جامعة الشرق الأوسط
MIDDLE EAST UNIVERSITY
Amman- Jordan عمان - الأردن

A Semantic Ontology based Concept for Measuring Security Compliance of Cloud Service Providers

قياس الأمتثال الأمني لمقدمي الحوسبة السحابية مبني على مفهوم الأنتولوجي الدلالي

By

Mustafa Nouman Murad Al-Hassan

Supervisor

Dr. Mohammad Alharibat

This thesis is submitted to the Department of Computer Information Systems, Faculty of Information Technology, Middle East University in partial fulfillment of the Requirements for Master Degree in Computer Information Systems.

Faculty of Information Technology
Middle East University
Amman, Jordan
(August, 2013)

Examination Committee Decision

This is to certify that the thesis entitled "A Semantic Ontology based Concept for Measuring Security Compliance of Cloud Service Providers" was successfully defended and approved on August 25th 2013.

Examination Committee Members	Signature

Prof. Reyadh Naoum

Department of Computer Science

Middle East University

R. NaoUm

Dr. Ahmad Kayed, (PhD)

Associate Professor

Department of Computer Science

Middle East University

Dr. Hussein Al-Bahadili

Associate Professor

Faculty of Information Technology

Petra University

إقرار التفويض

أنا ، مصطفى نعمان مراد ، افوض جامعة الشرق الأوسط بتزويد نسخ من رسالتي للمكتبات أو المؤسسات أو الهيئات أو الأفراد عند طلبها.

التوقيع:

التاريخ: ٣١ /٨ /٢٠١٣

Authorization Statement

I'm Mustafa Nouman Murad Al-Hassan, authorize the Middle East University to supply a copy of my thesis to libraries, establishments or individuals upon their request.

Signature:

Date: 31/08/2013

Dedication

To the light of my life …
My Wife ,
My Son “Ali”
And My Daughter “Ula” ..

Acknowledgement

I would like to express my sincere appreciation to Prof. Reyadh Naoum and Dr. Mohammad Alharibat for their guidance, support and motivation throughout my Master’s Thesis.

Also I’am particularly grateful to Dr. Ahmed Kayed helping and encouraging my efforts during this research.

I owe more than thanks to my family members which includes my parents, my relatives and my wife's family for their financial support and encouragement.

For the past seven months, my son and daughter kept on reminding me to go to office and write the thesis! I’m very grateful to them.

Finally, and most importantly, I would like to thank my wife. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past fifteen years of my life have been built. She was always there cheering me up and stood by me through the good times and bad. For her constant love, support, and remembering to feed me when I couldn't remember to feed myself!

Thank You All..

Abstract

Cloud computing is Internet-based computing, whereby shared resources, software and information, are provided with computers and devices on-demand. It also makes security problems more complicate and more important for Cloud Service Provider (CSP) and consumer than before. International standard organizations issue security-related standards and guidance which can be used in cloud environment such as ISO/IEC 27001. This thesis explores the possibility to measure security compliance for data breaches threat based semantic similarity measure between the documents of standard compliments and CSP response against data breaches threat.

We developed a model that measures security compliance of CSP with the major international standard organization against data breaches threat.

Our model consists of three stages: (1) Extracting ontology concepts of CC threat (2) Extracting ontology concepts of CSP (3) Matching Process among the both ontology concepts. The matching process has done by using semantic similarity measure. Also during our study, we collected and studied many documents and reports that discussed data breaches threat. Then we classified it into group of (Control Area), identify the items that cover each control area. Also tested 5 CSPs to measure their security compliance by collection their data related to each control area; then convert it into text file in order extracting ontology concepts.

Our results were promising (0.885 %) Mean Square Error (MSE) between our measure and human judges.

Keywords: Cloud Computing, Security Compliance, Ontology Concept, Semantic Similarity.

الملخص

تعتمد الحوسبة السحابية على أستخدام الشبكة العنكوبوتية (الأنترنت) ، وبواسطتها يمكن مشاركة الموارد ، التطبيقات ، المعلومات ، يتم توفيرها مع أجهزة الحواسيب حسب الطلب. وهذا بالتأكيد يجعل أمن المعلومات أكثر خطر وعملية مهمة لكل من مقدمي الخدمة والزبائن أكثر من اي زمن مضى. هناك مجموعة من المنظمات الدولية (مثل ISO/IEC 27001) التي تضع المواصفات القياسية لأمن المعلومات يمكن استخدامها في بيئة الحوسبة السحابية. ان في أطروحتنا هذه نستعرض أمكانية قياس الأمتثال الأمني لمقدمي خدمة الحوسبة السحابية مع هذه المواصفات القياسية ، فقد أخذنا قياس الأمتثال الأمني من ناحية تهديد الخروقات الأمنية. القياس الذي استخدمناه هو قياس تشابه دلالي بين مستندات مقدم الخدمة والأمتثال القياسي العالمي من ناحية تهديد الخروقات الأمنية. فقد طورنا نموذج يقيس الأمتثال الأمني لزود الخدمة مع المنظمات الدوالية الرئيسية التي تضع المواصفات القياسية.

هذا النموذج يتكون من ثلاثة أقسام: (1) أستخراج مفهايم للتهديد المطلوب قياسه (2) أستخراج مفاهيم مزود الخدمة (3) عملية التطابق بين المفاهيم المستخرجة. وأستخدمنا قياس التشابه الدلالي في عملية التطابق. وخلال بحثنا هذا جمعنا ودرسنا العديد من المستندات والتقارير التي ناقشت تهديد الخروقات الأمنية وبالتالي صنفناها الى مجموعة من (منطقة تحكم) ، وعرفنا العناصر التي تغطي كل منطقة تحكم. وأختبرنا عدد (5) من مزودي خدمة الحوسبة السحابية لقياس مدى أمتثالهم الأمني من خلال جمع البينات المتعلقة بكل منطقة تحكم ثم تحويليها الى ملف نصي لأسترخاج مفاهيم الأنتولوجي.

