DATA INTEGRITY FOR CLOUD COMPUTING WITH HOMOMORPHIC ENCRYPTION

RUBA N S AWADALLAH

UNIVERSITI SAINS MALAYSIA

2022

DATA INTEGRITY FOR CLOUD COMPUTING WITH HOMOMORPHIC ENCRYPTION

by

RUBA N S AWADALLAH

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

August 2022

ACKNOWLEDGEMENT

I praise Allah first of all the Almighty, for giving me the opportunity to go through this journey and carry on successfully in my work. To my father Eng. Naeem Awadallah and my mother Hamdiah Albalawi I would like to extend my heartfelt gratefulness and deep thanks for your ongoing prayers and for your continuous sacrifices in order to fulfill my dreams (may Allah prolong your life and health).

Once again, thank Allah who destined for me to be under the supervision of a well-educated, patient, and continuous supporter to complete this thesis. From the bottom of my heart, I would like to express my sincere appreciation to my supervisor Prof. Dr. Azman Samsudin for his thoughts, encouraging words, and his always faith had in me, he has always been like a father through this path. I would like to thank my colleagues in the Computer Science Society for their help in sharing valuable information.

I also extend my heartfelt thanks to my brothers and sisters who have always motivated me with their supportive words. A special thanks to my mother-in-law, my friends, cousins, and every person who left a beautiful footprint in my life.

Last but not least, my soulmate and the father of my children, Dr. Saleh Amarneh, thank you very much because you loved me sincerely and whatever I say, I will not give you your right to describe your fatigue with me through this journey. For the pure hearts and innocent laughs, to my kids; Omar and Masa, you both are my life, my source of happiness and inspiration to move forward as I wish you both a happy life.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AES	Advanced Encryption Standard
AWS	Amazon Web Services
BaaS	Blockchain as a Service
BC	Blockchain
BFT	Byzantine Fault Tolerance
BTC	Bitcoin's Currency
CA	Contract Accounts
CAIQ	Consensus Assessments Initiative Questionnaire
CCM	Cloud Controls Matrix
CSA	Cloud Security Alliance
CSP	Cloud Service Provider
DaaS	Data as a Service
DB	Database
DBaaS	Database as a Service
DES	Data Encryption Standard
DIV	Data Integrity Verification
DML	Data Manipulation Language
DOS	Denial of Service
EBaaS	Ethereum Blockchain as a Service

EOA	Externally Owned Accounts
ETH	Ethereum’s currency
EVM	Ethereum Virtual Machine
FHE	Fully Homomorphic Encryption
FIFO	First-In-First-Out
FIPS	Federal Information Processing Standard
HE	Homomorphic Encryption
HF	Hash Function
HTTPS	Hypertext Transfer Protocol Secure
IaaS	Infrastructure as a Service
IAM	Identity and Access Management
IBM	International Business Machines Corporation
IDC	International Data Corporation
IND-CCA2	Ciphertext Indistinguishability Under Adaptive Chosen-Ciphertext
IT	Information Technology
LwE	Learning with Error
NIST	National Institute of Standards and Technology
P2P	Peer to Peer
PaaS	Platform as a Service
PAP	Public Auditor Proxy
PDP	Provable Data Possession

PHE	Partial Homomorphic Encryption
PIR	Private Information Retrieval
POR	Possibility of Retrieval
POS	Proof-of-Stack
POW	Proof-of-Work
P2PKH	Pay To Pubkey Hash
QR	Quick Response
RC	Research Contribution
RLWE	Ring Learning with Error
RO	Research Objective
SaaS	Software as a Service
SDN	Software Defined Network
SEVS	Secure Electronic Voting Systems
SHA	Secure Hash Algorithms
SWHE	Somewhat Homomorphic Encryption
SSHP	Secure Shell protocol
TA	Trusted Authenticator
TPA	Third-Party Auditor
UTXO	Unspent Transaction Output
VM	Virtual Machine

LIST OF SYMBOLS

r	Ciphertext Range
λ	Encryption Security Parameter
$=$	Equality
$\neq$	Inequality
$\equiv$	Equivalence
$\times$	Multiplication
$+$	Addition
$\veebar$	XOR
$\sum$	Summation
( )	Parentheses
[ ]	Brackets
c_i	Ciphertext
Prk	Private Key
Puk	Public Key
c^*	Homomorphic Computed Result in Ciphertext
e	Small Error Terms or Noise Parameter
$Dec_{Prk}(c_i)$	Asymmetric Decryption of Ciphertext c_i using *Prk*
$Enc_{Puk}(m_i)$	Asymmetric Encryption of Plaintext x_i using *Puk*
m	Plaintext

f	Computation Function
$f(m)$	Function of m
k	Known Constant
N	Sequence of Plaintext
r_{sig}	Received Signature
p_{sig}	Signature Pre-Request
h	Hired CSPs
sig	DB Signature of Verification Order
c_{sig}	Comparative Received Signatures
s	PoW Nonce Variable
i	PoW Input Solution
P_x	PoW Produced Record Hash
q	Computation's Query

