THESIS PROPOSAL

TOPIC

SECURING CLOUD DATA ON MULTIPLE INFRASTRUCTURE USING ERASURE CODING, DISPERSAL TECHNIQUE AND ENCRYPTION

BY

FRIMPONG TWUM

INTRODUCTION

1.0. Background of Study

The need for Information Technology (IT) in industry has never been as high as today. However, organisations budget for IT is much stringent than ever. Businesses today invest in Information Technology and Information Systems (IS) to achieve one or several of the following strategic objectives: Improve decision making process; Survival; Customer and Supply Intimacy; Competitive Advantage over rivalry; New product, services and business model; and Operational excellence (Laudon and Laudon, 2010). To be competitive in this era, organisations ought to be strategic in their investments in IT and IS to ensure it supports key business processes that deliver the required outcome. This challenge has necessitated that organisations always find and adapt the most efficient and cost effective IT and IS solution. As a result, businesses today and even individuals are increasingly moving away from owning and managing their own IT infrastructures to using Cloud Computing technology which provides them with the ability to outsource their IT needs for example, Infrastructure as a service (IaaS), Platform as a Service (PaaS), or Software as a Service (SaaS) to the care of a remote third party provider on a subscription bases at a per usage fee on-demand (Khan and Yasiri, 2016).

Using Amazon Cloud Computing Infrastructure for instance, SaaS providers avoid having to invest heavily in IT infrastructure costs (for example, owning servers and setting up of data centres). SaaS is a software deployment model that enables the distribution of software over the Internet for subscribers' utilisation (Amazon Web Services, 2010). SaaS, which provides remote software solution to tenants (individuals or organisations) on either shared or un-shared environment on demand is what is usually referred to as cloud computing (Barry, 2000). SaaS applications offer users a centralised networked data storage facility that is hosted, maintained and managed by the Cloud Service Provider (CSP). Although cloud computing comes with many benefits, adopting cloud computing technology also means relinquishing control over key business data assets to a third party, the CSP. This is a cause of anxiety to the cloud customer and hence raises an issue of trust which is a major concern frustrating SaaS adoption.

The challenge therefore is how to create a secure and vigorous data architecture that can deal with both internal and external threats posed to the cloud tenant data resources and also satisfy subscribers legitimate concern of losing control over key business data asset to a third party

CSP (Chong et. al., 2006; Rao and Selvamani 2015) - the focus of this research. SaaS incorporates IaaS and PaaS.

1.1. Overview of Cloud Computing

The National Institute of Standard and Technology (NIST) define cloud computing as:

"A model for enabling convenient on-demand network access to a shared pool of configurable computing resources (e.g. Networks, servers, storage, applications, and other computing services) that can be rapidly provisioned and released with minimal management effort or service providers interaction" (Mell and Grance, 2011).

In other words, Cloud Computing is simply the delivery of computing services over the Internet.

Cloud Computing allows organisations or individuals to use software, hardware, and data storage facilities that are hosted and managed by third parties at remote locations. It is a technology with the potential to enhance collaboration, agility, scalability, availability of computing resources, and also give organisations or individuals the added benefit of reducing their Information and Communications Technology (ICT) cost through optimised and efficient computing (Khan and Yasiri, 2016). Examples of existing cloud services include:

- Online file storage services such as Dropbox, Google Drive
- Social networking sites for example Facebook
- Online business applications such as online salary processing systems

Cloud computing give user's accessibility to their data resources from any location where they have access to the Internet and this enhances productivity and work efficiency. Cloud computing gives individuals or organisations the ability to share resources as data storage space, networks, computing processing power, and specialised corporate hardware and software applications. Music, video clips, photographs that people use to keep on their own personal computers and other mobile computing storage internal devices are now all been stored on servers owned and managed by third party CSP's such as Amazon, Google, Facebook.

With cloud computing, a subscriber has to sign-up for a service model (SaaS, PaaS, or IaaS) and subscribe to one of the cloud deployment models (private cloud, public cloud, community cloud, or hybrid cloud) on a subscription-based payment agreement with a CSP to access the storage space or other computing resources required on-demand (Goswami and Singh, 2012).