نتائجنا توعد بنسبة خطأ (% 0.885) من خلال حسب متوسط مربع الخطأ ، بين قياسنا وحكم الأنسان.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACT	:	Accountability
API	:	Application programming interface
AWS	:	Amazon Web Service
CA	:	Control Area
CAIQ	:	Consensus Assessments Initiative Questionnaire
CC	:	Cloud Computing
CCM	:	Cloud Control Matrix
Cobit	:	Control Objectives for Information and Related Technology
CSA	:	Cloud Security Alliance
CSP	:	Cloud Service Provider
DG	:	Data Governance
DSS	:	Data Security Standard
FISMA	:	Federal Information Security Management Act
GQM	:	Goal Question Metric
HIPAA	:	Health Insurance Portability and Accountability Act
HITECH	:	Health Information Technology for Economic and Clinical Health
IaaS	:	Infrastructure as a service
IEC	:	International Electrotechnical Commission
IS	:	Information Security
ISO	:	International Organization for Standardization
ISP	:	Internet Service Provider
IT	:	Information Technology

ITIL	:	Information Technology Infrastructure Library
KAON	:	**KA**rlsruhe **ON**tology.
MSE	:	Mean Square Error
NIST	:	National Institute of Standards and Technology
NIST	:	National Institute of Standards and Technology
OS	:	Operation System
PaaS	:	Platform as a Service
PCI	:	Payment Card Industry
PCI	:	Payment Card Industry
QoS	:	Quality of service
SA	:	Security Architecture
SaaS	:	Software as a Service
SACS	:	Security Access Control Service
SANS	:	SysAdmin, Audit, Network, Security
SLA	:	Service Level Agreements
SOX	:	Sarbanes-Oxley
SSM	:	Semantic Similarity Measure
VM	:	Virtual Machine
WS4J	:	WordNet Similarity for Java

CHAPTER ONE

Introduction

1.1 Overview

In this chapter we review the thesis. A brief background about the scope of the thesis is given; Cloud Computing, Cloud Computing security, and semantic similarity measure. Then we give an idea about our research problem and how it has been addressed, our own contribution, and the outline of the thesis chapters.

1.2 Cloud Computing

Cloud Computing (CC) is a term used to describe both a platform and type of application. As a platform it supplies, configures and reconfigures servers, while the servers can be physical machines or virtual machines. On the other hand, CC describes applications that are extended to be accessible through the internet and for this purpose large data centres and powerful servers are used to host the web applications and web services (Boss et al., 2007). CC (Almorsy M., 2010) provides the next generation of internet based, highly scalable distributed computing systems in which computational resources are offered 'as a service'. The most widely used definition of the cloud computing model is introduced by NIST (Peter Mell, 2011) as “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.”.

Cloud Service Providers (CSPs) offer cloud platforms for their customers to use and create their web services, much like internet service providers (ISP) offer customers high speed broadband to access the internet. CSPs and ISPs both offer services. The cloud provides a layer of abstraction between the computing resources and the low level architecture involved. The customers do not own the actual physical infrastructure but merely pay a subscription fee and the CSP grants them access to the clouds resources and infrastructure. A key concept is that the customers can reduce expenditure on resources like software licenses, hardware and other services (e.g. email) as they can obtain all these things from one source, the CSP (Curran, K.,2011).

1.3 CC Security

CC is designed to be successful by reducing overhead and improving efficiency. With those improvements come the loss of control and possible security risk to the data (Townsend M., 2009) the leading U.S. market research firm Gartner released a report “Assessing the Security Risks of Cloud Computing” in June 2008. This report stated that cloud computing has great risk to data integrity, data recovery and privacy, etc. (Jing X, 2010).

There are still many open and interesting issues regarding CC paradigm and standards are still evolving. But, it is a general opinion that security is indeed one of the most important issues (Mell P.,2011). In the recent IDC report over 74.6% in 2008 (Figure 1.1) and 87.5% in 2009 (Figure 1.2) of users think that security is a dominant issue for widespread use of CC services.