INTEGRITI DATA UNTUK PENGKOMPUTERAN AWAN DENGAN ENKRIPSI HOMOMORFIK

ABSTRAK

Pengkomputeran awan adalah satu model pengkomputeran baharu yang sumber pengkomputerannya disediakan sebagai utiliti umum yang boleh disewa oleh pengguna melalui Internet secara atas permintaan. Sistem kripto Homomorfik Asimetri diiktiraf dan digunakan sebagai kaedah penyelesaian yang berpotensi tinggi untuk menyelesaikan isu keselamatan data kerana sistem ini ciri privasi dan kerahsiaan data. Penyulitan Homomorfik (HE) adalah satu sistem kripto yang membolehkan pengkomputeran awan melakukan pemprosesan ke atas data yang telah disulitkan. Walau bagaimanapun, skema HE tidak dapat memenuhi kriteria Indistinguishability Under Adaptive Chosen-Ciphertext Attack (IND-CCA2) kerana sifat enkripsinya yang tidak teguh. Tambahan pula, klien (pemilik data) tidak mempunyai kemampuan untuk membuktikan jika data telah dimanipulasi oleh Pembekal Perkhidmatan Awan (CSP) memandangkan CSP berkuasa mutlak ke atas data klien. Klien juga tidak dapat mengesan proses-proses yang telah dilakukan ke atas data setelah data tersebut disumberluarkan *(outsource)*. Klien juga tidak berupaya untuk mengesah kebolehpercayaan CSP dalam melakukan operasi-operasi yang sepatutnya dilakukan ke atas data tersebut. Oleh itu, melaksanakan HE sahaja tidak mencukupi untuk menentang pelbagai serangan integriti ke atas data. Dua skema pengkomputeran awan yang boleh disahkan kesahihannya dicadangkan dalam tesis ini. Skema pengkomputeran awan ditentusahkan pertama adalah yang menggunakan aritmetik modular untuk menghasilkan data dalam bentuk integer medan terhingga yang boleh digunakan untuk mengesahkan pengiraan HE oleh CSP. Prestasi skema yang dicadangkan berbeza-beza berdasarkan kriptosistem yang digu-

nakan. Walau bagaimanapun, berdasarkan kriptosistem yang diuji, skema ini mempunyai overhed penyimpanan 1.5 % dan overhed komputasi yang dapat dikonfigurasi untuk berfungsi di bawah 1%. Skema ditentusahkan kedua adalah berdasarkan teknologi blok rantai (BC), yang dapat memberikan perakaunan yang telus dan tidak boleh berubah pada data pelanggan yang berada di awan. Skema verifikasi berasaskan BC dilaksanakan dengan bermodelkan rangkaian desentralisasi dengan dua model penyebaran yang berbeza: Cloud-BC campuran dan BC ke atas pangkalan data hubungan (RDB) CSP. Kedua-dua model yang dicadangkan berfungsi secara rapat dengan skema HE untuk memberikan kerahsiaan data dan integriti data kepada perkomputeran awan. Kedua-dua model yang dicadangkan berfungsi secara bersama dengan skema HE dapat memberikan privasi, kerahsiaan, dan integriti data kepada pengkomputeran awan. Selanjutnya, skema Campuran Cloud-BC yang dicadangkan adalah berdasarkan matawang kripto awam sebagai perkhidmatan back-end dan tidak memerlukan tindakan persediaan tambahan oleh pelanggan selain dompet untuk matawang kripto yang dipilih. Sebaliknya, dua sistem yang berbeza muncul dari BC melalui CSP-RDB: RDB berasaskan BC tangkas dan RDB berasaskan BC yang selamat. Berdasarkan analisis prestasi dan keselamatan kedua-dua sistem, RDB berasaskan BC yang tangkas sangat disarankan untuk pengunaan kepada pangkalan data throughput yang tinggi. Seterusnya, RDB berasaskan BC yang selamat disyorkan untuk pengunaan kepada RDB yang mengandungi data sensitif dan prestasi daya pengeluaran yang rendah. Lebih-lebih lagi, RDB yang diperbaiki adalah fleksibel dan dapat dikendalikan berdasarkan spesifikasi pemilik data.

DATA INTEGRITY FOR CLOUD COMPUTING WITH HOMOMORPHIC ENCRYPTION

ABSTRACT

Cloud computing is a new computing model in which resources are provided as general utilities that users can lease through the Internet on-demand fashion. However, this technology has numerous data security concerns. Asymmetric Homomorphic cryptosystem has been acknowledged as one of the potential solutions for achieving secure cloud computing since it can provide data privacy and confidentiality. Homomorphic Encryption (HE) is a cryptosystem that allows cloud computing to operate computations on encrypted data. However, HE schemes are non-compliance to indistinguishability under adaptive chosen-ciphertext attack (IND-CCA2) because of their malleable nature. Moreover, the client (data owner) cannot prove if the data has been manipulated by the Cloud Service Provider (CSP) since CSP has absolute authority over the client's data. The client is also unable to trace the processes that are being applied to the data once the data is outsourced. The client can also not verify the CSP's reliability in applying the required operations on the required data. Therefore, implementing HE alone is not sufficient against various data integrity attacks. Two different verifiable cloud computing schemes are being proposed in this work to address these questions. The first verifiable cloud computing scheme uses modular arithmetic to produce verified data to verify the CSP's HE computations over a finite integer field. The performance of the proposed scheme varied based on the underlying cryptosystems used. However, based on the tested cryptosystems, the scheme has 1.5% storage overhead and a computational overhead that can be configured to work below 1%. The second verifiable scheme is based on blockchain (BC) technology, which

can provide transparent and immutable accounting on the client's data in the cloud. The BC-based verifiable scheme is being implemented over a decentralisation network with two different deployment models: Mixed cloud-BC and BC over CSP-relational database (RDB). Both proposed models that work hand-in-hand with HE schemes can provide data privacy, confidentiality, and integrity to cloud computing. Furthermore, the proposed Mixed cloud-BC scheme is based on public cryptocurrency as a back-end service and does not require additional setup actions by the client other than a wallet for the chosen cryptocurrency. On the other hand, two different systems emerged from the BC over CSP-RDB: agile BC-based RDB and secure BC-based RDB. Based on both systems' performance and security analysis, the agile BC-based RDB is highly suggested for the high throughput database. On the other hand, the secure BC-based RDB is recommended for RDB that contains sensitive data and low throughput performance. Moreover, the improved RDB is flexible and can be operated based on the data owner's specifications.

CHAPTER 1

INTRODUCTION

The demanding needs of modern computing have prompted many enterprises to scramble to outsource their data solutions to Cloud Service Provider (CSP). CSP provides different services to raise performance efficiency, easy maintenance, reprovisioning of resources, and cost-saving to the adopters (Badshah et al., 2019). Thus, enterprises can conveniently store, maintain, manage, and backup data files remotely from different geographic locations (Shaffer et al., 2019). Several established CSPs are developing innovative products to remain competitive. The number of businesses has almost doubled in the last few years, and it is anticipated that by 2025, more than 90% of corporate workloads will be in the cloud (Globenewswire, 2020). Apart from cloud storage, cloud computing has recently gained traction as a highly sought-after cloud service. At the same year, the cloud computing industry is projected to reach a value of US $832.1 billion (Marketsandmarkets, 2020).