'Public' cloud computing services (public SaaS) has been in use for years for example via client webmail services such as Yahoo, Hotmail, Gmail, and also most recently via online data storage services as DropBox, Google drive, OneDrive, and Box. These services have generally been offered at no cost to the subscriber although they are paid for in disguise for instance via advertisements presented to the subscriber online while using the services, or via the selling of subscriber's personal data and online surfing behaviour patterns to interested organisations (Barry, 2000).

Client webmail software is used to access electronic mails stored in a client's email storage accounts located on remote systems that usually belongs to other people or organisations (known or unknown) and are also stored at remote unknown locations. The remotely known or unknown individual or organisation (SaaS cloud provider) is entrusted with taking care of the subscriber's software and hardware needs, and also most importantly taking care of the subscriber's vital data assets which may include key business financial data, trade secrets, employee records, supplier's contacts, product lines, and customer's data.

This poses serious security threats to the cloud subscriber in terms of data security as the subscriber relinquishes control over the management and maintenance of the data to a third party CSP who might use the data for other purposes for which the data owner permission has not been sought.

Hence, there is a challenge with the technology ability to ensure data confidentiality, data integrity, and even data availability (Rao and Selvamani, 2015; Ali et. al., 2015). Therefore, the question of the security of the data and other resources outsourced for cloud storage remains to be a major hindrance to the technology widespread adoption (OpenCirrus, 2017).

1.2. Problem Statement

Cloud computing technology is an invention in the ever changing computing technology that has come to save organisations from setting up, owning, and maintaining high cost computing equipment and other ICT infrastructure. Benefits includes cost savings (in terms of hardware, software, personnel, etc.), ability to access resources from anywhere at any-time provided there is an Internet enabled device and connectivity to the Internet, and paying per usage among others. Cloud computing gives users huge storage capacity via storage facilities hosted on the Internet that are usually owned and managed by third party Cloud Service Providers (CSP's). These storage facilities usually are publicly accessible referred to as Public Cloud, or may be configured for an individual subscriber's private use referred to as Private Cloud, or configured explicitly for a group of organisations usage referred to as Community Cloud, or may be a composite of two or more of the specific cloud deployment models referred to as Hybrid Cloud. The CSP has access and control over the data whether encrypted or un-encrypted as the responsibility for the data maintenance (such as data backups and data restore) is usually mandated to them.

Although the cloud tenant outsourcing its IT functions enables them to focus on their core business processes, they also put their vital data resources at risk in the hands of the third party provider who may use it for their own gains. For example, selling the data to a competitor, or using it for other purposes other than has been agreed.

Cloud computing at the onset came with security challenges as a result of its resource pooling and multi-tenancy characteristics where multiple customers share the same resources, same application, same databases or in some cases same tables (Youssef and Alageel, 2012; Khatri et. al, 2013). As an example, a cloud provider computing resources may be pooled to serve multiple subscribers and this may put data at risk of getting into unauthorised hands through accidental or intentional disclosure. Thus, the CSP may accidentally or deliberately leak data or other vital resources to a competitor as they serve multiple subscribers (Khan and Yasiri, 2016; Shapland, 2017). A study by Trigueros-Preciado (2013) found cloud computing security to be of a supreme concern to subscribers and this discovery in 2017 remains unchanged as confirmed by Ahmed (2017) study. The Treacherous 12 (2017) survey identifies data security breaches such as:

- The two Yahoo! data breaches reported in September and December 2016 (affecting 3 billion user accounts, leading to a drop of $350 million in the acquisition price of Yahoo! which was earlier valued at $4.8 billion) (McMillan and Knutson, 2017),
- Data loss such as malicious CSPs or malicious users intentionally corrupting the user's data inside the cloud by modifying or deleting (Chauhan, 2015; Sailaja and Usharani, 2017),
- Malicious insiders such as the theft of 1.5 million T-Mobile customers' data by an employee at their Czech offices (thehackernews, 2016),
- Denial-of-Service (DoS) attacks such as the Australian Bureau of Statistics denial of service (ABS, 2016) as concerns of cloud computing security.

Another issue that arises from the use of Cloud Storage as a Service is the use of customer data for marketing and personal profiting such as leaking it to competitors (Chauhan, 2015).