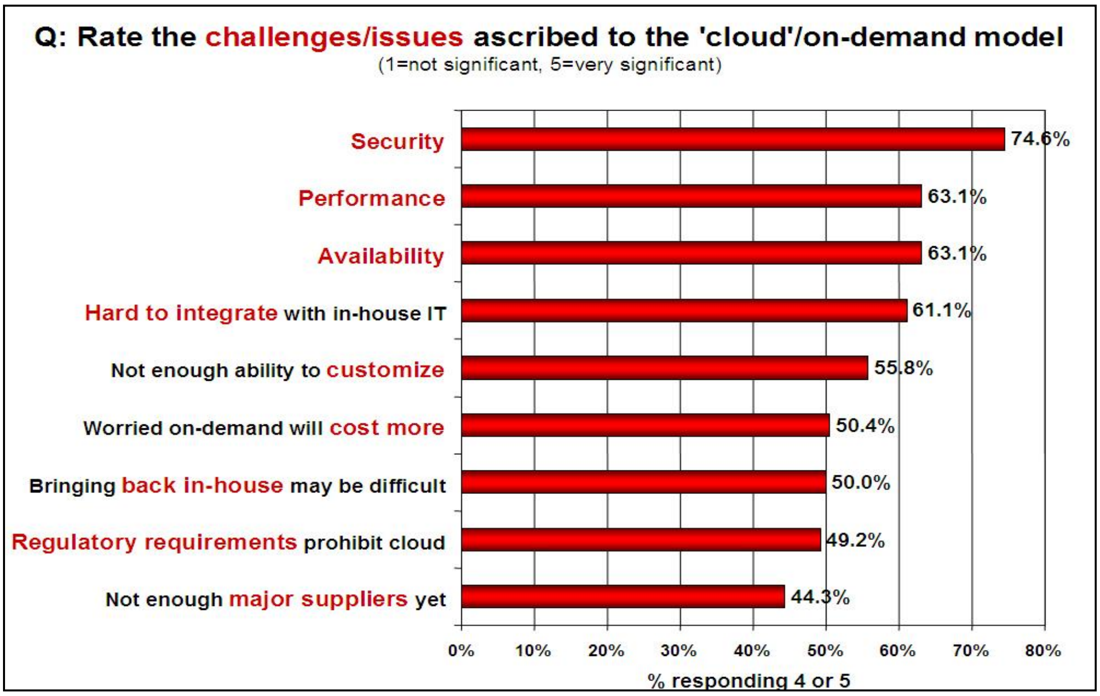


Figure 1.1: Importance of Security for CC Environments 2008

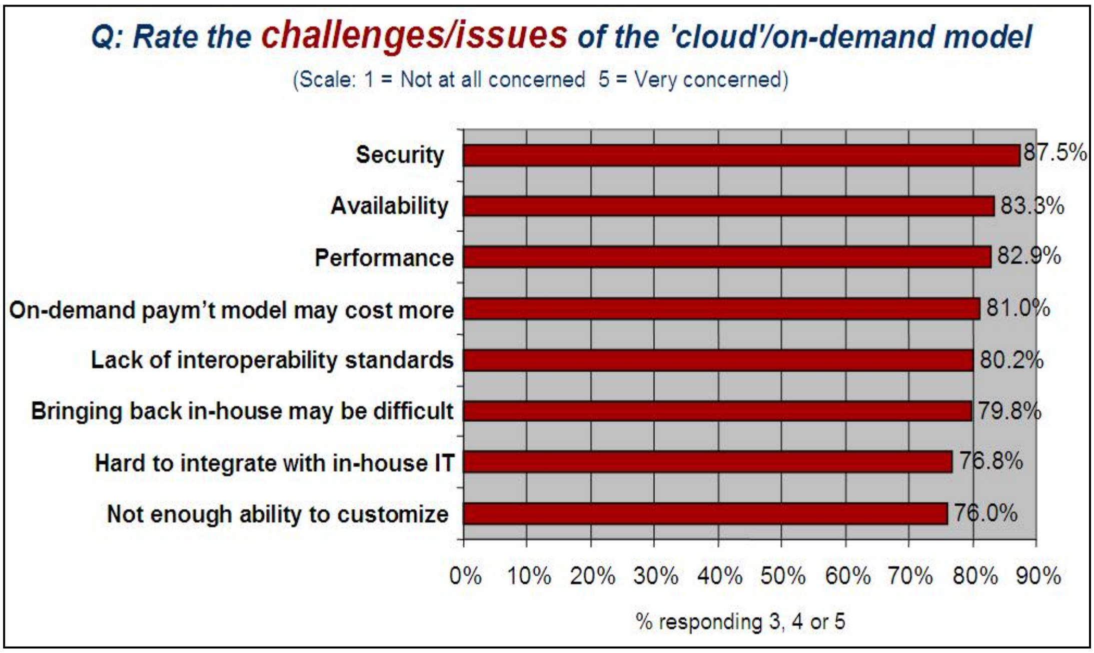


Figure 1.2: Importance of Security for CC Environments - 2009

Like traditional computing environments, CC brings risks and security concerns to the organizations that need to be considered appropriately. Such risks and security concerns include challenges in handling privileged user access, ensuring legal and regulatory compliance, ensuring data segregation, maintaining data recovery, difficulty in investigating illegal activities, and lack of assurance of long-term viability of the (CSP) (Kandukuri, B. R., 2009). CSP has recognized the cloud security concern and are working hard to address it. In fact, cloud security is becoming a key differentiator and competitive edge between cloud providers. By applying the strongest security techniques and practices, cloud security may soon be raised far above the level that IT departments achieve using their own hardware and software (Kumar V., 2012).

In CC environment, overall security issues can be evaluated from the points of CSP and the consumer. While the CSP focus on the continuity of their services against configuration updates for performance and QoS, spam and virus threats and proper customer accountability, clients mainly look for the security of their data and the reliability of the provider (Yildiz M.,2009).

Academics and security products manufacturers are actively studying CC data security (Shuanglin R., 2010). However there still exist many problems with cloud computing today. A recent survey shows that data security and privacy risks have become the primary concern for people to shift to cloud computing (Shuanglin R.,2012). Due to the above challenges and threats cloud customers therefore need to institute mechanisms to measure and improve the security of their information assets operating in the cloud. Among the alternatives available to the cloud customer for monitoring, measuring and hence improving information security of the assets managed in the cloud is to develop information security metric (Putri N. R.,2011).

1.4 Data Breaches Threat

A Data Breach is the intentional or unintentional release of secure information in an untrusted environment. Other terms for this phenomenon include unintentional information disclosure, data leak and also data spill (Wikipedia, 2013). In CC and according the Cloud Security Alliance (CSA) report “Top 10 threats in Cloud Computing” (Cloud Security Alliance, 2011); the data breaches threat is ranking No.5 in 2011 and No.1 in 2013 (Cloud Security Alliance, 2011); and it's in the high risk level Risk Matrix (Figure 1.3,1.4).

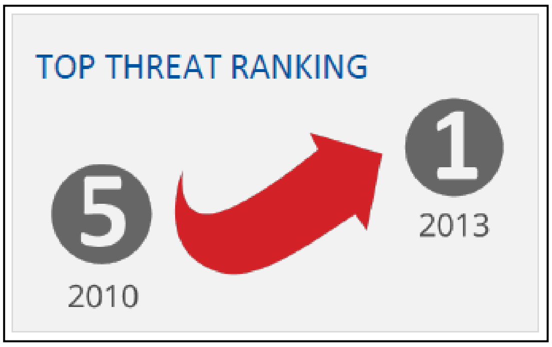


Figure 1.3: Data Breaches Threat Top Ranking 2013

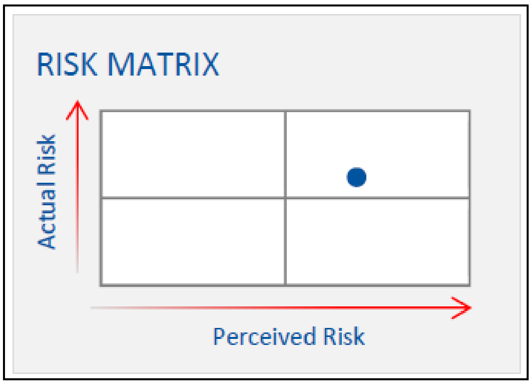


Figure 1.4: Data Breaches Threat in very high in Risk Matrix

1.5 Security Compliance

Security compliance distinguishes it from security itself. While security refers to a mechanism that have to be used in order for a system to be in a safe state from prospective threats, security compliance refers to a state of compliance with a given set of security requirements. Therefore, while security it is used to protect a system from threats, security compliance has nothing to do with this protection. Rather, security compliance ensures that the security measures taken to protect the system are compliant with the necessary requirements (Ullah K. W. ,2012). Klaus Julisch from IBM Research has defined the security compliance as follows (Julisch K., 2009) :

“Security compliance, in IT systems, is the state of conformance with externally imposed functional security requirements and of providing evidence (assurance) thereof."

These days security compliance generally indicates the compliance with industry accepted security standards such as NIST, ISO 270001/27002, HIPAA, PCI, etc. While compliance helps drive security, it does not equal actual security.

