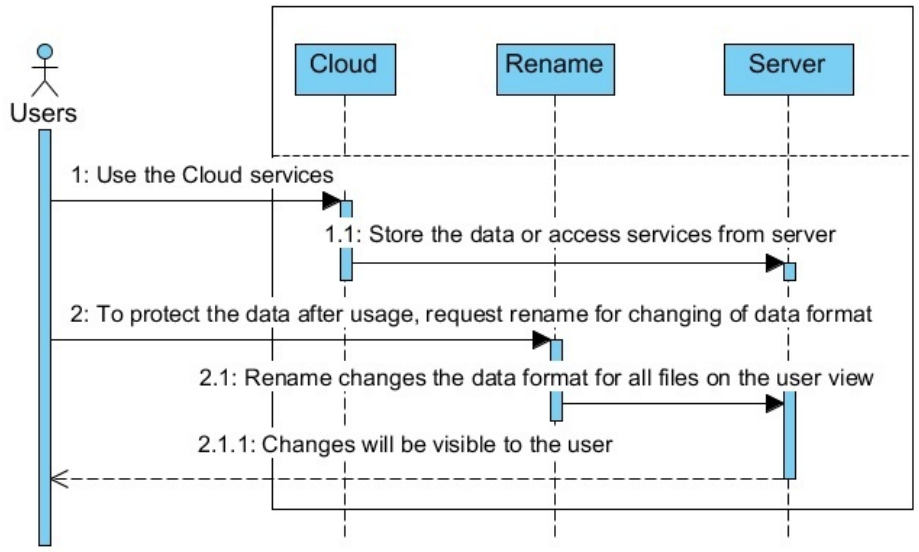

Figure 2: Sequence diagram for protecting data with proposed framework using rename

- Once the user had decided to protect the data, they need to call Rename module the path of the files which needs to be protected.
- Based on the format of these files, Rename module will enforce renaming of all the files in the given path to some other format not matching the original format. Once the rename operation is done, it will be hard even if someone has recreated this files. As data is in the wrong format it would be impossible to know the information present in the files without changing them back to original format.

4.4 Execution Time

Let us analyze the proposed method in terms of amount time it will take to complete the task of renaming the data in Cloud environment. To protect the data, each type file with any type of data needs to be renamed to some other type. So, let us assume time for renaming

the file from any format to any format will take the same amount of time (T_{rename}) in the Cloud then, time renaming a file is:

$$T_1 = T_{rename} \tag{5.1}$$

Where, T_{rename} is the time required to name rename each file and T_1 will be total time taken to rename a single file. In order to protect the complete data, all the files need to be renamed.

$$T_{total} = N * T_{rename} \tag{5.2}$$

Where, N represents the total number of files which need to be renamed to protect the customer data in the Cloud and T_{total} is the total time required for renaming the total content. It is clear from the equation 5.1 and 5.2, the time required for renaming is directly proportional to the number files needs to be renamed and this represents the extra cost for the customer once they are doing using the cloud services.

5 Discussion

In this work, we have discussed the framework to secure the data using data recovery tools or techniques.

(1) **Is it possible to recover the data once it has been deleted in a cloud environment?** Yes, it is possible to reconstruct the data using data recovery tools based on the memory analysis and carving techniques.

(2) **Does these data recovery techniques are going to raise security challenges in the cloud?** Yes, the data recovery techniques will raise the security concern for the customer data as it would be possible to reconstruct the data even after the data deletion with access to infrastructure.

(3) **How to fix the data recovery problem in a cloud environment?** With the adoption of the proposed framework, the customer can rename their data types with some unknown or other data types. Even someone with access to the cloud infrastructure it would be hard to know the actual format of the data and understand the data without knowing the format of the data.

6 Conclusion

In this paper, we have proposed a simple framework to protect the customer data in the cloud once they are done using their data in a cloud environment. The proposed framework works effectively by using the concept of file format. It will be hard to understand the content of the files in wrong data format unless these files are modified to original data formats. We have discussed our approach with various steps to achieve the proposed framework. In the future work, we would like to check the applicability of the proposed framework on real-time scenarios.

References

[1] **Forensics wiki**. *Tools:Data Recovery*. `http://www.forensicswiki.org/wiki/Tools:Data_Recovery`. [Online; accessed 31-Jan-2018].

[2] **Yelp**. *Yelp photo dataset*. `https://www.yelp.com/dataset/challenge`. [Online; accessed 31-Jan-2018].

[3] **Jayachander Surbiryala, Chunlei Li, and Chunming Rong**. "A Framework for Improving Security in Cloud Computing." In: *2nd IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA 2017)*. IEEE. 2017.

[4] **Kevin Hamlen, Murat Kantarcioglu, Latifur Khan, and Bhavani Thuraisingham**. "Security issues for cloud computing." In: *Optimizing Information Security and Advancing Privacy Assurance: New Technologies: New Technologies* 150 (2012).

[5] **C Grenier**. "Photorec." In: *URL http://www.cgsecurity.org/wiki/PhotoRec* (2011).

[6] **Peter Mell, Tim Grance, et al.** "The NIST definition of cloud computing." In: (2011).

[7] **Åsmund Ahlmann Nyre and Martin Gilje Jaatun**. "A probabilistic approach to information control." In: *Internet Technology Journal* 11.3 (2010), pp. 407–416.

[8] **Scott Paquette, Paul T Jaeger, and Susan C Wilson**. "Identifying the security risks associated with governmental use of cloud computing." In: *Government information quarterly* 27.3 (2010), pp. 245–253.

[9] **Meiko Jensen, Jörg Schwenk, Nils Gruschka, and Luigi Lo Iacono**. "On technical security issues in cloud computing." In: *Cloud Computing, 2009. CLOUD'09. IEEE International Conference on*. IEEE. 2009, pp. 109–116.

[10] **Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb**. "A Taxonomy and Survey of Cloud Computing Systems." In: *NCM* 9 (2009), pp. 44–51.

Paper III: Secure Customer Data over Cloud Forensic Reconstruction

Secure Customer Data over Cloud Forensic Reconstruction

J. Surbiryala[1], C. Rong[1]

[1] Department of Electrical Engineering and Computer Science,
University of Stavanger, Stavanger, Norway

Abstract:
Cloud computing has been evolving over the last couple of years. Adoption of the cloud services for usage in real time has been increasing ever since in the era of Big Data and Internet of Things (IoT). Adoption of Cloud is evident and can not be stopped by the individuals and small organizations. It offers many features, at low or reduced cost and provides the great potential. Cloud computing provides that extra support to move applications from the traditional local systems to cloud service provider's storage locations. But there are many security aspects which need to be considered while moving to Cloud, which are yet be solved to be completely secure in the Cloud. In this paper, we look into new security issue in the Cloud forensics point of view. Can we secure the data over cloud forensic reconstruction of the private data once that customer has done with the services and taken out their private data? The main focus of this paper is to look for an alternative ways to protect the data in Cloud.

1 Introduction

Modernization and computerization has led to many developments and various trends like distributed processing, resource sharing, access controls and so on in the Cloud computing. This has even led to Big Data analysis and Internet of Things (IoT) in the Cloud. Cloud computing has been evolving as to provide various services for managing the data and applications in the Cloud. Cloud services allow the customers to use the services required and pay for the services they have used as per their requirements. Cloud services can be provided by the cloud service providers or can be managed by the individuals or organizations of their private cloud infrastructure [6].

Cloud computing has been evolving with many of advantages provided by it, to all of the customers using the various services provided by the Cloud service providers. Cloud computing can be treated as a combination of various traditional services including the Infrastructure services which provide the actual hardware requirements to support the Cloud, Platform services which cover various hosting environments, and Software services which support the various software or web applications available in the Cloud. From all of the existing and present developments clearly indicate the evolution of the Cloud as future [7].

Many of the services in the Cloud are provided by various cloud service providers, it is challenging for the customers to choose the appropriate service provider to meet their requirements in the Cloud. For some customers or small organizations it might be more challenging if they are working on their own applications, they will be more worried about the security of their data in the Cloud. Especially for organizations working on their own products, it will be more challenging. Based on the security provided by various service providers, will be one of the factor in choosing the cloud service providers for customers or organizations in the Cloud [5].

Security is one of the main issues which need to be solved in the Cloud. Security issues need to be solved at every step in the Cloud to encourage more of the customers to adopt the services. Most of

the responsibilities in Cloud are owned by the cloud service providers, and most of the security issues need to be solved by the service providers only. Some of the issues can be solved with the third party applications also [4]. In [1], we have proposed to use the homomorphic encryption for improving the security of operations in Cloud. But, as suggested in that work it might take some more time in adopting homomorphic encryption in real time cloud environment.

The main focus of this paper is to solve one of the problems posed by the Customers data once the customer has stopped using the cloud services. Customers will be thinking more of the security for the data while using the cloud services. But even when the customers are done using the cloud services, they will delete all of their confidential information or data from the Cloud. Can the customers be sure that their data is not available in the cloud after they have moved out of it? It raises one important question **Problem:** Is the customer data has been completely deleted from the Cloud, once they have stopped using the Cloud services?