1.1 Introduction to Cloud Computing

While the fundamental structure for cloud computing is cloud storage that benefited the masses, cloud computing is more valuable to businesses since it provides better collaboration, efficiency, and transparency (Novais et al., 2019). As a result, many enterprises invest in cloud computing to benefit from scalable information, technology services that offer cost-saving, and agile technology (Abdel-Basset et al., 2018).

The cloud computing ecosystem consists of cloud users, CSPs, and the network infrastructure that connects the users and the CSPs. This simple ecosystem provides software, infrastructure, and a platform as a service. Mainly these services provided through the Internet using multi-tenancy technology and resource virtualisation techniques. According to the National Institute of Standards and Technology (NIST) definition Mell et al. (2011) of cloud computing, there are five essential characteristics, three service models, and four deployment models associated with the technology adoption. Typically, there is an agreement between the cloud users and CSPs to agree on the type and quality of the service provided by CSPs. Cloud computing platforms keep the client data inside cloud databases (DB), whether structured or unstructured. The structured data is attributed to being inside the cloud relational databases (RDB), the most famous renewed DBs available (Saravana et al., 2021). In comparison, unstructured data is kept in non-relational databases (NoRDB), most popular in social media content, photos and videos (Dechev et al., 2019).

The continuous pressure to reduce operating expenses has prompted companies to adopt cloud technology to increase profits. Cloud computing offers attractive financial savings in IT costs, in which most of the infrastructure cost is transferred to the CSPs, and users only pay for what is used (Mather et al., 2009). The cloud computing unique features, such as no initial capital investment, pay-per-use, flexibility, accessibility, lower operating costs, easy deployment, accelerated provisioning, scalability, guaranteed service stability, low-cost disaster recovery, minimised business risks, and others, have propelled this technology to rapid deployment and acceptance by different entities (Ali et al., 2015; Phaphoom et al., 2013; Subashini & Kavitha, 2011; Zhang et al., 2010).

Even though the outlook for cloud computing is positive, the security risks associated with cloud computing should be appropriately studied. The issue lies in the principle of cloud computing, where enterprises need to delegate the task of protecting their data to CSP (Ramachandra et al., 2017). Whether the data stored in RDB or NoRDB, data sovereignty is lost once the data is stored in a remote CSP. This absence of control for data security introduces data protection problems.

1.2 Motivation

Cloud computing helps enterprises become more competitive through on-demand scalability, pooled shareable resources, customised self-service, agile computing platforms, and reliability in data accessibility. With cloud computing technology, IT departments are taking advantage of economies of scale. In addition, the move to the cloud has freed existing infrastructure and resources for more strategic tasks (Sabbah et al., 2019). All the advantages of cloud computing have made it a highly sought technology in the market. The figure shows the value of the global cloud computing market from 2021 to 2030, which is expected to reach US $1,614.10 billion at a compound annual growth rate of 17.43% from 2022 to 2030.

Even with all the advantages of cloud computing technology, it still faces some challenges, according to the International Data Corporation (IDC) 2018, the most prominent of which is DB security. As Tabrizchi and Rafsanjani (2020) survey showed, the cloud database's internal threat is one of the main problems delaying many companies in adopting this technology.

Figure 1.1: Cloud Computing Market Size, 2021 to 2030 (USD Billion) (PrecedenceResearch, 2021)

According to the Computer Security Institute (CSI) Survey, it was found that the percentage of internal violations is more severe and costly than its foreign counterparts (Richardson & Director, 2008).

More particularly, as stated in Panda Security (2021) report, the current average annual cost of an internal attack comes in at US $13.7 million per year for risks that take more than 90 days to resolve. In contrast, those that take less than 30 days to fix cost an average of US $7.12 million.

Internal threats are said to be the leading cause of more than 60% of data breaches (Hunker & Probst, 2011). As demonstrated by Cloud Security Alliance (CSA) analysis, for the fourth time in a row (Cloud Security Alliance, 2011, 2013, 2017, 2020), data breaches topped the list of threats in the cloud (see Figure 1.2). Data is considered breached once its information is disclosed, manipulated, or used by unauthorised parties.

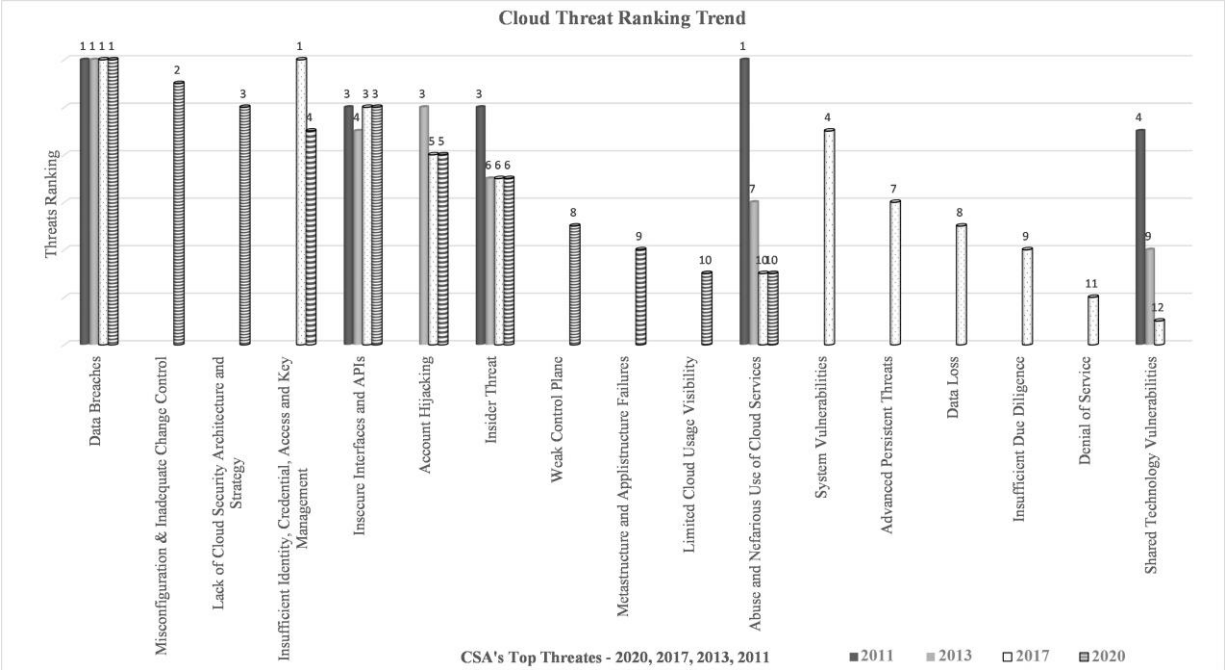

Figure 1.2: Cloud Threat Ranking Trend Analysis
Adopted from: (Cloud Security Alliance, 2011, 2013, 2017, 2020)