The 2011 Data Breach Investigation Report (Baker W., 2011) outlined the fact that non-compliance is one of the main reasons for data breaches in the Payment Card Industry (PCI). In this report, it was stated that 96% of the companies that suffered the breach have not achieved compliance with the PCI DSS. Only the remaining 4% of companies were still under attack despite having achieved the compliance with PCI DSS. This is a clear indication of how much difference can it make to have the security compliance.

1.6 Ontology and Semantic Similarity

Ontology; is an abstract description system for knowledge composition in a certain domain. By organizing concepts (terms) in a domain in a hierarchical way and describing relationships between terms using a small number of relational descriptors, an ontology supplies a standardized vocabulary for representing entities in the domain. (Jiang, R. 2013).

Semantic similarity is a concept whereby a set of documents or terms within term lists assign a metric based on the likeness of their meaning / semantic content. Various semantic similarity techniques are available which can be used for measuring the semantic similarity between text documents (Nagwani N.,2011).

This thesis explores the possibility to measure security compliance for data breaches threat based semantic similarity measure between the documents of standard compliments and CSP response against data breaches threat.

1.7 Problem Definition

The CC offers dynamically scalable resources provisioned as a service over the Internet. Each CSPs has own security requirements. We need to unify and measure the majority of cloud security compliance and requirements in order to the evaluate the level of the security of his cloud. This will lead to several problems to be identified as follows:

1- How can we classify the different parameters of CC threats?

2- How can we build a unified classification measure of the data breaches threat?

3- How to automate the measure (semantically) for the each CSP compliances to mitigate data breaches threats?

4- How to deploy semantic similarity measure to rank the CSP according CC threat.

1.3 Questions

This research represents a new approach in combining between CC security and ontology concept extraction in unified model that CSPs can benefit from this technology. Thus, this research will answer the following questions:

1. How can we identify the aspect of the cloud threat and challenges?
2. How can we extract ontology concept of each (Control Area) related to the threat?
3. How can we extract ontology concept of each CSP?
4. How can we measure the similarity concept between the above extract ontology concepts?

1.4 Contribution

This thesis contributes the following:

1- Collections and classify data breaches of CSP security response against data breaches threat by using ontology concept.

2- Explains how to measure CSP security plans for different CC threats.

3- Support CSPs toward asset their security issues.

1.4 Motivation

CC is an evolving paradigm with lots of benefits. It can be seen as an integration of traditional computing technologies and network technologies. The CC security has become a hot topic in the industry and academic research. In particular data security is concerned with organizations and users which use cloud computing.

Thus motivates this to propose a new measure depend on matching of ontology concepts semantically in each CA of data breaches threat with CSPs.

1.5 Objectives of the Thesis

The main objective of this research is:

1. Unify security compliance against data breaches threat of CSP.
2. Provide cloud developers and users with a model to evaluate the security respond of different type of cloud.
3. Spread awareness of the security requirements for the CSP.
4. Collaborate with the researcher in the field to develop this research and the tool in the future.

1.6 Methodology

The methodology that will be used to develop our model contains the following phases:

- Study and Analysis Phase.
- Design and Implementation Phase
- Test Phase

Study and Analysis Phase

The first step in the studying and analysis will be the collection of data, acquiring information and knowledge for the following:

1- Different security issues in CC and focusing on the data security in CC and protect it from threats.

2- CSP’s security response or their security actions against the security compliance.

3- Studying and understanding different traditional and CC security measure method.

Design and Implementation Phase

In this phase we are going to design a model and implantation it with the necessary collection data and information in order to create ontology concepts of data security for each parameter (CA) of it. Also create ontology concepts for the CSP security respond.

Test Phase

This phase consists testing our model, to effectiveness measure of CSP security responsiveness.

1.7 Organization of the Thesis

Chapter 2: Presents a theoretical background about the CC ,its security, semantic similarity matching measure and focusing on data security aspects, what is security compliance for CSP, what is the security measure method in cloud computing.

Chapter 3: Presents the process of applying the effectiveness measure the security of CC. Also describe in detail of our approach to measuring the CSP security actions.

Chapter 4: Presents in detail the component of our model and matching process in details also presents the experimental our proposed model.

Chapter 5: Conclusion and future works.

CHAPTER TWO

Literature Review & Related Works

2.1 Introduction

This chapter gives a brief idea about the most relevant work in the literature that relates to our study. We provide a background and literature review of the four main concepts covered by this research, namely, Cloud Computing, General Security Issued in CC, Data Security Issued in CC, Ontology and Semantic Similarity Measure. The most important related studies in the field of measurement CC security in Section 2.3. Finally, in Section 2.3, we present the tools that are used for our research.

2.2. Literature Review

This part investigation of existing study and research which is relevant to our theme and present some background reading required to give context to our research.

2.2.1 Cloud Computing (CC)

(Curran K., 2011) provided an overview of the key aspects of CC which has five key attributes which grant it some advantages over similar technologies and these attributes include: •Multitenancy (shared resources): Unlike previous computing models, which assumed dedicated resources dedicated to a single user or owner, cloud computing is based on a business model in which resources are shared at the network, host and application level. •Massive scalability: Cloud computing provides the ability to scale to tens of thousands of systems, as well as the ability to massively scale bandwidth and storage space. •Elasticity: Users can rapidly increase and decrease their computing resources as needed, as well as release resources for other uses when they are no longer required. • Pay as you go: Users pay for only the resources they actually use and for only the time they require them. • Self-provisioning of resources: Users self-provision resources, such as additional systems (processing capability, software & storage) and network resources.

(Zissis D, 2012) explained the available service model in CC: Infrastructure as a Service (IaaS) provides the consumer with the capability to provision processing, storage, networks, and other fundamental computing resources, and allow the consumer to deploy and run arbitrary software, Platform as a Service (PaaS) provides the consumer with the capability to deploy onto the cloud infrastructure consumer created or acquired applications, produced using programming languages and tools supported by the provider and Software as a Service (SaaS) provides the consumer with the capability to use the provider’s applications running on a cloud infrastructure. These services are delivered and consumed in real-time over the Internet. Also this research presents the CC four deployment models: Private Cloud, Public cloud, Hybrid cloud and Community cloud.