Our main contribution in this work is to protect the customer data from the Cloud forensic reconstruction [9]. We have identified the one of the most common problem which can arise from using the Cloud services from various cloud service providers. Once the customers have done with their usage of the cloud service, can the customers will be sure that their data will not be reconstructed from the infrastructure traces using the forensic applications [9]? No, the customer can ever trust anyone else apart from them selves. To solve the problem of reconstruction using the cloud forensic applications, we propose a framework to solve the problem of reconstructing the customer data once it has been deleted or stopped using the cloud services.

This paper is organized as follow: Introduction to the Cloud and one of the problem associated with cloud is presented in Section 2. In Section 3 a new framework is proposed to address issue along with approach, requirement, and cost analysis. Finally, we draw the conclusion in Section 4 with some future work.

2 Introduction to cloud

"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of computing resources (e.g., networks, servers, storage, applications, and services) that can be provisioned and released with minimum effort or service provider interaction" [3] as per National Institute of Standards and Technology (NIST) standard definition. The cloud models should have on-demand self-service, network access, resource pooling, rapid elasticity, and pay per use are five main characteristics. It has three service models namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Cloud models will have four deployment models namely private, public, community, and hybrid cloud.

Cloud computing is a combination of various technologies, on which it depends directly or indirectly for the proper functioning of the cloud services. As the cloud computing is a combination of various technologies, security issues which are present in all those technologies will still be applicable for the cloud computing. So, there are many security issues which need to be solved in the Cloud [2]. We will not be presenting all of the security issues posed by the cloud computing. Apart from all those security issues with the combination of other technologies, cloud computing will pose new security challenges which will emerge with a combination of various technologies.

2.1 Problem

As discussed in the Section 1, usage of the cloud services leaves a lot of traces of the customer data on various servers of the cloud service providers. Someone with access to the actual cloud infrastructure and can use the cloud forensic applications to generate the confidential customer data with some of the memory reconstruction techniques. This leaves with a huge risk for the customer using the cloud services, with the leakage of their confidential information.

3 Proposed Framework

With the rapid increase in the adoption of cloud services to achieve the personal or organizational goals usage of the cloud services are increasing at the faster rate. Once the usage of cloud services has been done, the customer will collect their analysis or data anything which is present in the cloud and move out. But can customers be sure if their data has been completely taken out of the cloud?

For that purpose, to protect the customer data even though the customer has moved to the Cloud services, we propose a framework in cloud computing to solve the problem and protect the customer's data after their usage has been done.

3.1 Components

To have the proper understanding of the proposed framework, let us first present all the components present in the proposed framework. Various components are shown in the Fig. 1 are:

- **Users** are the general people or organizations who want to use or using the cloud services provided by the cloud service providers to achieve their objectives in the Cloud. These customers will store, process, and analyze their data using the cloud services.
- **Cloud** is a combination of various services provided by different cloud service providers over the internet. For a better understanding of the issue, here we will consider it as one of the cloud service providers services.
- **File Manager** looks after the file management in the Cloud. It keeps track of all the files locations stored on various servers in the Cloud. It looks after how the files are stored, managed, and usage of storage space for efficient space management.
- **Servers** are the backend infrastructure used by cloud service provider to provide cloud services. Usually, location of these

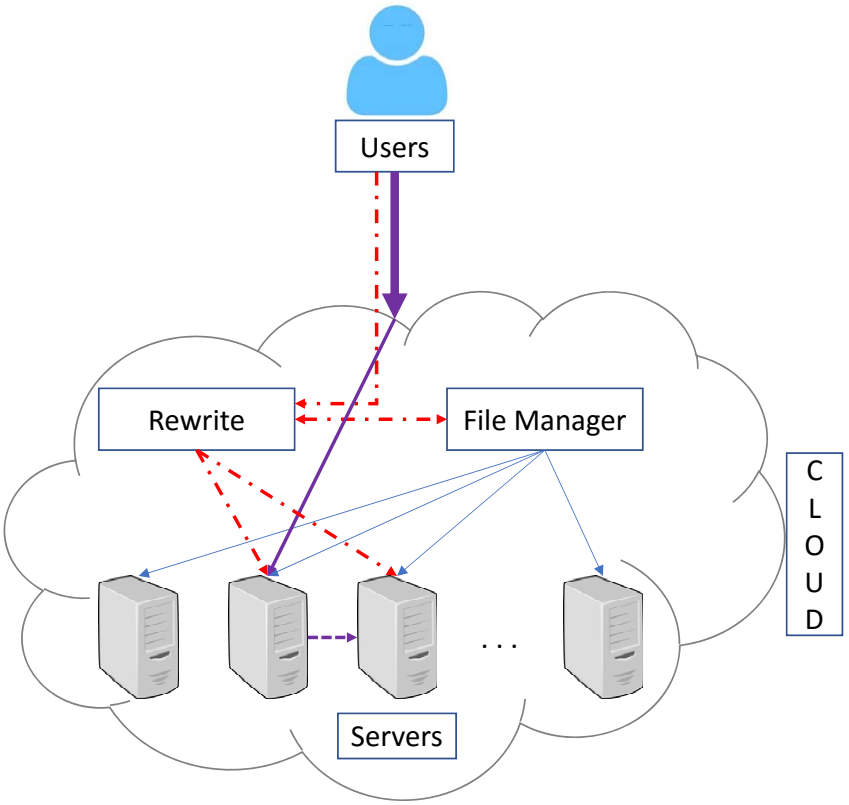

Figure 1: Overview of proposed framework

servers will be unknown to the customers, but their personal or confidential data will be stored and processed on any of the service providers servers.

- **Rewrite** is the proposed module, the application of this module in the Cloud is to provide two main functions.
  - **Delete data while using Cloud :** While using the cloud services, if the customer wants to delete some of the data from the cloud services, they can use the module to protect the data against reconstruction using forensic applications.
  - **Delete data to stop using Cloud :** Once the customer has decided to stop using the cloud services. They can use

this module to rewrite all their data with random content and erase the data from Cloud.

This module makes sure that at the actual location where these files are stored will be replaced with the random data. If someone tries to read the content from these memory locations, they will not be able to get the actual data, rather they will get some nonrelated information or bits which they will not be sure and will not be able to extract anything useful from it.

3.2 Approach

Here we present a new approach for protecting the customer data once they are done using the Cloud services. The proposed framework includes rewriting of the original data on storage devices so that even if someone gains access to the data they will not be able to read the original content. Our proposed system includes the following steps:

- First we need to identify the files which need to be deleted on the Cloud.
- Then locate all those files on the Cloud servers.
- Enforce the rewriting the data on all those locations in the Cloud.

When the files are deleted in some of the file systems, the actual content of the present in the memory will not be deleted. But these memory locations will be marked as deleted and space (size of the files) will be added to unallocated memory in the system, effectively reflecting the available space in the system with the deleted space. Data is still available in those memory locations without any change until these memory locations are used or overwritten [10]. This is one of the main reason for adopting the rewriting process to protect the customer data in the Cloud.

Our proposed framework consists of a new component for rewriting apart from the existing cloud architecture. Various components of the proposed framework are presented in section 3.1. Fig. 1 presents

the proposed framework overall view. If the customer has decided to delete some of the content in the Cloud, they will send a request to the rewrite module, it will communicate with the file manager and get the location of all the files on various servers. Then it will rewrite data. Once the modification has been done, then the customer will be informed that rewrite process on all the servers is completed. Customers can go ahead with deletion of data which is present in the Cloud.

For example, If one of the Big Data file stored in the Cloud is no longer required for the customer for analysis in Cloud, they can start deleting. If a customer has decided to protect the content they will request for rewriting the Big Data file using the proposed method. Once the customer has initiated the rewriting request, it will identify all the files which need to be protected and rewriting has to be done in the Cloud. The proposed rewriting module will locate those files and start rewriting the content so that even if someone gains access to this memory location will not be able to reconstruct the original data.

Some times data might be stored in chunks when there is not enough contiguous space to store the complete data. So, rewriting these chunks might not completely delete the data from all of these locations. For that, we need to rewrite all of these chunks several times to completely delete the Big Data files. As discussed in [11], we will adopt several rewrites to protect the information even in small chunks of data. Once this operation has been done, space will be available for reuse in unallocated space for the Cloud services.

Fig. 2 present the sequence diagram for rewriting the customer data in multiple servers of cloud service providers. Once the customer has been done using a particular file or data in the Cloud, they send the request to rewrite module to protect the data. This rewrite module will communicate with the file manager and gets the actual location of the data on their servers. Once the location of all the data files is acquired by the module, it will locate them and start rewriting the data on these locations. Once the rewrite operation has been done, customer can start deleting their data.