This weakness could be explained by the cloud metastructure in which data is processed in the cloud, which may be physically located at any of the data centres worldwide linked as regulated by the Custom Solutions Group (CSO) (Chen & Zhao, 2012). In summary, the responsibility for data is transferred entirely from the client's hands to the burden of CSP. The central managing manner creates fear in users for losing control of their data. The reason is that insiders know the system and attack valuable records, while outsiders steal what they have access to.

Searching for a trusted CSP is not an easy task since cloud computing management is centralised control and cannot guarantee the reliability of their employees (Anciaux et al., 2007). Also, Pearson and Benameur (2010) indicates that the CSPs themselves have the right to amend terms and conditions that may affect client data whenever they want.

Data privacy, confidentiality, and integrity is part of security risk assessment. Therefore it is crucial to ensure the potential risks is alleviated. Privacy refers to the access control that the clients have over their data. Confidentiality means only authorised parties can access the data. Furthermore, data integrity refers to the assurance of data consistency over its entire life cycle.

When data transfer to CSP, standard encryption techniques protect stored data and the operations involved. However, problems arise when there is a requirement to perform additional computations on stored data which definitely requires decrypting process before proceeding to any functions. As a result, several researchers pointed out that Homomorphic Encryption (HE) is the best way to maintain data privacy and confidentiality while processing data on external servers (Acar et al., 2017; Ibtihal & Hassan, 2020; X. Liu et al., 2016). Thus, to allow the CSP to perform different computations on client's encrypted data at their request with complete confidentiality and privacy. For example, in processing sales patterns, without disclosing raw data or seeing its contents (Sultan & Yasen, 2018).

HE is a mechanism of converting data into ciphertext in which the process is capable of performing operations on encrypted data without access to the private decryption key; the data owner is the only one in control of the private key. Once arithmetic operations are performed on encrypted data, the same results can be achieved for raw data. Therefore, HE allows corresponding third parties to compute complex mathematical operations over encrypted data without deciphering them, making secure computation delegation to third-party possible. In addition, HE cryptosystems are malleable in design which can perform multiplication or/and addition operation. Depending on the

supported homomorphism features, HE cryptosystems can be divided into three categories, Partial Homomorphic Encryption (PHE), Somewhat Homomorphic Encryption (SWHE), and Fully Homomorphic Encryption (FHE) (Fun & Samsudin, 2016).

The first process of HE is key generation *KeyGen*, where the data owner creates the public-key pair (a public key *Puk* and a private key *Prk*). The next is the encryption process *Enc*, which involves applying the encryption algorithm onto the data $c = Enc_{Puk}(m)$ before sending it to the cloud server. At the cloud server, the *Puk* and the encrypted data are stored in a database. When instructed by the client, the cloud server performs the requested calculation on the encrypted data before sending the result back to the client in its encrypted form. This is known as the evaluation process, *Eval*. With the corresponding *Prk*, the client can process the decryption function, *Dec*, to recover the plaintext. To sum up, HE has four primary operations: *KeyGen*, *Enc*, *Eval* and *Dec*.

Works of literature have proven the quality of various HE cryptosystems in providing outsourced storing and computing services with the utmost privacy and confidentiality (Alanwar et al., 2017; B. Kumar et al., 2016; Maral et al., 2016). All HE cryptosystems achieve data confidentiality by giving the public key information to apply the assigned mathematical operations. Also, these cryptosystems have proven their competence in preserving data privacy by not allowing unauthorised parties to modify the data. However, the HE cryptosystems were not successful in verifying the data integrity, so the primary pursuit of many studies was to find a method that could be linked with HE to ensure the integrity of the processed data, as shown in Section 2.3.2.

Although HE protects data from attackers, it does not prevent data from falling under exchange or elimination attacks (Li et al., 2017). Furthermore, HE is malleable in nature, which makes it non-compliance to the indistinguishability under adaptive chosen-ciphertext attack (IND-CCA2) security notation (Döttling et al., 2012). Subsequently, HE does not impose security commitment on CSP over the outsourced computations. Also, the client has no authority to verify the accuracy of CSP activities, in which the CSP can purposely manipulate the calculation or miscalculate the data due to negligence. Therefore, data integrity is at stake with only HE itself versus centralised cloud data management.

1.3 Problem Statement

In a cloud computing environment, control delegation can expose client data to several internal attacks. Such vulnerabilities endanger data confidentiality, privacy and integrity. Therefore, HE cryptosystems have been proposed for different privacy and confidentiality preserving protocols. Furthermore, these cryptosystems have the valuable property of enabling anyone to execute computations over encrypted data. Nevertheless, if such data is important enough to protect its privacy and confidentiality, data integrity would also be worth protecting simultaneously in many scenarios.

The following illustrates that the application of HE to the data processed in the cloud computing servers is not sufficient to maintain the integrity of this data.

HE is malleable in nature, making it impossible to achieve IND-CCA2 security notation (Yi et al., 2014). The IND-CCA2 attack model can be divided into two parts. The first ciphertext indistinguishability (IND), where the adversary cannot learn any

information about plaintext m_i underlying a challenge ciphertext c_i. The second is adaptive chosen-ciphertext attack (CCA2), the adversary is given the public key plus the access to a decryption oracle, who can use the decryption oracle even after the challenging ciphertext is given, as long as the adversary does not ask for the decryption of the challenging ciphertext (Döttling et al., 2012).