2.2.2 General Security Issue in CC

(Subashin S.,2011) surveyed of the different security risks that posed a threat to the cloud. Also surveyed more specifically to the different security issues that has emanated due to the nature of the service delivery models of a CC system and define the elements key for each. Security issues in SaaS **:** the following key security elements should be carefully considered as an integral part of the SaaS application development and deployment process: Data security, Network security, Data locality, Data integrity, Data segregation, Data access, Authentication and authorization, Data confidentiality, Web application security, Data breaches, Virtualization vulnerability, Availability, Backup, Identity management and sign-on process. Security issues in PaaS: in the PaaS model, the provider might give some control to the people to build applications on top of the platform. Hackers are likely to attack visible code, including but not limited to code running in user context. They are likely to attack the infrastructure and perform extensive black box testing. Security issues in IaaS : the security responsibilities of both the provider and the consumer greatly differ between cloud service models. Some CSP offering, includes vendor responsibility for security up to the hypervisor, meaning they can only address security controls such as physical security, environmental security, and virtualization security. The consumer, in turn, is responsible for the security controls that relate to the IT system including the OS. This survey also presented the current security solutions from several groups and organization are interested in developing security solutions and standards for the cloud. CSA is gathering solution providers, non- profits and individuals to enter into a discussion about the current and future best practices for information assurance in the cloud.

(Jensen M., 2009) focused on technical security issues arising from the usage of Cloud services and especially by the underlying technologies used to build these cross domain Internet-connected collaborations. The authors presented a selection of issues of CC Security. They investigated ongoing issues with application of XML Signature and the Web Services security frameworks (attacking the Cloud Computing system itself), discussed the importance and capabilities of browser security in the Cloud Computing context (SaaS), raised concerns about Cloud service integrity and binding issues (PaaS), and sketched the threat of flooding attacks on Cloud systems (IaaS).

(Che J.,2011) the authors proposed some security strategies against common security issues of cloud computing. These security strategies as follows: (1) Securely Construction Strategies of Cloud Computing (a) Traditional Security Practice Mechanism (b) Virtualization Security Risks Assessment (c) Development Outsourcing Risk Control (d) Portability and Interoperability (2) Securely Operation Strategies of Cloud Computing (a) Business Continuity Assurance (b) Attack Proactive Alerting (c) Data Leak Prevention (d) Security Accident Notification & Response (e) Security Incidents Audit.

(Parek M. D. H., , 2013) this study is presented and so as to effectively refine the crude security issues under various areas of cloud. This study also aimed at revealing different security threats under the cloud models as well as network concerns to stagnate the threats within the cloud, facilitating researchers, cloud providers and end users for noteworthy analysis of threats. Also represented the schematic diagram showing the hierarchy of the cloud computing, with security challenges as showing in Figure 2.1:

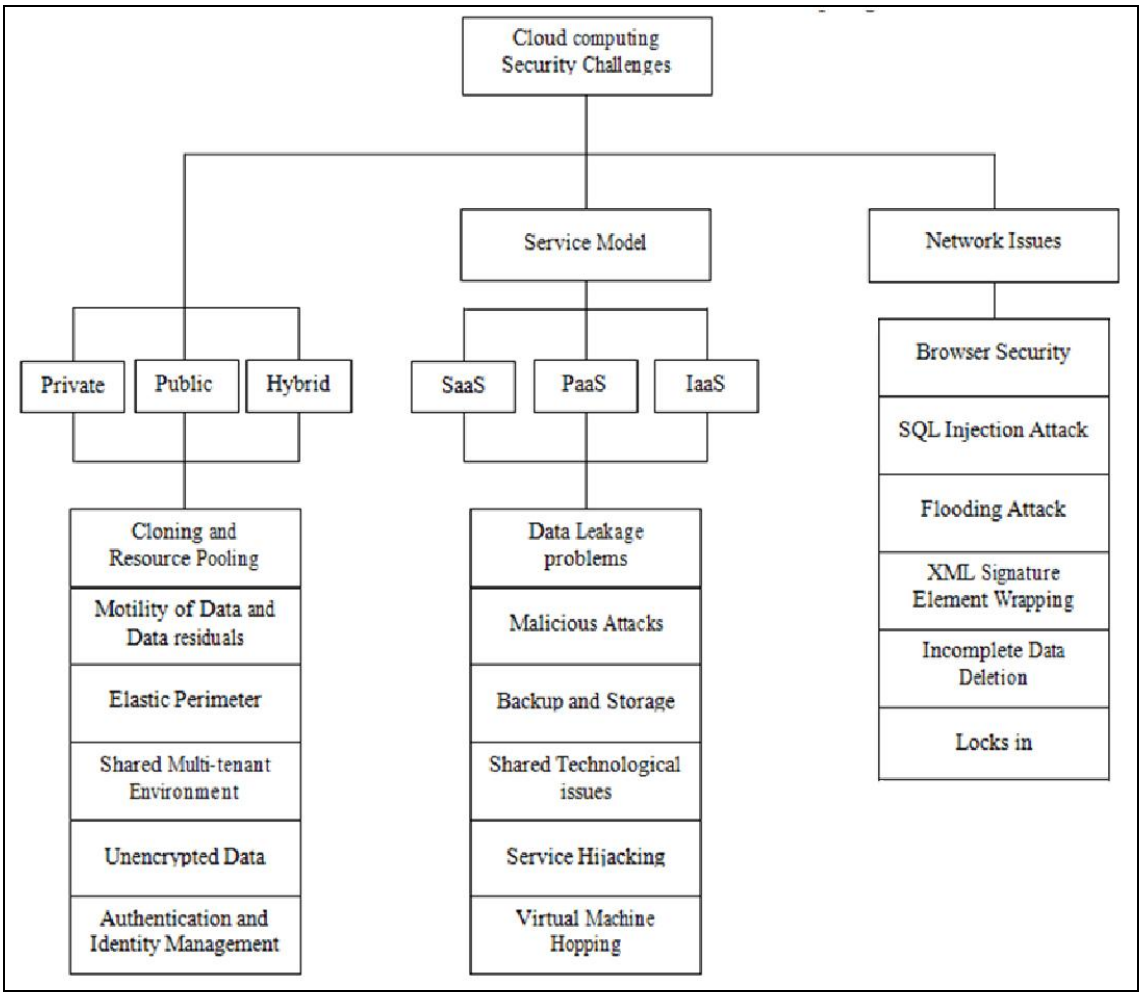


Figure 2.1: The hierarchy of the CC, with security challenges

(Pal D. G., 2012) proposed a framework consists of eight domains which can be further divided into sub domains. All these domains should comply with various regulations and government policies like SOX, FISMA, HIPAA, COBIT, ISO/IEC 27001/2, etc. accordingly. (Figure 2.1) shows their proposed framework for cloud security.