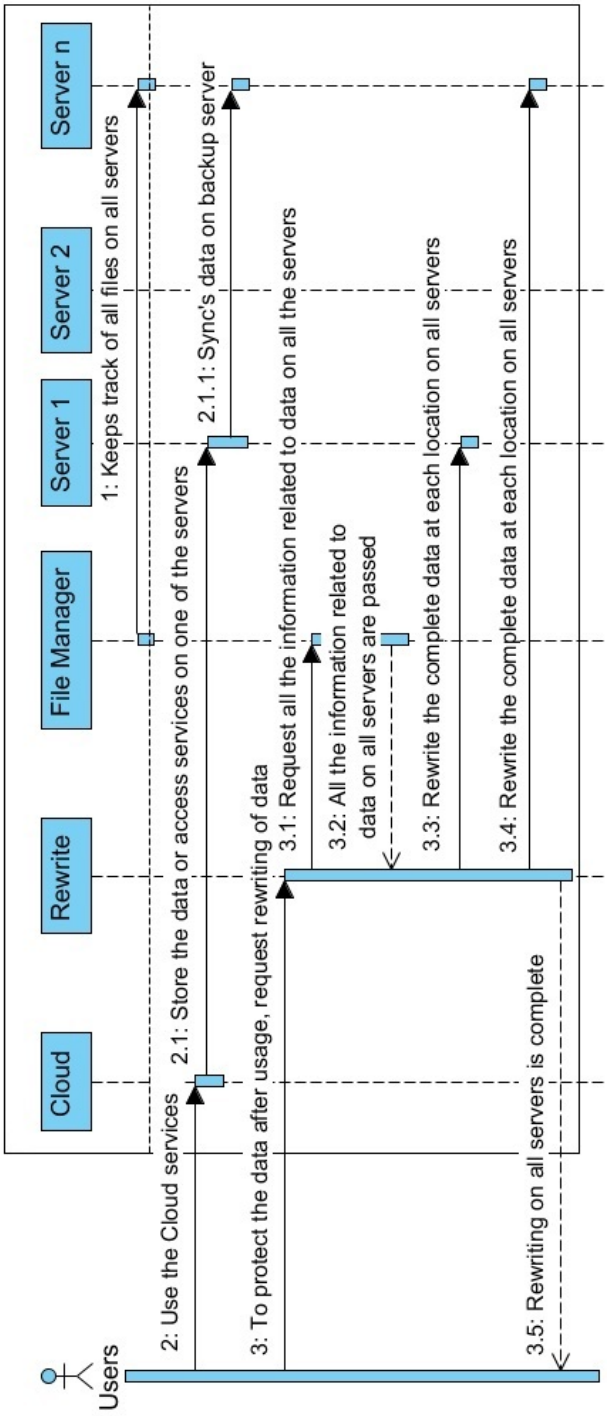

Figure 2: Sequence diagram for deleting data with proposed framework

3.3 Implementation requirements

Proper care has to be taken care while implementing the deletion of the data with rewriting and data should not be re-constructable using forensic applications. At first, we have to make sure that we will identify the exact data location on cloud service provider's servers. It can be stored in various locations in the Cloud. Once the data is modified and we have to make sure, nothing is readable. Data carving is one the most used technique in forensic investigations. carving is the process of extracting the small amounts of data from the storage area (memory). Using the header and footer values files will be carved from the memory [8].

The proposed framework consists of several commands from the customers for deleting or protecting their data in the Cloud, when data stored in the Cloud needs to be protected even after deletion. The various steps involved in protecting the customer data with the proposed framework in the Cloud are:

- Identify the location of customer data on the server in the Cloud.
- Locate them on the server.
- Start rewriting the data on those locations.

For the purpose of identifying the location of customer data on various cloud service providers servers, rewriting module needs to communicate with the cloud service providers internal file system (file manager) and fetch the location of all the files on their servers. Even if the files are located on various servers and different locations. Then these file locations need to be identified on their various servers and at last, we have to modify the content on all of those files on all the locations on these servers to protect the customer data. The rewrite will overwrite the existing information or data which is present in the Cloud. Once this operation has been done, it should not be possible to reconstruct the original data even using the forensic applications.

3.4 Execution Time

Let us analyze the time taken to delete complete data using the proposed techniques by overwriting the complete data. To delete the data, random bits are been written in all the locations where the data is present. The data deletion time for rewriting the data can be calculated as:

$$T_1 = M_{deleted} * T_{rewritting} \tag{5.1}$$

Where, $T_{rewritting}$ is the time required for rewriting 1MB of data on one of the servers, $M_{deleted}$ is M MB of data which needs to be deleted or protected on the cloud service providers servers, and T_1 is the total time required for rewriting the complete data on a single server. As discussed in section 3.2, we need to rewrite several times to protect the data against reconstruction at various locations of chunks of data [11].

$$T_{several} = S * M_{deleted} * T_{rewritting} \tag{5.2}$$

Where, $T_{several}$ is the time required for rewriting several times on a single server, S is the total number rewrites carried out to protect the data. S can be chosen as per the customer requirement. It is evident from the equation 5.1 and 5.2, the time for overwriting depends on the total number of rewrites and the size of the data which needs to be deleted from a single server. But our proposed method depends on rewriting the customer data on all of the servers. So, the total time of complete deletion on all servers is:

$$T_{total} = N * S * M_{deleted} * T_{rewritting} \tag{5.3}$$

Where, N is the total number of servers on which the customer data has been maintained by the cloud service provider and T_{total} is the total time required for rewriting the complete data on all server with various number of rewrites. Equation 5.3 presents the over all time

for rewriting the customer data on all of the servers. This overall extra time represents the extra cost for the customer.

4 Conclusion

In this paper, we have presented one of the security issue posed by the customers during or after the usage of the Cloud services related to their confidential data stored or processed in the Cloud. To solve this issue, we have proposed a new framework for Cloud. The proposed framework works similar to the forensic applications, which works based on the memory reconstruction of the deleted files. To prevent this issue, we have proposed to rewrite the complete data in the Cloud which needs to be protected on all memory locations of data stored on various servers of cloud service provider.

We have discussed our approach with various steps to achieve the proposed framework along with requirements and cost associated with the adoption of proposed framework. In the future work, we would like to check the applicability of the proposed framework on real time scenarios.

References

[1] **Jayachander Surbiryala, Chunlei Li, and Chunming Rong**. "A Framework for Improving Security in Cloud Computing." In: *2nd IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA 2017)*. IEEE. 2017.

[2] **Kevin Hamlen, Murat Kantarcioglu, Latifur Khan, and Bhavani Thuraisingham**. "Security issues for cloud computing." In: *Optimizing Information Security and Advancing Privacy Assurance: New Technologies: New Technologies* 150 (2012).

[3] **Peter Mell, Tim Grance, et al.** "The NIST definition of cloud computing." In: (2011).

[4] **Åsmund Ahlmann Nyre and Martin Gilje Jaatun**. "A probabilistic approach to information control." In: *Internet Technology Journal* 11.3 (2010), pp. 407–416.

[5] **Scott Paquette, Paul T Jaeger, and Susan C Wilson**. "Identifying the security risks associated with governmental use of cloud computing." In: *Government information quarterly* 27.3 (2010), pp. 245–253.

[6] **Meiko Jensen, Jörg Schwenk, Nils Gruschka, and Luigi Lo Iacono**. "On technical security issues in cloud computing." In: *Cloud Computing, 2009. CLOUD'09. IEEE International Conference on*. IEEE. 2009, pp. 109–116.

[7] **Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb**. "A Taxonomy and Survey of Cloud Computing Systems." In: *NCM* 9 (2009), pp. 44–51.

[8] **Simson L Garfinkel**. "Carving contiguous and fragmented files with fast object validation." In: *digital investigation* 4 (2007), pp. 2–12.

[9] **Antonio**. *Ddrescue - Data recovery tool*. `http://www.gnu.org/software/ddrescue/ddrescue.html`. [Online; accessed 31-July-2017]. 2004.

[10] **Brian Carrier, Eugene H Spafford, et al.** "Getting physical with the digital investigation process." In: *International Journal of digital evidence* 2.2 (2003), pp. 1–20.

[11] **Peter Gutmann**. "Secure deletion of data from magnetic and solid-state memory." In: *Proceedings of the Sixth USENIX Security Symposium, San Jose, CA*. Vol. 14. 1996, pp. 77–89.

Paper IV: Improve Security over Multiple Cloud Service Providers for Resource Allocation

Improve Security over Multiple Cloud Service Providers for Resource Allocation

J. Surbiryala[1], B. Agrawal[2], C. Rong[1]

[1] Department of Electrical Engineering and Computer Science,
University of Stavanger, Stavanger, Norway
[2] Analytic Innovation Center (DNVGL)
Oslo, Norway

Abstract:
From the last couple of decades, computations has changed from the client and server side to Cloud. Most of the Data generated today is completely processed in the virtual environment compared to the traditional systems. The physical location of these servers is unknown to many of the users using these services. Customers might not be aware of the storing of their data on backup servers to provide high availability in case of any failures in one of their data centers. Users are losing full control of their data by using Cloud services compared to processing their data on personal computers. What will happen to customers data once they have stopped using Cloud services? Customers data is completely deleted from all Cloud servers? Even if the data is deleted from all of their servers, can customers will be sure their data will not be reconstructed by cloud service provider (CSP) using forensic applications? To solve these problems, we have proposed a framework to solve the problem of reconstructing the data from the deleted servers using forensic applications.