Clarification of this case depends on the HE mechanism; the client startup the coding operation by generating the keys *KeyGen*. Whenever the public key *Puk* is ready, the client will use it to encrypt *Enc* the data and send these resulted ciphertexts $Enc_{Puk}(c_i)$ to the cloud servers. As such the CSP will have ciphertexts c_0 and c_1, and has received a challenge $Enc_{Puk}(c_b)$ from which it must be established if $b = 0$ or $b = 1$. The CSP may therefore asks the encryption of some known constant k, and consequently deploy the homomorphic features of the cryptosystem to calculate $Enc_{Puk}(c_b) \times Enc_{Puk}(k) = Enc(c_b + k)$. The CSP then relays a decryption query to the oracle in order to understand $c_b + k$, and can simply establish which of c_0 or c_1 is the challenge.

Therefore, HE cryptosystems alone do not guarantee data integrity. Besides, cloud computing is a third party controlling client data without oversight. Data integrity can still be compromised by CSP and go undetected.

Unlike confidentiality and availability, once integrity is compromised, there is no way to restore the original data. Therefore, data integrity needs to be enforced on such outsource computations.

1.4 Research Questions and Objectives

The Research Questions are further listed as follows:

1. How to incorporate data integrity in homomorphic encryption calculations to trace the CSP operations step by step?
2. What is the suitable data integrity scheme for each type of homomorphic encryption?
3. How to use immutable ledger technology as one of the data integrity schemes for homomorphic encryption?
4. What is the effect of incorporate data integrity to the homomorphic encryption calculations cost?

The research aims to develop a verification scheme that can verify the HE calculations performed by CSP. Integer-based HE has been widely researched and used. Therefore the first objective is to look at a specific method to address integer-based HE efficiently. At the same time, the second objective proposes a scheme that can work on any HE (not necessarily on integer-based HE only). This goal also aspires to de-centralise the CSP and keep data on the blockchain (BC).

Next, the third objective achieves a method for tracking cloud DB signature in a distributed manner and not relying on the central CSP authority using the core components of the BC technology. Finally, the fourth objective is to justify the functionality against the cost of the proposed schemes.

1. To achieve computations integrity on integer-based homomorphically encrypted data using modular arithmetic.
2. To provide a verification mechanism in a distributed manner that achieves computations integrity on homomorphically encrypted data using blockchain network.
3. To provide a verification mechanism applied in the cloud relational DB in a distributed manner that achieves computations integrity on homomorphically encrypted data using blockchain technology.
4. To gauge the overhead of implementing the proposed schemes on homomorphic encryption calculations.

1.5 Research Scope

This study focuses on developing data security in cloud computing and ensuring the integrity of the CSP's activities on client data. For cloud computing, the threats to data integrity can be many and varied. However, this work is focused on data integrity against CSP manipulation activities. The main aim of the design of the proposed schemes is to ensure the confidentiality and privacy of data and the integrity of the processes applied to the data to meet the client's requirements. The first design uses modular arithmetic to effectively fulfil the applied data computations' integrity over integer finite fields. Simultaneously, the second and third schemes are based on BC technology to ensure the comprehensiveness of computation integrity for all HE cryptosystems.

The research's scope is bounded by the study and proposition of verifiable cloud computing schemes using HE. The proposed new schemes should contain the necessary data security requirements, which are highlighted as follows:

- **Data privacy**: To ensure that the new scheme provides a client's rights to have control over how its data is handled.
- **Data confidentiality**: To ensure that the new scheme does not disclose sensitive client data to any unauthorised entity.
- **Data integrity**: To ensure that the new scheme provides a commitment that the new manipulation is not random and as requested by the client.

1.6 Research Methodology

The study is conducted in five phases: Firstly, the existing literature of verifying CSP computations schemes is studied to exhibit the shortcomings and difficulties. The weaknesses of these schemes are listed and studied. Secondly, the problem statement of this study is identified, and the research gab is highlighted. In the third phase, three new verifiable cloud computing schemes are proposed based on two different methods as follows:

- Firstly, a verifiable cloud computing scheme using modular arithmetic achieves computation integrity for homomorphically encrypted data over integer finite field.
- Secondly, a verifiable scheme of mixed cloud computing with BC technology that reaches computation integrity for all homomorphically encrypted data; by

surpassing the CSP absolute authority over the data and providing immutable verification data.

- Lastly, an enhanced cloud computing DB based on BC technology component scheme provides data integrity for all homomorphically encrypted data and surpasses the absolute CSP computation authority.

To depict the cost and desired properties of the proposed schemes, evaluation in terms of data security and performance is undertaken in the fourth phase. In the fifth phase, discussion and comparison of the proposed schemes findings are made. Figure 1.3 shows an overview of the steps involved in this study, and Table1.1 provides a summary of the five phases of the research methodology.

This research seeks to find ways to verify the operations applied to client data in the outsourced DB in the cloud servers. The first proposal scheme contributes to the goal of having a system that validates the functions assigned to an integer-based HE. At the same time, the second and third proposal schemes based on BC technology find a solution for all the different HE cryptosystems. They also achieve the goal of clients obtaining a non-tamper-proof signature of the external DB with a distributed visualisation that does not depend on the centralisation of CSP authority. All three of these proposals are subject to a feasibility study for their overall cost-effectiveness, flexibility and ease of implementation.