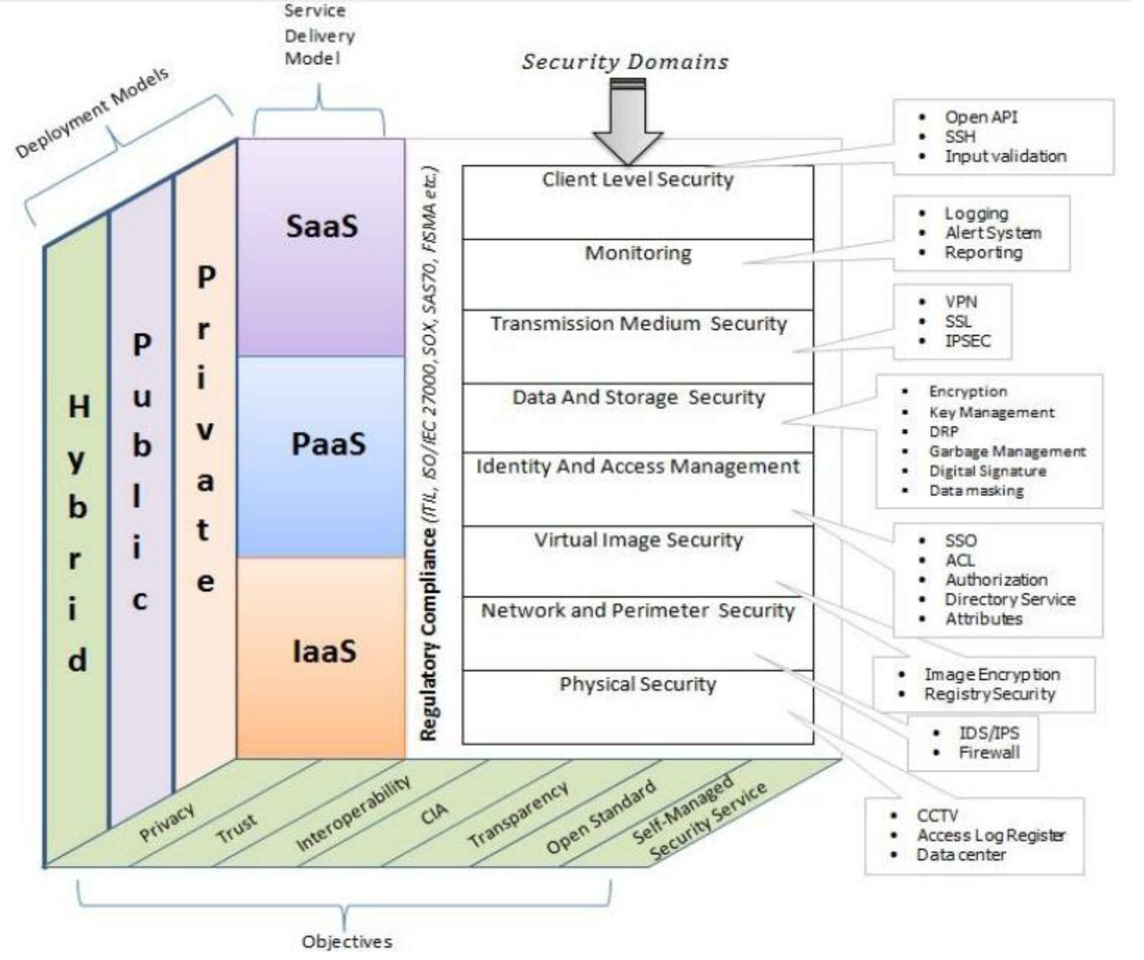


Figure 2.2: Proposed Cloud Security Framework for end-to-end security

2.2.4 Data Security Issue in CC

This section presents the necessary theoretical background for understanding the data security aspects and challenges in CC:

(Malik A., 2012) this paper defined a methodology for cloud providers that will protect users' data, information which is of high importance. When data deleted without any backup or encoding key loss/unauthorized access, data is always in danger of being lost or stolen. To provide a solution for this, we need to: • Implement fault free API access control. •Mechanism used for encryption and protection of data should be secure. •Data protection analysis done at both design and run time. •Provider backup and preservation strategies must be defined.

(Shuanglin R., 2012), put forward management ideas of user data classification and designed a cloud-based data security policy through user demand for data security protection. The whole strategy is divided into technical support and management to protect the entire strategy more effective. (1) Technical support section (a) Strong Authentication (b) Classification of data evaluation (c) Filtering sensitive information (d) CC security gateway (2) Management to protect (a) To establish a safety management system (b) To establish rules and regulations (c) Safety education.

(Yuefa D., 2009) built a data security model for cloud computing. The model used three level defence system structure, in which each floor performs its own duty to ensure that the data security of cloud layers. The first layer: responsible for user authentication, the user of digital certificates issued by the appropriate, manage user permissions. The second layer: responsible for user's data encryption, and protect the privacy of users through a certain way. The third layer: The user data for fast recovery, system protection is the last layer of user data.

(Mohamed E. M., 2012) proposed a new data security model (in the previous research) based on studying of cloud computing architecture. The model used three-layer system structure. The first layer: responsible for user authentication, almost this is two factor authentications. The second layer: responsible for user's data encryption, and protect the privacy of users through a certain way by using one symmetric encryption algorithms. Also allow protection from user. The third layer: The user data for fast recovery this depends on the speed of decryption. They implemented software to enhance work in a data security model for cloud computing. This software provided to the cloud provider and implemented with two factor authentication. This software compares between eight modem encryption algorithms. This comparison based on Statistical Tests to get the most security algorithms. This software gets the faster and the highest security algorithm based on cloud infrastructure. They proposed to CSP the suitable, more security encryption algorithm to its platform. Finally, by this evaluation the authors ensured that data retrieve faster to the user and ensure security of user data. In addition, they make software to the cloud user. This software allows user to choose between eight encryption algorithms to ensure data security. This software gives the cloud user some advices to select the most security or most security and faster algorithm that suitable to its cloud infrastructure.