1 Introduction

Cloud computing has been evolving over the last couple of decades to provide the services over the Internet. It has various architectures modeling based on the services they provide [8]. Usage of the cloud for storage and data analysis has been increasing [2]. So, in Cloud data is being processed on servers which are unknown to the customers.

Cloud computing provides many advantages to business, because of that reason many of the organizations have started moving to Cloud or building their private Cloud infrastructure. Cloud computing and usage toward Big Data storage is increasing at faster with many of the advantages provided by the evolution of cloud services. This scenario can further be improved with the more adoption of scalable and deploy-able services of the cloud for usage toward Big Data [2].

Various cloud service providers (CSP's) provide the cloud services but it is challenging for customers to choose the appropriate cloud service provider(CSP) which provides all the services required by them. It is a bit hard for the customers to evaluate the various service providers and relate with their requirements. Adoption of the cloud for various applications is not an easy task, there are many problems which need to be solved to move these organizations to Cloud. Many of these challenges are related to the security provided in the Cloud for the customer data stored in the Cloud [11].

Security is one of the major concerns in cloud computing. One of the most important factors for customers is to establish trust with the service providers. If the customer does not have confidence in the service provider, then the customer might think their confidential data may be at risk [10]. Customers need to have trust in the service providers to provide the service availability and security.

The service level agreements (SLA) between the CSP and the customer is the only legal agreement between them for the services. Customers can trust CSP based on their SLA only. But CSP discusses usage, availability, and security aspects for accessing the Cloud services [13]. Once the customers have stopped using the cloud services, what will happen to their data? Can customers trust someone with their

data, which can be reconstructed using the forensic applications [14]? Resource allocation or resource sharing among various cloud service providers is one of the main research topic which has been studied in Cloud environment [15, 12, 9, 7, 6, 5, 3, 4, 1]. All these aspects raise following security questions such as:

(1) All of the Customer related data is completely deleted from all of CSP servers, once the customer has stopped using the cloud services?

(2) How customers can be sure that their data will not be reconstructed, once they have stopped using cloud services provided by CSP?

(3) What will happen to the customer's data, if they have decided to use another CSP over the present CSP?

In this work, we have studied how can we improve the security for customer data even after they have stopped using the cloud services. To study that, we have identified some of the problems which can arise once the customer has stopped using the cloud services or moved to another cloud service provider. We have introduced the resource allocation problem between various CSP's and to address the problems possed by these techniques we have proposed a framework. This framework need to be implemented in case of resource sharing or resource allocation scenarios between the various cloud service providers. With the adoption of the proposed framework, customers need not worry about the security of their data after they have stopped using the Cloud.

The rest of the paper is organized as follows: Section 2 presents the overview of the resource allocation problem with all of its components and security problems posed by it from multiple cloud service providers. Section 3 describes the proposed framework using the user shredder module and presents how the proposed framework would solve the issues in real time. Then we discuss the advantages and disadvantages of the proposed framework in Section 4 and finally, Section 5 draws the conclusion.

2 Overview of Resource Allocation

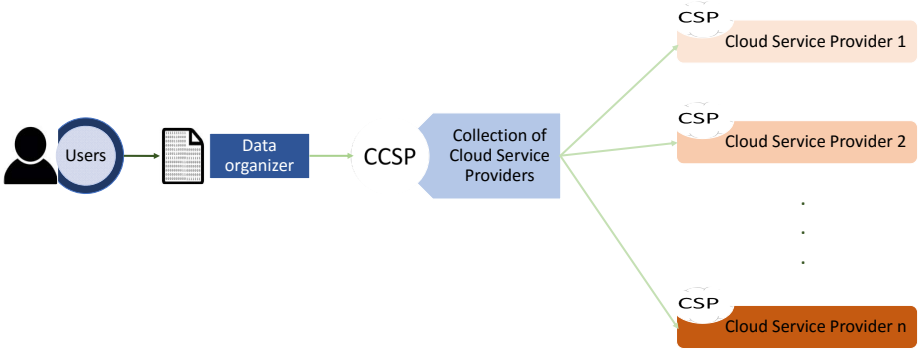

Figure 1: Overview of typical resource allocation framework

Resource allocation or resource sharing is one of the areas which is coming into existence with the evolution Cloud services [15, 12, 9, 7, 6, 5, 3, 4, 1]. Resource allocation is widely used to reduce the cost of the operations to the cloud customers at a real time. To have a better understanding of the Resource allocation problem and how this to be managed, we need to understand all the components present in it. Components of resource allocation are shown in the Fig. 1 are:

(1) **Users:** are typical customers or small organizations who would like to use the cloud services to achieve their individual or their organizational goals in the Cloud.

(2) **Data Organizer:** is service which is mainly used by the customers to manage their needs for Cloud requirement. Which makes sure they have enough resources required for the task which they would like to perform in the Cloud.

(3) **Collection of cloud service providers (CCSP):** is a service which keeps track of all the service providers of Cloud. Apart from that, it will also manage all the details related to cost and the services provided by all the cloud service providers.

(4) **Cloud service provider (CSP):** is a typical cloud service provider, who is available to provide the Cloud service for all the customers.

2.1 Scenario

Customers wishing to use and perform some operations on their personal data in the Cloud will contact data organizer. Data organizer will inform the collection of cloud service providers (CCSP) about the requirements of customers. CCSP will have information about all the existing cloud service providers (CSP) capabilities in infrastructure, platform, software, and their pricing. If some of the information is not available, CCSP will communicate with the CSP and collect all the information as shown in the sequence diagram for resource allocation in step 1 Fig. 2.

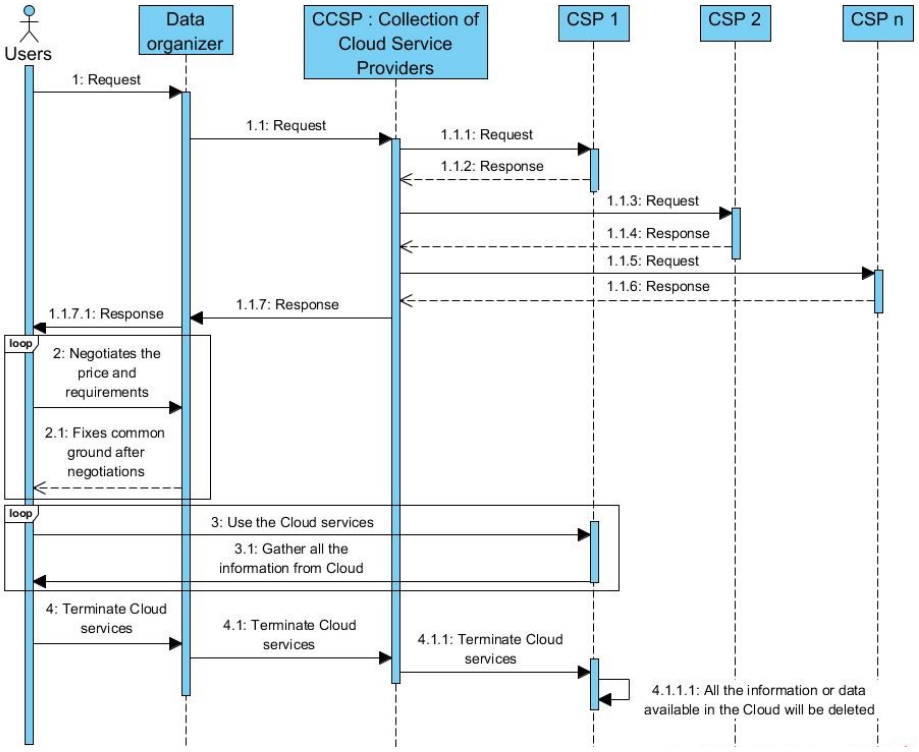

Figure 2: Sequence diagram for resource allocation framework

Once the requirement from the customer has been received by the CCSP, it will calculate the price based on the requirement for all existing CSP's. CCSP will communicate this information with the

Data organizer. Once the information about the price for particular requirements in the Cloud has been gathered by the Data organizer, the customer will start biding their requirements and start to negotiate their requirements for the cloud services with a minimum price. These requirements include the resources (Infrastructure, Platform, and Software), the time required to finish the job, and so on.

Once the negotiation has started they will find the common grounds for all their requirements and price to start working with one or more cloud service providers as shown in the sequence diagram in step 2 Fig. 2. The customer can then start their analysis using the services provided by CSP. Once the customer has finished their work in the cloud, the results from all the machine will be saved by the customers as indicated in step 3 Fig. 2. Customer will initiate the request for terminating the contract with the cloud service providers, which leads to deleting all customer data from the CSP data centers as shown in step 4 in Fig. 2.