Ahmed (2017) study established cloud computing poses security threats to the subscriber in terms of:

- Who has access to the data/resource (accessibility)
- What other use is the data/resource been used for (usage)
- Where the data/resource is located (location)
- Who has ownership over the data/resources outsourced to the cloud (ownership)
- And also ensuring accuracy of the data outsourced for cloud storage (accuracy)

And these raise questions as follows:

- How can cloud data be secured to prevent unauthorised access?
- In what ways can a cloud subscriber prevents their data from been used for other purposes by the cloud provider?
- In what ways can the cloud subscriber ensure that their outsourced data is not vulnerable as a result of the data location since different countries have different data privacy laws?
- In what ways can the cloud subscriber ensure that they have sole ownership of their data outsourced for cloud storage?
- In what ways can the integrity of data outsourced for cloud storage be maintained?

In relation to ownership there is the risk in terms of what happens to the data on contract termination or in the event of conflict between the cloud subscriber and the cloud provider. For

example, when a CSP refuses to grant a subscriber access to their data in the event of a dispute over say the subscriber's subscription payments.

With the issue of location, accessibility, and usage of the data resource, cloud computing distributes data across servers setup and managed by CSPs across the globe and this makes it difficult for the cloud subscriber to find in which country(s) their data is been stored, who has access to the data, and for what unauthorised use (Rao and Selvamani, 2015). Finally data outsourced for cloud storage can be altered in transmission by man-in-the-middle attack or modified inside cloud provider's storage facilities by a malicious insider attack (Sailaja and Usharani, 2017).

These issues are making it unattractive for organisations and individuals to subscribe to cloud services. Although traditional counter security measures such as using encryption techniques (for confidentiality), using hash functions (for integrity), and using firewall, anti-virus, intrusion detection and prevention systems (for availability) have been employed, they have been inadequate to securely protect vital organisation data against attacks. Malicious attackers have found ways of going round them to compromised vital business data asset using network security attacks as Dos/DDoS, U2R attack, R2U attack, Probing attack, MITM attack, Message replay attack, and Brute-Force analysis attack (thehackernews, 2017).

According to Wang (2009), cloud computing technology distributes data on multiple servers belonging to a single CSP but the challenge as noted by Ahmed (2017) is implementing a distributed protocol architecture that assures of a robust secured cloud data security in a defence-in-depth design.

1.3. Aim of Research

This research therefore seeks to propose a cloud data security solution framework and implement an algorithm whereby data outsourced for cloud storage will first be sliced into chunks of data fragments and then encrypted on the subscriber's gateway system before being distributed to multiple different CSP's storage nodes (storage servers).

1.4. Specific Research Objectives:

i. To enhance security of data outsourced for cloud storage by ensuring the data is useful to only the data owner
ii. To propose a cloud data security solution framework and an algorithm for securing cloud data that alleviates the cloud subscriber's fear of the data/resource's privacy, usage, location and ownership.
iii. To implement and test the propose cloud data security solution framework and algorithm.

1.5. Research Questions

i. How can one ensure data outsourced for cloud storage is useful only to the data owner?
ii. Can a proposed cloud data security solution framework and algorithm alleviate the cloud subscriber's fear of the data/resource's privacy, usage, location and ownership?
iii. How does the proposed algorithm compare in terms of strength and performance?

1.6. Justification of choice of the study

As outlined in the problem statement, cloud computing comes with numerous benefits but also faces several security issues especially in terms of data privacy, data integrity, and data availability. Cloud computing characteristics of resource pooling, multi-tenancy, on-demand self-service, broad-network access, and rapid elasticity introduces new security threats in terms of data accessibility, data ownership and data accuracy and hence demand new approaches for dealing with them.

Traditional counter security measures have been found to be inadequate for dealing with cloud security issues, an example been that encrypting data before sending it to a single CSP does not protect the data from been decrypted, deleted, or altered (O'Reilly, 2017).

A survey conducted by the Cloud Security Alliance CSA in 2016 identifies twelve security concerns of cloud computing including data breaches, data loss, malicious insiders and Denial of Service among others (The Treacherous 12, 2017). Other cloud security issues includes: the cloud provider profiting from using the subscriber's data entrusted in their care for advertising, or using the data to learn more about the subscriber for their own interest or gains. Although research suggests that cloud security threats from multi-tenancy architecture have been reduced

by major CSPs such as Amazon and Microsoft, the threats are still real especially for smaller CSP's (Shapland, 2017).