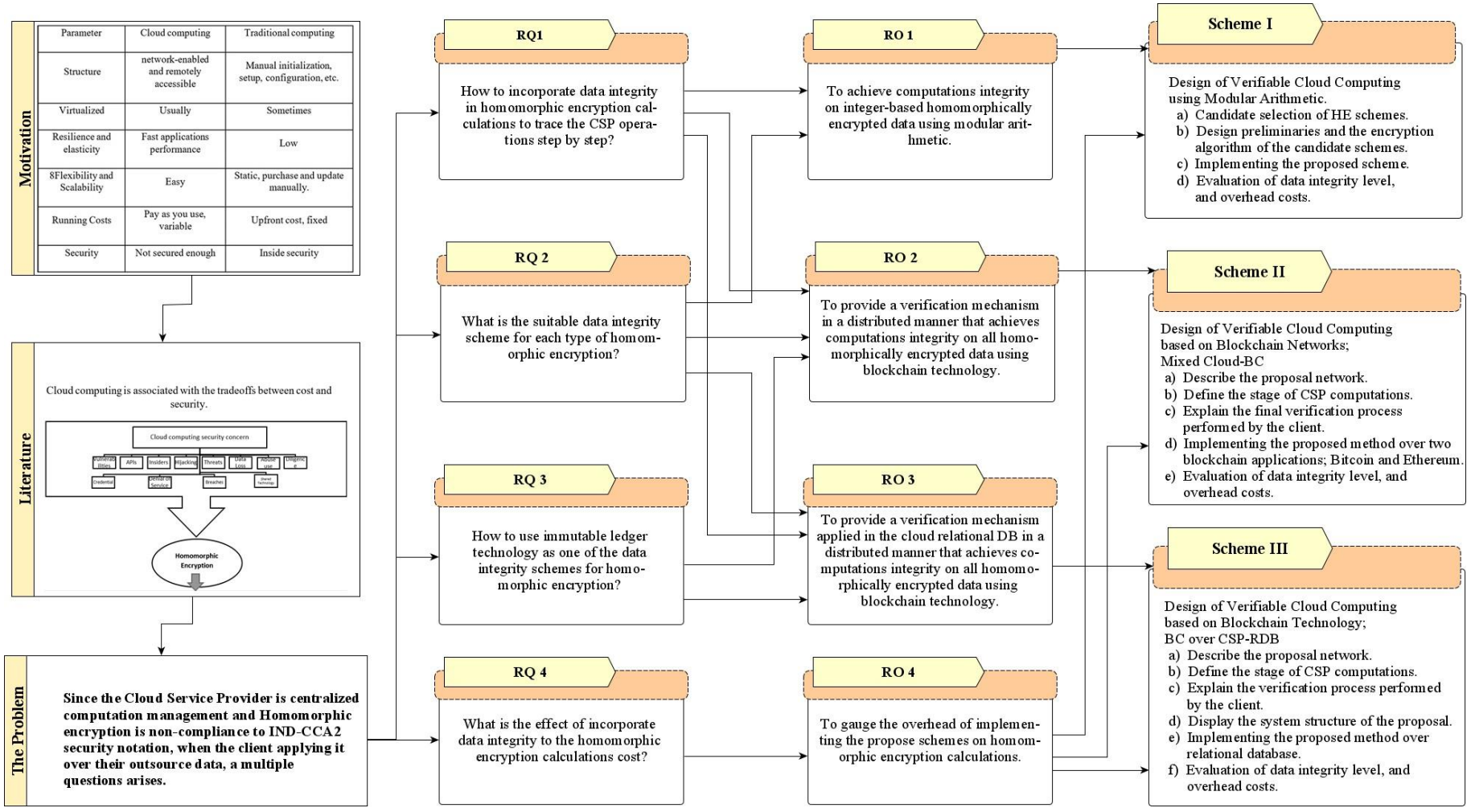

Figure 1.3: Research Methodology

Table 1.1: The Five Phases of Research Methodology

Phase	Details
1. Analysis of existing schemes	Many schemes focused on HE as a method that has been proposed in the literature to enhance cloud computing data security. This stage of the process sheds light on studying such schemes in current literature, addressing their weaknesses, and minimising them by developing new schemes in terms of data integrity.
2. Problem identification and research gab	Despite HE preserves data privacy and confidentiality, HE cryptosystems alone don't guarantee full data security. HE cryptosystem is malleable, and therefore it is not IND-CCA2 secured by design. Consequently, data integrity remains a research challenge.
3. ***Proposed schemes***	Proposed schemes This phase describes three distinct schemes that can provide solutions to the identified issues facing existing schemes in order to achieve the research objectives. The three proposed schemes are verifiable cloud computing using modular arithmetic, cloud computing mixed with blockchain technology, and blockchain over cloud computing relational database.
4. ***Proof of concept***	The proposed schemes that will improve the data integrity in cloud computing are implemented to achieve the research objective.
5. Evaluation and comparison	This phase analyses and discusses the quality efficiency of the proposed schemes by assessing the results. In addition, comparisons are made in the studies between the proposed schemes in terms of data integrity level, general cost, performance and implementation.

1.7 Research Objectives and Proposed Schemes

In order to achieve the objectives of the research, this work generated three independent design schemes. Each one fulfils the integrity of the computations applied to client data. Figure 1.4 shows the relationship of each proposed scheme with each research objective. Based on security countermeasures literature in cloud computing, several approaches and technologies are used to develop security. According to the scope of the research, this work adopted HE, BC technology, and modular arithmetic as counter methods to overcome the research problem. Therefore three independent schemes with different designs and results have resulted. The first design relied on HE along with modular arithmetic operations to provide data integrity for particular HE cryptosystems. The second and third schemes contributed to merging HE with BC.

Thus, producing two completely different proposals in design, results, and even the level of data integrity. In particular, the second research objective requested to provide a verification mechanism in a distributed manner using BC. Consequently, It is led to the design of the second scheme for storing data in BC networks known for transparency and immutability. While the third research objective seeks to provide a validation mechanism implemented in cloud relational DB in a distributed manner using BC, which contributes to the third scheme. The design concept of this scheme is to build a relational cloud DB based on the main BC components. The difference in the design building base for each scheme results in the difference in the consequences of security, performance, and implementation, as is the fourth objective of this research. Accordingly will show how each scheme sought to validate the computations over all the encrypted data and the differences between each scheme.