2.2 Problems

As discussed in section 2.1, resource sharing might look as simple as that. But it might lead to security concerns in the Cloud such as:

(1) Once the customer has stopped using the Cloud services, all of the customer related data is completely deleted from all of CSP servers? or Can the customer be confident that their personal data has been deleted from all backup servers of CSP?

(2) How customers can be sure that their data will not be reconstructed using the forensic applications, once they have stopped using cloud services provided by CSP?

(3) What will happen to the customer's data, if they have decided to use another CSP over the present CSP?

3 Proposed Framework

To protect the customer's data even after they have stopped using the Cloud services, we propose a framework which needs to be adopted in Cloud computing scenarios. Apart from all the existing components in resource allocation, proposed framework will have an another module known as **User Shredder Module (USM)** as shown in the Fig. 3.

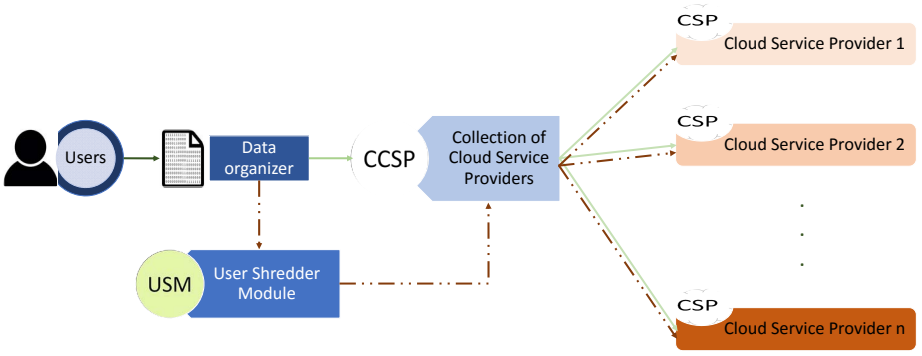

Figure 3: Proposed framework using user shredder module

USM will have an extra privilege to interact with the Data organizer and CCSP. Whenever there is any situation for the customer to stop using the Cloud services. Data organizer will inform USM about the data or content which need to be shred in the Cloud. Then USM will interact with the CCSP and gain access to those files and modify all the files with some random data such that no one can understand the information available in Cloud and can not be used for any other purposes. Sequence of steps involved in this process are presented in the sequence diagram in Fig. 4.

To understand effectiveness of the proposed framework, let us look into two scenario's and discuss how proposed framework would react to such scenarios, when they occur in real time.

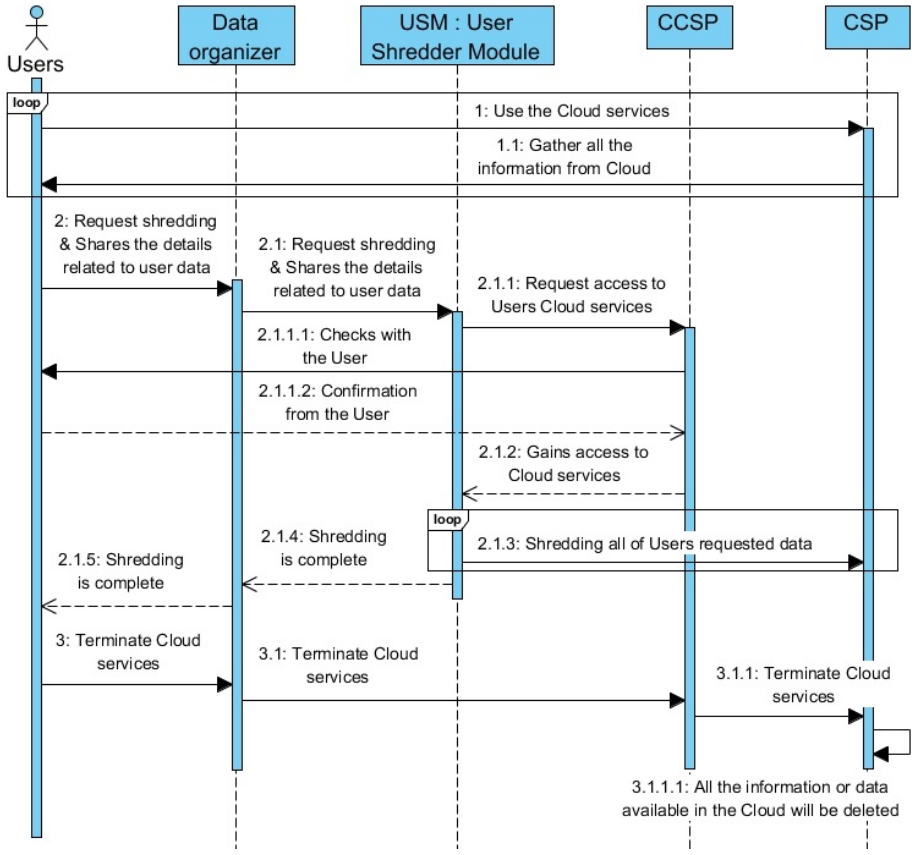

Figure 4: Sequence diagram for proposed framework

3.1 Scenario 1: What if customer has decided to use another CSP over the present CSP

Initially, when the customer has decided to use the cloud services, they start bidding process with the data organizer to use the cloud services. The CSP will be chosen from the various available CSP's from CCSP. As the customer has decided about various factors and started using the CSP. Over the time because of market changes, there are several changes in current service provider and other service providers. Customer's need might also changed over the time. Customer's will keep on checking for better solutions and to meet their requirements. So, in such scenario customer might find a better deal

from another CSP and choose to shift from the present CSP. In such scenario, the first customer needs to migrate all their data which is present in first CSP to another CSP.

Once the customer has shifted all of their data and service to new CSP, the customer can start to initiate the shredding of their data on the first CSP using USM. For that, the customer will communicate with the data organizer saying that the data which is present in the following CSP is no longer required. Please go ahead and start shredding of all those contents. Data organizer will send the request to shredder module with all the required information to shred their contents. USM will collect the information related to the data and CSP's. Then, USM will communicate with the CCSP to gain access, but CCSP will cross check with the customer. As shown in step 2.1.1 Fig. 4, if the customer confirms this request, then USM gain the access to CSP instances which need to be protected (make sure, no one can read or recover its contents later point of time) .

Once the shredder has access to all the data, which needs to be shredded. It will start modifying the original data files with some random information as shown in step 2.1.3 loop of sequence diagram in the Fig. 4. Then, this edited data will be synced on all of the backup servers of CSP to provide high availability in case of any failures in one of their data centers. Once all the data has been shredded, the customer can leave the cloud services for some more time, to make sure the data stored on backup servers are replaced with the data with unrelated information. After this, the customer can start deleting all the data from old CSP.

By using the proposed framework with user shredder module, customers confidential data can be protected after they have stopped using the cloud services from any CSP. Even someone with access to actual infrastructure will not be able to reconstruct the original data using forensic application.

3.2 Scenario 2: If the customer has decided to stop using the Cloud services, after using it for sometime

Customers when they want to use some of the cloud services, they initiate the process of getting the access with Data organizer from any of the cloud service providers. Once the customer has chosen their CSP, they will use their cloud services to achieve their goals in the Cloud.

As the customers are done with the services in the Cloud, they will collect all the information required by them and data which is present in the Cloud. Once they have all their required information and data from the Cloud, they will send a notification to data organizer saying that we no longer need the cloud instances in the CSP and data which need to be shredded in the Cloud. Data organizer will forward all the details to user shredder module. USM will interact with CCSP and gain access to the cloud instances which need to be terminated from various CSP's.

Once the USM has access to the cloud instances, it will start shredding all the data files by modifying it with some random data in those places. This modification will be done such that, random data will not be related to original data in any manner. After editing these files on Cloud, the customer has to leave their data on Cloud for some time to update on all backup servers used by CSP to provide high availability of cloud services. As soon as this modification has been completed in the Cloud by the USM, customers can start deleting their cloud instances. As the confidential data is no longer available on all of their servers. Then, customer can stop using their cloud services.

4 Discussion

There are some advantages and disadvantages using the proposed framework. Advantages of using the proposed framework are:

(1) **Security issues posed by post usage of Cloud services:** The proposed framework solves all security issues posed by customer data after the usage of Cloud services presented in section 2.2.

(2) **Security for customer data which is no longer required in the Cloud:** If the customer no longer requires some of data which is present in the Cloud. If they want to protect the information present in the data, they can also use the proposed framework to delete the content without stopping their usage of cloud services. In such scenario's customer will not request to stop cloud services as shown in sequence diagrams Fig. 4 at the end.

Disadvantages of the proposed framework are:

(1) **Cost:** It will increase the cost for the customers, in order to protect the customer's data. Customers need to maintain the cloud services for some more time to replace the data and leave it to get an update on all backup servers while using the proposed framework.