(Chen Z., (2010) discussed the evolvement of CC paradigm and present a framework for secure CC through IT auditing. The research approach is to establish a general framework using checklists by following data flow and its lifecycle. The checklists are made based on the cloud deployment models and cloud service models. The contribution of the paper is to understand the implication of cloud computing and what is meant to secure cloud computing via IT auditing rather than propose a new methodology and new technology to secure cloud computing. Their checklist focuses on the following: Data location Aware, Data ownership aware, Data protection plan and best practice, Data processing isolation, Data Lock-in, IaaS IT architecture, Regulatory Compliance.

(Dai Yuefa W. B., 2009) the authors built a data security model for cloud computing. The model used three level defense system structure, in which each floor performs its own duty to ensure that the data security of cloud layers. The first layer: responsible for user authentication, the user of digital certificates issued by the appropriate, manage user permissions. The second layer: responsible for user's data encryption, and protect the privacy of users through a certain way. The third layer: The user data for fast recovery, system protection is the last layer of user data.

(El-Khameesy N., 2012) highlighted the security aspects of data storage from perspectives of threats and attacks from one side and approaches for solutions from the other side. The paper proposed an effective and flexible security policy and procedure with explicit data support, including block update, delete, and append. Control Access Data Storage that includes the necessary policies, processes and control activities for the delivery of each of the Data service offerings. The collective control Data Storage encompasses the users, processes, and technology necessary to maintain an environment that supports the effectiveness of specific controls and the control frameworks. The Security, correctness and availability of the data files being stored on the distributed cloud servers must be guaranteed by the following: • Providing Security Policy & Procedure for Data Storage. • Defense in Depth for Data Storage in cloud computing. • Layer 1 – Devices on the Storage Network. • Layer 2 – Network connectivity.

2.2.5 Ontology and Semantic Similarity Measure

(Jiang, R. 2013) defined the Ontology; is an abstract description system for knowledge composition in a certain domain. By organizing concepts (terms) in a domain in a hierarchical way and describing relationships between terms using a small number of relational descriptors, an ontology supplies a standardized vocabulary for representing entities in the domain. They reviewed a majority of existing methods that rely on ontologies to calculate semantic similarity between terms. Authors classified existing methods into five categories: methods based on semantic distance, methods based on information content, methods based on properties of terms, methods based on ontology hierarchy, and hybrid methods.

(Talib A. M, 2012) presented the semantically structure of cloud computing security knowledge, an ontology based security approaches have been increasingly adopted by several expertise from diverse security domains. CC issues represent the main class concept (Figure 2.3) containing several sub concepts such as Legal Issues, Flexibility/Elasticity, Compliance, Open Standard, Security, Freedom, Reliability and Privacy. Slots (not shown in the figure) are attributes associated with a class.

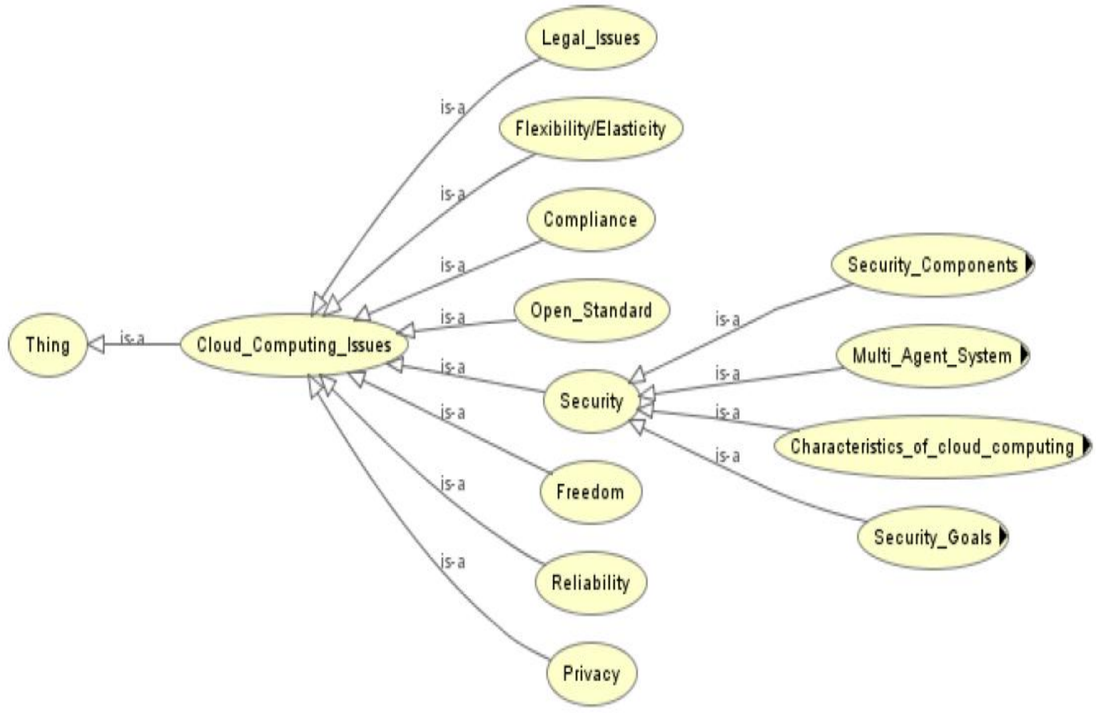


Figure 2.3 Simple conceptual definition of cloud computing issues

Also explained a generalization and specification ontology for the (Cloud Data Storage) security can be met. They have described an OWL-based ontology with its core concepts cloud asset, cloud threat, security goal, cloud users and CSPs. All the core concepts are subclasses or instantiated to provide the domain vocabulary of information security. Ontology can be developed based on three main steps, 1) domain, purpose and scope setting, 2) important terms acquisition, classes and class hierarchy conceptualization and 3) instances creation. Their approach used to enhance the security goals, CSPs and cloud users collaborations, and ontology to maintain consistency within a heterogeneous cloud computing environment.

(Michelizzi J., 2005) described the different types of the semantic similarity: 1) Path Length Similarity Measures: a- Path b- Wu & Palmer c- Leacock & Chodorow 2) Leacock & Chodorow 2) Information Content Similarity Measures: a- Resnik b- Lin c- Jiang & Conrath 3) Semantic Relatedness Measures a- Extended Gloss Overlaps (Adapted Lesk) b- Context Vectors c- Hirst & St-Onge. Also define the WordNet as a machine readable dictionary created at the Cognitive Science Laboratory at Princeton University. Unlike most dictionaries, WordNet contains only open-class words (nouns, verbs, adjectives, and adverbs). WordNet does not contain closed-class words such as pronouns, conjunctions, and prepositions.