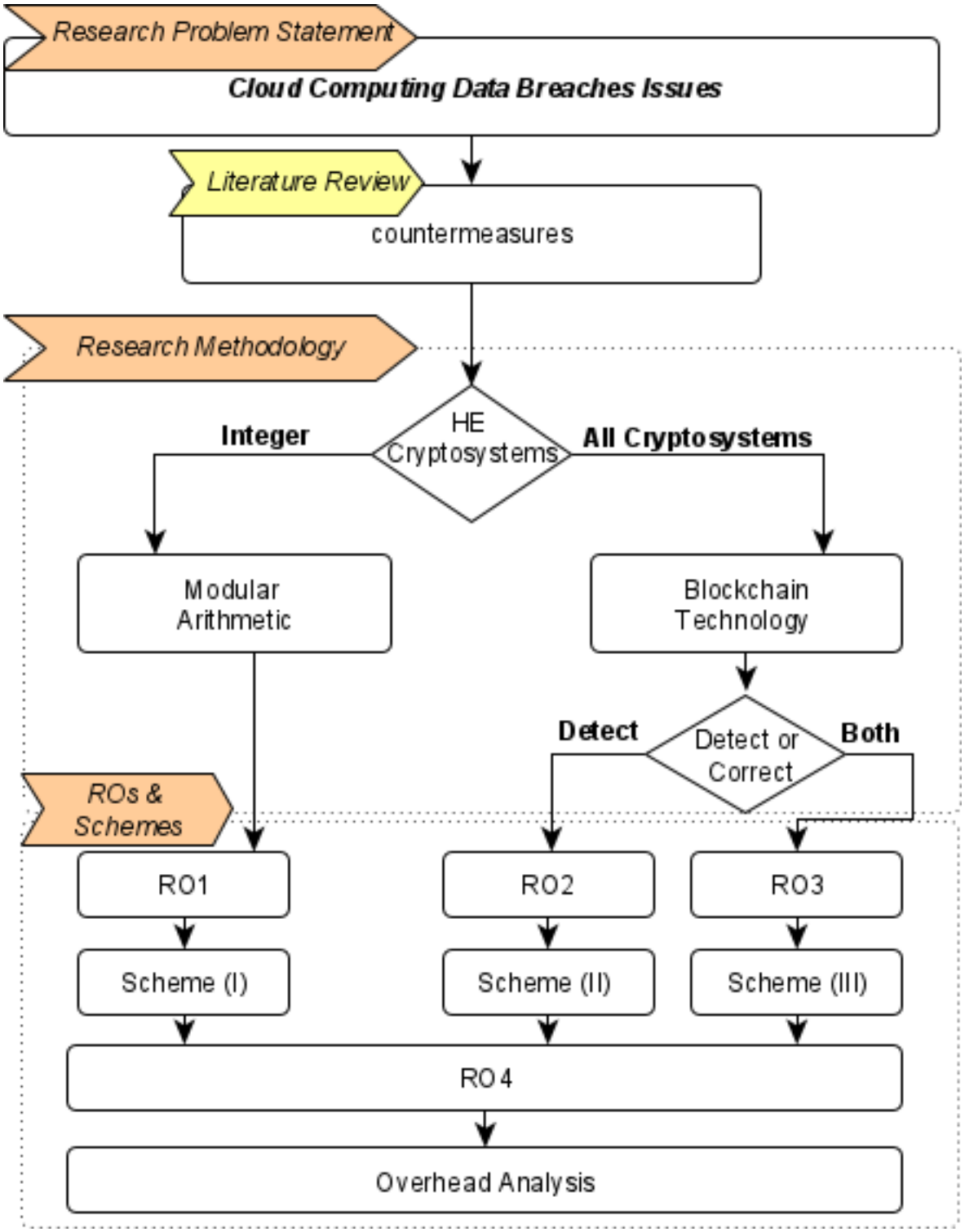

Figure 1.4: The Relationship of the Research Objectives with the Proposed Schemes

1.8 Criteria for Evaluating Proposed Schemes

In this work, the results of the proposed schemes will be compared according to the following criteria:

1. Data Security: the level of data privacy and confidentiality for all schemes is based on the properties of the HE. In contrast, the schemes are varying in data integrity. The research assessment of data integrity for each scheme relies on the trust model based on CSA that evaluate the strength of data security in cloud computing.

2. Performance: the results will be compared with the standard framework to evaluate the performance of cloud computing services. The performance is based on the amount of energy wasted by each proposal scheme.

3. Implementation: the difficulty of building each proposed scheme and its requirements are the main core of comparison.

4. Cost: the conceptual construction of each proposal will give an overall impression of the cost of each proposal separately.

1.9 Research Contributions

The significance of this study is summarised as follows:

1. **The development of the data integrity of HE computation over integer finite field.** The client authorises the CSP to apply various operations to client data without ensuring the integrity and correctness of that calculations. If contra-

dictory are detected on the data, this will affect the client's confidence in cloud technology and ultimately lead to retreat from delegating operations to an insecure intermediary. To solve this issue, a verification cryptosystem based on modular residue is proposed to validate HE computation over integer finite field Z_p^* to be used in cloud computing so that all of data confidentiality, data privacy and data integrity can be enforced during an outsourced computation.

2. **The development of the client's self-verification of his/her data integrity stored in the cloud.** The cloud computing environment has many problems with maintaining data security. The HE schemes ensure data privacy and confidentiality are not violated, but it is insufficient to prevent tamper with data integrity. Homomorphically encrypted data requires a technology that supports its integrity, regardless of the origins of the encryption process. To address this issue, a mechanism based on BC applications represented in Bitcoin and Ethereum is proposed to verify DB integrity stored in the cloud servers.

3. **The development of the central authority of the CSP over the client data.** Centralised cloud computing management can expose homomorphically encrypted client data to many administrative risks that may cause data loss or tamper in cloud-owned DBs. Consequently, the client cannot trust the primary relationship structure with the cloud provider. To solve this problem, Byzantine Fault Tolerance (BFT) consensus is proposed to create a distributed computation network.

4. **The development of the DB layout to enhance the client ability to verify CSP performance with the stored and processed data.** The traditional DB cannot keep track of the processes applied to the data, thus exposing the data to many different attacks. A new DB layout is proposed to solve this issue based on the concept of the BC philosophy.