(2) **Time to write:** As the proposed framework are based on rewriting the complete data in the Cloud. If the customers have Big Data in Cloud, it might take a lot of time for writing the complete data.

5 Conclusion and Future work

In this paper, we discussed the idea of resource allocation between the various cloud service providers. We presented security issues posed by using the cloud services for resource allocation from the various CSP's. To solve these security issues we have proposed a framework. The proposed Framework uses the user shredder module to protect the customer data in the Cloud. The effectiveness of the proposed framework is presented with real time scenarios and discussed with the sequence diagrams. This clearly shows that adoption of the proposed

framework will protect the customer data even after they have stopped using the cloud services. It will encourage more customers to use the cloud services over the Internet. In the future, the proposed framework can be implemented and deployed to check the effectiveness of the framework in real time scenarios.

References

[1] **Bikash Agrawal, Jayachander Surbiryala, and Chunming Rong**. "Resource Allocation in Cloud-Based Distributed Cameras." In: *Big Data (BigData Congress), 2017 IEEE International Congress on*. IEEE. 2017, pp. 153–160.

[2] **Bichitra Mandal, Ramesh K Sahoo, and Srinivas Sethi**. "Scalable Big Data Analysis in Cloud Environment: A Review." In: *IJRCCT* 5.12 (2017), pp. 623–630.

[3] **AT Saraswathi, YRA Kalaashri, and S Padmavathi**. "Dynamic resource allocation scheme in cloud computing." In: *Procedia Computer Science* 47 (2015), pp. 30–36.

[4] **Rafael Xavier, Hendrik Moens, Bruno Volckaert, and Filip De Turck**. "Design and evaluation of elastic media resource allocation algorithms using CloudSim extensions." In: *Network and Service Management (CNSM), 2015 11th International Conference on*. IEEE. 2015, pp. 318–326.

[5] **Yi Zhu, Yan Liang, Qiong Zhang, Xi Wang, Paparao Palacharla, and Motoyoshi Sekiya**. "Reliable resource allocation for optically interconnected distributed clouds." In: *Communications (ICC), 2014 IEEE International Conference on*. IEEE. 2014, pp. 3301–3306.

[6] **Zhen Xiao, Weijia Song, and Qi Chen**. "Dynamic resource allocation using virtual machines for cloud computing environment." In: *IEEE transactions on parallel and distributed systems* 24.6 (2013), pp. 1107–1117.

[7] **Mansoor Alicherry and TV Lakshman**. "Network aware resource allocation in distributed clouds." In: *INFOCOM, 2012 Proceedings IEEE*. IEEE. 2012, pp. 963–971.

[8] **Michael Armbrust, Armando Fox, Rean Griffith, Anthony D Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, et al.** "A view of cloud computing." In: *Communications of the ACM* 53.4 (2010), pp. 50–58.

[9] **Marian Mihailescu and Yong Meng Teo**. "Dynamic resource pricing on federated clouds." In: *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*. IEEE Computer Society. 2010, pp. 513–517.

[10] **Åsmund Ahlmann Nyre and Martin Gilje Jaatun**. "A probabilistic approach to information control." In: *Internet Technology Journal* 11.3 (2010), pp. 407–416.

[11] **Krešimir Popović and Željko Hocenski**. "Cloud computing security issues and challenges." In: *MIPRO, 2010 proceedings of the 33rd international convention*. IEEE. 2010, pp. 344–349.

[12] **Sivadon Chaisiri, Bu-Sung Lee, and Dusit Niyato**. "Optimal virtual machine placement across multiple cloud providers." In: *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific*. IEEE. 2009, pp. 103–110.

[13] **Pankesh Patel, Ajith H Ranabahu, and Amit P Sheth**. "Service level agreement in cloud computing." In: (2009).

[14] **Antonio**. *Ddrescue - Data recovery tool*. `http://www.gnu.org/software/ddrescue/ddrescue.html`. [Online; accessed 31-July-2017]. 2004.

[15] **Christos H Papadimitriou**. "On the complexity of the parity argument and other inefficient proofs of existence." In: *Journal of Computer and system Sciences* 48.3 (1994), pp. 498–532.

Paper V:
Method to Solve a Privacy and Security Issue in Cloud for Energy Informatics

Surbiryala, J., Chung, R. (2020) Method to Solve a Privacy and Security Issue in Cloud for Energy Informatics, In: J.P. Liyanage, J. Amadi-Echendu, J. Mathew (Eds.) *Engineering Assets and Public Infrastructures in the Age of Digitalization: Proceedings of the 13th World Congress on Engineering Asset Management.*

This paper is not in Brage due to copyright.

Paper VI: A Framework for Improving Security in Cloud Computing

A Framework for Improving Security in Cloud Computing

J. Surbiryala[1], C. Li[1], C. Rong[1]

[1] Department of Electrical Engineering and Computer Science,
University of Stavanger, Stavanger, Norway

Abstract:
Cloud computing has been evolving over a couple of years with the increased use of cloud-based services like Amazon Web Services (AWS), Dropbox, Office 365, and so on. It revolves around the several technologies, for developing and supporting these services. With the evolution of new technologies, it leads to widespread use of cloud-based applications in real time. Processing powers are increased tremendously, more efficient, and centralized in the cloud environments. With the launch of services in a cloud environment has decreased the price of various services at the cost of security in these services. It is inevitable for most of the individuals and small organizations to use these services at reduced or less secured cloud services. There are many open challenges in the cloud environment, which needs to be addressed. In this paper, we propose a framework to solve various ethical and security aspects related to cloud computing.

1 Introduction

Cloud computing is one of the areas in information technology (IT), many things change in the blink of an eye. There are many aspects of cloud computing; people tend to use as per their convenience. A lot of research is happening around cloud computing [12, 16, 6, 5]. Cloud computing can affect the society, individuals, and firms in a way no one can imagine. However, it would be appropriate to think in security direction, how it is going to affect everyone if security in the cloud computing has not been taken seriously. It is suitable for everyone using the cloud services to understand the real security issues associated with these services.

In today's world, humans are marching towards the Internet of Things (IoT) [11] with many devices connected to the internet, the cloud can be used as one of the metaphors. Whereas, the internet is still the collection of servers (can be treated as devices) which are well connected through fiber optic cables or through some other means to provide or share the data between the connected nodes or devices. In other words, it can be stated as personal devices connecting to remote devices to perform some operation without using or limited use of private infrastructure can be called as cloud computing. Which is entirely different when compared to the traditional use of resources to perform the same computations or operations.

This model helps the individuals and organizations who would not like to invest huge amounts for short-term purposes in the platform, infrastructure, and software. Cloud computing is a model which helps to reduce the cost of operations and pay for the services which are used for the required or used period of services.

With a shift in the business model from the traditional to the cloud-based environment, there will be scope for expansion or reduction of the resources as per the requirements at a faster rate and helps to handle the resources effectively and cost efficiently.

The paper is organized into seven sections: the first section is the Introduction. Section 2 presents details about cloud computing along with various delivery models, and deployment models, Section 3 talks

about ethics with their analysis and considerations. In Section 4, a brief understanding of security in a cloud environment is presented. In Section 5 we discuss the ethical and security aspects related cloud. Section 6 describes the proposed framework using homomorphic encryption for a cloud environment, along with applicability of proposed framework and finally Section 7 draws the conclusion.

2 Cloud Computing

As per the National Institute of Standards and Technology (NIST) standards [14], there are five main characteristics which need to be available in any service to be treated as cloud service which is:

(1) **On-demand self-service:** user can carry out operation whenever he wants to use the service without any interference from anyone else.

(2) **Network access:** is accessible with any internet-connected device.

(3) **Location independent resource pooling:** resources should be shared across the users irrespective of their location.

(4) **Physical transparency:** user can change their resource capacity as per their requirement.

(5) **Pay peruse:** customer need to be charged based on the resources used.

Cloud computing can be classified based on service delivery models or deployment models. Fig. 1 **The NIST cloud definition framework** represents both models based on NIST cloud definition framework [14]. In the cloud, there are three kinds of service delivery models namely:

(1) **Infrastructure as a Service (IaaS):** users rent out the infrastructure provided by service providers such as processing power, disk storage, network and other computing resources.

The user can use this infrastructure for any of their purposes to develop their own platform and software applications.

(2) **Platform as a Service (PaaS):** users rent out the platform to perform various operations like developing and managing their own applications without any concern regarding infrastructure.

(3) **Software as a Service (SaaS):** users rent out the software applications. They just use the service without worrying about infrastructure and platform.

Cloud services can be deployed in several ways such as:

(1) **Private cloud model:** an organization with their infrastructure to set their own cloud services.

(2) **Community cloud model:** infrastructure is shared among the several organizations with shared objectives.

(3) **Public cloud model:** cloud service provider will provide the infrastructure for other organizations or normal people.

(4) **Hybrid cloud model:** where two or more models will be combined to provide the services to users or organizations.