In addition, although different countries have different privacy and security laws, acts, and regulations that govern the protection of data for example, the Asia Pacific Economic Cooperation (APEC) privacy framework, the Organisation for Economic Corporation and Development (OECD) privacy framework and the European Economic Area (EEA) data protection laws, the actual responsibility of ensuring that data and other resources outsourced to the cloud are secured and protected against data loss, damage, misuse usually rest with the custodian of the data - the CSP (CSA, 2011; OpenCirrus, 2017).

However, given that data is the life blood of every serious organisation and that with cloud computing the subscriber's vital data asset is to be outsourced to third party organisation, it is critical the cloud subscriber take key interest in ensuring the safety of their data been outsourced for cloud storage. Hence, this study is of the same view with Fahmida (2016) that with cloud computing, the data owner (cloud tenant) even bears paramount responsibility in ensuring security of its data than the custodian of the data. Especially where critical business data such as trade secrets, financial data, employee data, or health data are been transferred for cloud storage. This assertion is more critical because when it comes to cloud computing service provision there is a chain of inter-dependency of services provisioning and hence tracing data leakage(s) could be extremely difficult. This study seeks to propose secure data security architecture and system that ensures data is useful only to the owner.

1.7. Methodology

The study will be carried out following the design research methodology and will employ a plan-driven incremental development and delivery method in which the sub-systems and its components specifications are planned ahead but the design and development processes are carried out as a series of increments to deliver the system as increments. The software increments are programmed and provision to users for their rapid feedback. The subsystems are then integrated and tested to deliver the system. The proposed cloud security solution framework will be developed into software in an experimental lab-setup using JAVA, SQL, and PHP.

The major steps involved in the development process are outlined.

i. To apply *erasure coding technique* to first sliced data objects outsourced for cloud storage into chunks of data fragments (Tashi and Ponsam, 2016; Plank, 2013).
ii. To apply an *encryption algorithm* to encrypt the chunks of data fragment (Goswami and Singh, 2012).
iii. To apply *data dispersion technique* to shuffle the encrypted data fragments and distribute to multiple CSP's storage nodes (servers) (CSA, 2011).
iv. To ensure efficiency especially during data retrieval as different CSP's storage nodes host the data fragments and hence may be operating at different data rates, a *buffering technique* is used to buffer the data fragments from the fast storage nodes as a waiting mechanism until the data fragments from the delayed storage nodes are received and assembled for onward delivery to the subscriber.

By employing above measures, this study hope to address the issues raised in the problem statement and assure the cloud subscriber the security of their data as the encrypted data fragments will be of no value to the CSP.

v. Finally the study foresees performance to be likely affected as security is strengthened and hence seeks to cater for performance by employing the use of *Metadata server* to keep track of the data fragments and where they are distributed so as to ensure accuracy of the cloud subscriber's data resources (Dell Power Solutions, 2005).

In effect, the study hopes the proposed data security solution framework and algorithm will assure cloud subscribers of the confidentiality, integrity and availability of their data resources outsourced for cloud storage.

1.8. The study organisation

The thesis is organised as follows:

Chapter 1: Introduction- gives brief introduction to the research, and then presents an overview of cloud computing technology, the statement of problem to be address, the overall study aim, the specific objectives, the research questions, and justification for the research.

Chapter 2: Literature Review- presents a review of related research work. The chapter reviews related published articles on the subject and also review related materials from other sources including Internet, Books, Journal Articles, among others.

Chapter 3: Methodology- presents the methodology employed for the study (Design Research), the software process method, and also presents the proposed model. A description of the proposed algorithm for securing cloud data on multiple CSP's storage nodes using Erasure coding, Encryption, Data Dispersion technique, Buffering technique, and Metadata works is presented.

Chapter 4: Implementation and Testing- implements the proposed data security framework and algorithm and test the result to ascertain the algorithm performance and most importantly its ability to assure cloud subscribers of the security of their data resource outsourced for cloud storage.

Chapter 5: Discussions and Results - present an insight into the results and compare the proposed system to other existing related systems in terms of its security strength.

Chapter 6: Conclusions and Recommendations- throw more focus on what the results actually means for both the cloud subscriber and the CSP especially in terms of benefits and suggest recommendations based on the research findings for the cloud subscriber and as well as suggest recommendations for future research work on the subject.