1.10 Thesis Organisation

The rest of the thesis is organised as follows; Chapter 2 is a general overview of the background, related concepts and the existing schemes. Chapter 3 provides the step-by-step methodology to build the proposed verified cloud computing calculations with the HE cryptosystem. The proposed schemes' proof of concept process will be explained in-depth in Chapter 4. Chapter 5 discusses the evaluation of the applied proposed schemes. Finally, Chapter 6 concludes this thesis with open questions pointing to the future directions of the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Cloud computing has become a desirable significant service for Internet computing users for its multiple advantages (Azad & Navimipour, 2017). Moreover, implementing cloud computing empowers numerous online service offerings to meet different needs (Chiregi & Navimipour, 2018). Thus, cloud computing becomes one of the hottest topics of emerging technology in applications and research. Among the literature, cloud computing is estimated to be one of the essential strategies.

Homomorphic Encryption (HE) is being proposed to enhance cloud computing security in which the data owner can encrypt the data with various HE schemes before sending it to the cloud. Against standard encryption, limiting the cloud's role to simple storage, HE allows a party that holds ciphertexts to perform certain operations on the ciphertexts, which mirror the plaintexts' corresponding operations.

In contrast, because cloud service provider (CSP) is a centralised management approach, HE alone cannot fulfil IND-CCA2 security notions against tampering. Therefore, client data is exposed to administrative risks that may cause data loss or undisclosed manipulation in the database (DB). To incorporate data integrity into HE, modular arithmetic and blockchain (BC) technology have been proposed as solutions.

This chapter studies the cloud computing benefits and security challenges, HE categories implementations and challenges, and finally, the BC security benefits. Firstly, a conclusive study on cloud computing and a comparison between cloud computing and traditional IT computing, and the illustration of security concerns over various cloud computing domains and how related work promise security for outsourcing sensitive data are presented in Section 2.2. Section 2.3 discusses the need for HE with different features to preserve data confidentiality and privacy in the cloud and the different challenges facing the HE. Section 2.4 defines modular arithmetic and clarifies the research papers it uses in data security. Section 2.5 presents an anatomical explanation of BC technology's components and its uses in cloud security. Finally Section 2.7 concludes the chapter.

2.2 Cloud Computing Definition and Features

The invention of the personal computer caused the first IT revolution between the 1980s and 1990s. In the early 21st century, the second IT industry revolution has shaped after the invention of the Internet and smartphones. In recent years, cloud computing has developed the third global IT industry revolution (Krutz & Vines, 2010). In 2006, Amazon introduced Elastic Compute Cloud, which ensures utility computing resources and infrastructure management to the consumer as needed (Espadas et al., 2013). The same year, Google presented the "cloud computing" concept. After that, most IT major enterprises jumped to the "cloud" in order to produce various kinds of cloud services. In the last quarter of 2007, Google and IBM started to promote cloud computing research plans in different Universities. As a result, cloud computing technology is constantly increasing, with other concepts of flexibility and computing utility

on-demand usage (Armbrust et al., n.d.). In 2010, the number of major IT enterprises worldwide had reached half a million for using the cloud computing platform to replace the traditional system. Others have allowed cloud computing as a strategic core (Marston et al., 2011).

Despite all this extensive publicity and use, the definition of cloud computing was somewhat vague (Vaquero et al., 2008) and required many years to standardise a complete definition by NIST (Mell et al., 2011). The term cloud computing has been defined as follows:

Definition 2.1. Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

NIST associated Cloud Computing with five essential characteristics, three service models, and four deployment models provided by a distributed shared pool of configurable IT resources on-demand, through complex networked infrastructure, on a pay-per-use or subscription basis. They are summarised in visual form in Figure 2.1.

The roles of the core parties involved in cloud computing technology consist of the client and CSP. The client could be an enterprise or end-user who maintains a business relationship with the cloud provider.

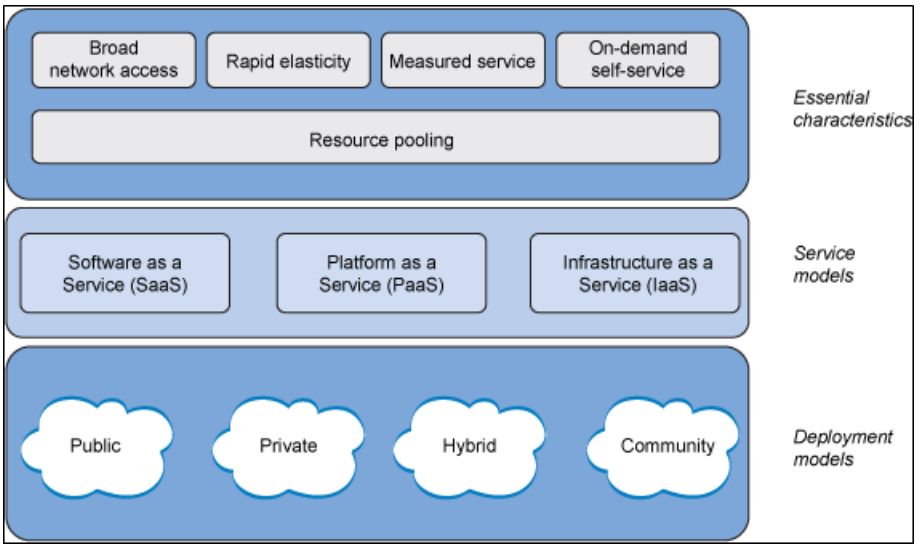

Figure 2.1: NIST Visual Model of Cloud Computing
(Naveen et al., 2016)

It allows him/her to use various services. At the same time, CSP could be any entity responsible for making a service available to interested parties over a network.

Cloud computing technology successfully makes an example shift from traditional IT in the building structure and the maintenance capabilities of distributed computing systems, multi-processors, virtualisation technology, and distributed data storage. Upon the literature review (Cloud Security Alliance, 2017a; Masud & Huang, 2012; Nandgaonkar & Raut, 2014; Velte et al., 2010) different researches show the benefits of cloud computing over classical IT infrastructure in different factors which summarised in Table 2.1.

Metastructure is a crucial difference between cloud computing and classical IT infrastructure. CSP is responsible for managing cloud computing systems remotely, running the cloud smoothly, providing resources as needed, and ensuring security measures. In contrast, traditional computing management is within the company and requires the hiring of professional staff. Cloud computing contributes to double up on