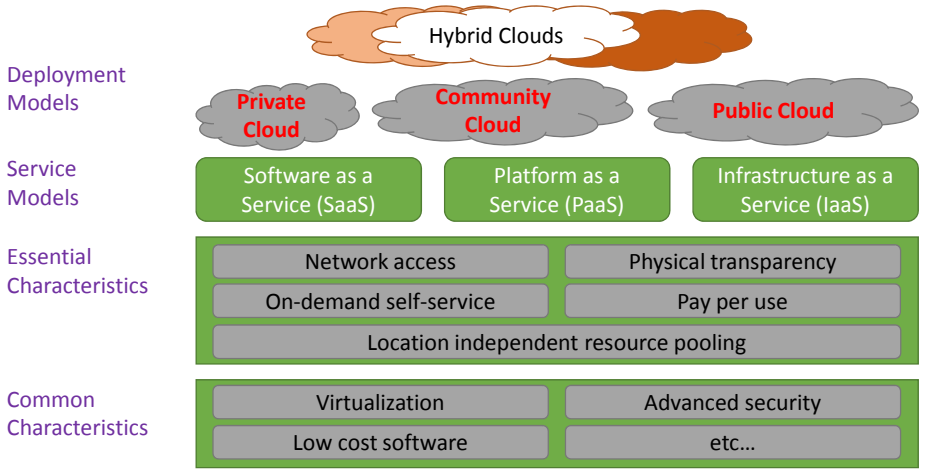

Figure 1: The NIST cloud definition framework [14]

Cloud Computing depends on various technologies, which exist in

information technology even before the cloud service started to exist. An example of such technologies are virtualization and distributed processing. Any further improvements in these technologies will also support and boost the Cloud Computing. Some of the information technology organizations, even though they were using the underlying technologies, they might call their service as cloud services without following all the requirements mentioned above [17]. Sometimes cloud computing can be treated as some fancy word used to attract new customer base to sell their product. Such behaviour can be stopped by making international standards and making sure; every organization follows those standards. NIST has already released some standards; it is profitable if all the organizations follow these standards.

3 Ethics

Ethics is one of the widely-studied topics in cloud computing after its introduction to the real world to store the information. Digital information has provided us with a lot of previous media information to understand the importance of it [7]. Cloud environment presents different ethical challenges that need to be understood properly, with unlimited access to the information, there is a probability that this information can be misused. Cloud computing is one of the things which leads to several debates on ethical aspects, as cloud environments have access to larger confidential information. Illegal or improper use of such data leads to ethical problems [19].

3.1 Ethical Analysis

Here we present the idea of ethics related to IT on various topics in connection to cloud computing. Readers who are more interested in in-depth details pertaining to the topic can refer the article [9]. IT systems lead to several issues related to ethical aspects.

Addressing the ethical aspects of cloud computing is somewhat complicated so, these issues can be dealt indirectly. There are three

accepts that are appropriate for ethical analysis in cloud computing that plays a central role in addressing ethical issues of cloud computing are [10]:

(1) **Control from technology to third parties for the cloud:** lead to loss of direct control in case of disasters like unauthorized access, infrastructure failure, or data loss.

(2) **Multiple storage locations for redundancy:** can lead to privacy issues across the various border. For example, what happens if one of the storage location has been compromised with access to confidential data to an unauthorized person?

(3) **The connection of services across different providers for a service in the cloud:** might lead to delays with an unexpected behaviour.

3.2 Dealing with Ethical Issues

Ethical issues in cloud computing include security and privacy which are obvious. It is evident that anyone will not be able to understand or identify all the ethical issues related to cloud computing. Like any normal person, cloud computing will also be facing real time scenarios that will be hard to determine. However, some basic understanding of ethical issues might help in real time scenarios in evaluating them to solve problems.

Being proactive about the ethical issues in cloud computing might solve some of the problems. In other words, we call it as the precautionary principle, which prevents the damage of unknown parameters without affecting the progress. For further knowledge on the precautionary principle, readers can refer to articles [15].

3.3 Ethical Frameworks

In this section, we will look into various theories related to ethics. These models make our analysis simple for making any decision.

They are utilitarianism, and contractarianism are two approaches for analysis, which finds out whether the actions executed are correct or not [21].

Utilitarianism [21] works on the principle of utility, which states the ethically correct action is the outcome of the welfare of all concerned actions. In the ethical analysis, we often have several options. Utilitarianism will be useful in such scenarios to evaluate those alternative options, by comparison, to produce the better result for each scenario.

On the contrary, contractarianism works with social contract [21], which serves as a counter-measure to utilitarianism. It works with the mutual agreement to get results for a person with the lesser concern of mutual service.

Another important concept is slippery slope reasoning [20]. Slippery slope presents with many problems in situations where decisions need to be taken. Sometimes it will be difficult to take decisions. It is better to avoid the contradictory frameworks like contractarianism and utilitarianism. Otherwise, we will end up with unimportant aspects for decision making as well.

As discussed in the previous subsection 3.2, core precautionary principle can be used to compare the risks with respect to lesser worst outcome [18]. It can be used in scenarios where it is difficult to make any decisions based on utilitarianism and contractarianism. In some situations, we might not be able to predict the result so, it is appropriate to use ultraconservative precautionary principle. It says that any operation should be stopped if it is going to create any problem.

3.4 Ethical Considerations

For any of the ethical issues, the computer cannot be responsible; it should be an organization that provides the service will be responsible for those mistakes. That means the organization has not met the obligatory requirements to provide the services. Organizations providing services needs to understand the consequences of these

mistakes.

If more than one approach is followed, it is better to choose their relation between them carefully. The ultraconservative precautionary principal is one of the most accepted principles in real time to solve the problems. In Section 6, we will present a new framework for the cloud environment, which can solve most of the issues.

4 Security

Security is a subjective topic according to ethical and social aspects. Security in the cloud is a process of securing data (or confidential information) from others getting access by interference or illegally without knowledge of the owner. Security in the cloud can be achieved in several ways depending on the type of controls used with the service. Many of the cloud service providers implement their own security controls to protect services and data associated with it. Sometimes organizations or paid users of the cloud services can choose security controls based on their requirements.

As discussed in the previous section some of the cloud services depend on many technologies, so security concerns associated with such technologies will still be applicable to the security issues in cloud computing. In other words, we can say that challenges faced by organizations using traditional services with other technologies are still applicable for cloud-based services. These security risks need to be managed properly and mitigated without any risk.

4.1 Security in Cloud

With the increased use and deployment of the cloud service, security issues in the cloud are also increasing. Organizations need to consider the following security issues in the cloud environment [8]:

- **Information in cloud computing environments:** can affect the organizations regarding security aspects in the cloud.

Based on confidential information stored and processed in the cloud might lead various attacks.

- **Attackers and their capabilities:** attackers can be classified into internal and external attackers. The internal attacker has inside knowledge and access to many resources in the cloud. An external attacker might not have access to internal details but, they exploit vulnerabilities to attack cloud and gain access to confidential data.
- **Risk management in the cloud:** organizations need to understand the risk of moving to cloud environment in various aspects. Cloud service providers have access to their data, if it is not encrypted. For redundancy purposes, they store data on several servers. Deleting their data doesn't mean complete deletion of data on all servers and geographic location of these servers are unknown to the users or organizations using the cloud service.
- **New cloud security risks:** organizations need to be prepared for new types of attacks which might not exist in traditional systems such as side channel attacks, social networking attacks, mobile device attacks.
- **Existing cloud security issues:** need to be well understood, and proper defense mechanisms need to be implemented.

4.2 Standardization in Cloud Computing

For a proper understanding of security measures in a cloud environment, it is better to follow standardization in cloud environment across various cloud service provider. Some of the organizations working in cloud standardization are NIST cloud standards, Cloud Security Alliance (CSA), IEEE Standards Association (IEEE-SA), and International Telecommunication Union (ITU). IEEE-SA have formed two working groups $P2301 - cloudprofiles$ and $P2302 - intercloud$ for improving standards to address migration, management, and interoperability across various cloud platforms. ITU has studied cloud

computing and standardization under study group 13 [3].

5 Issues in Cloud Computing

Cloud computing raises numerous ethical issue that arises when the control of the data is transferred from provider to a third party. Therefore, there is a need to provide security and privacy for the users using cloud services, and to make it is possible for authorized persons to have control over the data.

Using cloud services, users exposed to several risks associated with their data, computation, and analysis. Cloud computing imposes challenges while transferring the data for any further action or analysis. Any such challenges need to be addressed and adequately discussed with users and cloud provider.

Security in IT is a combination of confidentiality, integrity, and availability known as CIA model.

(1) **Confidentiality:** refers to access to the data to only authorized persons.

(2) **Integrity:** related to the data modification is not possible.

(3) **Availability:** ensures that data is accessible whenever it is required.

With the increased use of cloud services from various organizations and individuals, it is an ongoing process to meet the security requirements and get up to date security with the latest trend in technology. Authentication, authorization, and ownership (identification) are some of the main security concerns in cloud services.

Protection of data from unauthorized access is one more challenge. Data protection can be achieved with the help of proper data encryption protocols. As the data has to be exchanged over the internet, any attack possible on network protocols will be still applicable to confidential information in the cloud.

Another prominent issue in any IT system security is data availability. Cloud service provider needs to ensure the high availability of cloud services and data. Data needs to be protected from various factors including unforeseen circumstances without compromising security features like unauthorized access to the data in the cloud. Cloud provider needs to make sure of redundant backups in case of failures. It is better if the users and organizations using the cloud services have their own backup in regular intervals, in the case of any such emergencies. These backups can be used to verify the integrity of the data even if there is a failure in the cloud service provider.

Vulnerabilities and failures are common in any systems, which leads to security issues in the system. This is also applicable to the cloud services. Most of the issues can be easily identified and fixed based on their logging system management. Logs help in identifying, and isolating the problems where the issue has taken place.

One more problem with cloud based services is trust. Can an individual trust these cloud service providers with their personal and confidential data? How can customers be sure, service providers are not going to access the data and sell it to the third party?

6 Proposed Framework

To solve some of these issues in a cloud environment, we propose simple and yet powerful framework using homomorphic encryption of data to store and perform secure operations in the cloud. Gentry has proposed a fully homomorphic encryption [13] which allows us to carry computations on encrypted data (ciphertext), leads to encrypted results. When the results are decrypted will match the actions performed on plain data.

Fig. 2 **Overview of the proposed framework** shows the block diagram of the proposed framework to be implemented in a cloud environment. The proposed framework consists simple steps to achieve better security and solve many of the ethical and security issues in a cloud environment.

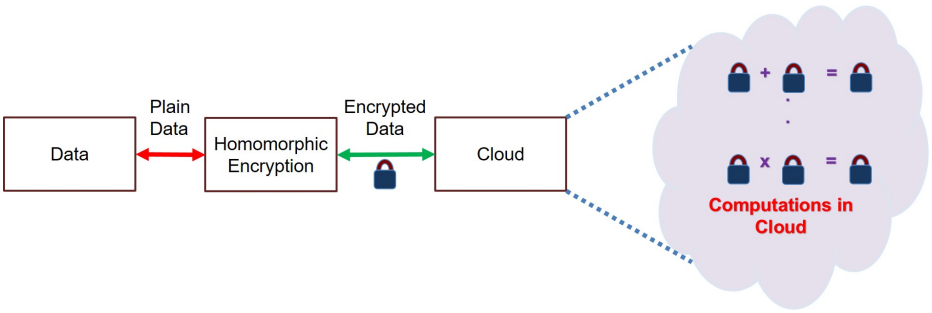

Figure 2: Overview of the proposed framework

First, we need to encrypt the plain data using the homomorphic encryption technique. Once the data has been encrypted, we can store the data in the cloud. As the data is encrypted using homomorphic encryption, we can perform computations in the cloud. The computations carried on encrypted data will same as the computations carried on the plain data because of the usage of homomorphic encryption. To get the plain data from the cloud, we need to decrypt the data using the homomorphic method with same parameters.

6.1 Homomorphic Encryption

As shown in the Fig. 2 **Overview of the proposed framework**, Homomorphic encryption allows us to perform the operations on encrypted data without decrypting in the cloud environment. A Fully Homomorphic encryption scheme should allow two basic operations on encrypted data without using the private key.

For example, if m_1, m_2 are two messages and $Encrypt$ and $Dcrypt$ are encryption and decryption process defined for a Homomorphic encryption scheme then, following operations are valid.

$$C_1 = Encrypt(m_1)\ and \tag{5.1}$$

$$C_2 = Encrypt(m_2) \tag{5.2}$$

$$m_1 + m_2 = Dcrypt(C_1 + C_2) \tag{5.3}$$

$$m_1 * m_2 = Dcrypt(C_1 * C_2) \tag{5.4}$$

6.2 Applicability of Proposed Framework

If the cloud service providers adopt to the proposed framework using homomorphic encryption, it would solve some of the issues in the cloud but not limited to:

(1) It provides security for the data stored in the cloud.

(2) Confidentiality of the data stored in the cloud will be achieved.

(3) Control and authorization of the data can be maintained as the parameters used for homomorphic encryption will only be known to the customer.

(4) Unauthorized access to data might not lead to comprise of the information stored in the cloud.

(5) Apart from all the issues addressed by the proposed framework, it allows customers to carryout computations on encrypted data in the cloud.

6.3 Reality

In cloud computing, most of the security measure need to be taken by cloud service provider. When data is stored, managed, processed, and analyzed in the cloud, security requirements are obligatory. Even security at physical locations needs to be considered. By adopting the proposed framework by big cloud service providers like Amazon EC2 [1], Google Cloud Platform [2], Microsoft Azure [4] will encourage customers to rely on them. Even though the idea of homomorphic encryption exists from a long time, it is still not developed entirely to adopt to cloud environments because of the complexity of operations involved to perform actions on encrypted data. Rapid increase in computation power possessed by cloud service providers, homomorphic encryption can be adopted in cloud environments.

7 Conclusion

Cloud computing provides many advantages for individuals and small organizations; it can also create some serious security issues with personal and confidential data. Cloud service providers should take proper security measures to prevent all security related issues. We have proposed a simple yet powerful framework using homomorphic encryption for solving data security problems in cloud environments. Adoption of the proposed framework will solve many of the issues in cloud environment related to ethical and security aspects.

Acknowledgment

The authors are grateful to the anonymous reviewers for their constructive suggestions to improve the quality of the paper.

References

[1] **Amazon Elastic Compute Cloud**. URL: `https://aws.amazon.com/ec2/` (visited on 03/15/2017).

[2] **Google Cloud Platform**. URL: `https://cloud.google.com/` (visited on 03/15/2017).

[3] **IEEE**. *Standards in Cloud Computing*. URL: `http://cloudcomputing.ieee.org/standards` (visited on 03/15/2017).

[4] **Microsoft Azure**. URL: `https://azure.microsoft.com` (visited on 03/15/2017).

[5] **John W Rittinghouse and James F Ransome**. *Cloud computing: implementation, management, and security*. CRC press, 2016.

[6] **Chunming Rong, Son T Nguyen, and Martin Gilje Jaatun**. "Beyond lightning: A survey on security challenges in cloud computing." In: *Computers & Electrical Engineering* 39.1 (2013), pp. 47–54.

[7] **Vanessa Ratten**. "Entrepreneurial and ethical adoption behaviour of cloud computing." In: *The Journal of High Technology Management Research* 23.2 (2012), pp. 155–164.

[8] **CPNI**. *Cloud computing - information security briefing - 01/2010*. 2010. URL: `http://www.cpni.gov.uk/Documents/Publications/2010/2010007-ISB_cloud_computing.pdf` (visited on 03/15/2017).

[9] **Scott Paquette, Paul T Jaeger, and Susan C Wilson**. "Identifying the security risks associated with governmental use of cloud computing." In: *Government information quarterly* 27.3 (2010), pp. 245–253.

[10] **Job Timmermans, Bernd Carsten Stahl, Veikko Ikonen, and Engin Bozdag**. "The ethics of cloud computing: A conceptual review." In: (2010).

[11] **Rolf H Weber and Romana Weber**. *Internet of Things*. Vol. 12. Springer, 2010.

[12] **Qi Zhang, Lu Cheng, and Raouf Boutaba**. "Cloud computing: state-of-the-art and research challenges." In: *Journal of internet services and applications* 1.1 (2010), pp. 7–18.

[13] **Craig Gentry**. "A fully homomorphic encryption scheme." PhD thesis. Stanford University, 2009.

[14] **Peter Mell and Tim Grance**. "Effectively and securely using the cloud computing paradigm." In: *NIST, Information Technology Laboratory* (2009), pp. 304–311.

[15] **Wolter Pieters and André Cleeff**. "The precautionary principle in a world of digital dependencies." In: *Computer* 42.6 (2009), pp. 50–56.

[16] **Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb**. "A Taxonomy and Survey of Cloud Computing Systems." In: *NCM* 9 (2009), pp. 44–51.

[17] **Maria Spinola**. "An essential guide to possibilities and risks of cloud computing." In: *Retrieved March* 24 (2009), p. 2011.

[18] **Stephen M Gardiner**. "A core precautionary principle." In: *Journal of Political Philosophy* 14.1 (2006), pp. 33–60.

[19] **Oliver Freestone and V Mitchell**. "Generation Y attitudes towards e-ethics and internet-related misbehaviours." In: *Journal of Business Ethics* 54.2 (2004), pp. 121–128.

[20] **Simon Blackburn**. *Ethics: A very short introduction*. Vol. 80. Oxford University Press, 2003.

[21] **Alan P Hamlin**. "Rights, indirect utilitarianism, and contractarianism." In: *Economics and philosophy* 5.02 (1989), pp. 167–188.