(Budanitsky, 2006) proposed to evaluate similarity measurements based on WordNet. However, the authors have evaluated five measurements lexical semantics distance. The authors mentioned that most of their work was limited to the narrower notion of similarity measures. These relationships include not just hyponymy and the nonhyponymy relationships in WordNet such as meronymy but also *associative* and *ad* *hoc* relationships. As the authors mentioned, these can include just about any kind of functional relation or frequent association in the world.

2.2 Related Work

(Fenz S., 2010) proposed a methodology for automatically generating ISO 27001-based IT-security metrics and showed how the security ontology can be used to generate concrete and organization-specific knowledge regarding existing control implementations. In the example of ISO 27001, author showed that the developed methodology supports organizations in evaluating (1) their compliance to information security standards, and (2) the effectiveness of existing control implementations.
The authors took some concept from this research to create ontology for the controls of the data breaches threat on the CC base on major compliance international standards.

(Tariq M. I., 2012) discussed security issues of cloud computing, and proposed basic building blocks of information security metrics framework for cloud computing. The information security metrics framework has four major stages: 1) Metrics Preparation the IS metrics preparation phase involves information security metrics development team to develop useful information security metrics. 2) Threat Identification and Analysis The 2nd Phase of this proposed framework is about threat elicitation and analysis. In this phase, the threats are identified from information security metrics and different techniques like threat tree are applied to analyze the threat 3) Threat Processing After Analysis of threat, this phase is defined different activities that help cloud users to process on identified IS threats. This phase is very critical & technical and required due concentration of the threat solving team.4) Application The last act of this framework focuses on the use of the security metrics and threat severity levels by the decision makers. They evaluate the security and take suitable actions. This research determined the (Drive information security metrics) and (Security requirements) in phase 2 from standard and guide for guidance in metric development. The author was using the majority of this guidance as data security requirements in this project. The author has mentioned a set of international accepted frameworks; standards and guides are available for guidance in metrics development. IT Infrastructure Library (ITIL) and Control Objectives for Information and related Technology (CobiT) are renowned frameworks. International Organization for Standardization (ISO) has ISO/IEC 27002 information security and control standards which also can be used to drive information security metrics. At present, SANS has also published an information security metrics guide which is very helpful for cloud users to drive information security metrics.

(Bhensook N., 2012) presented an initial attempt to assess security requirements compliance of CSP by applying the Goal Question Metric (GQM) approach to quality measurement and defining a weighted scoring model for the assessment. The security goals and questions that address the goals are taken from CSA, Cloud Control Matrix (CCM) and CAIQ (Figure 2.1), then transform such questions into more detailed ones and define metrics that help provide quantitative answers to the transformed questions based on evidence of security compliance provided by the cloud providers. The scoring is weighted by the quality of evidence, i.e. its compliance with the associated questions and its completeness. This research proposed scoring system architecture.

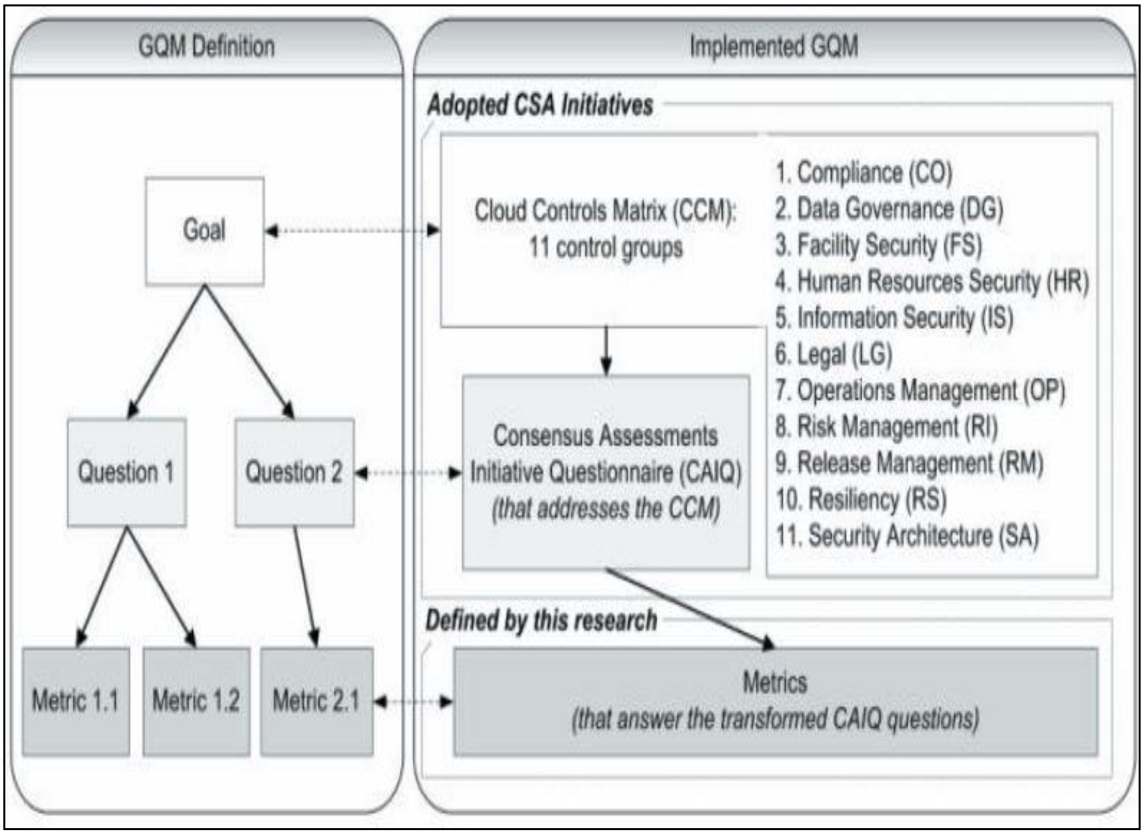

GQM Definition
Implemented GQM
Adopted CSA Initiatives
Goal
Cloud Controls Matrix (CCM): 11 control groups
1. Compliance (CO)
2. Data Governance (DG)
3. Facility Security (FS)
4. Human Resources Security (HR)
5. Information Security (IS)
6. Legal (LG)
7. Operations Management (OP)
8. Risk Management (RI)
9. Release Management (RM)
10. Resiliency (RS)
11. Security Architecture (SA)
Question 1
Question 2
Consensus Assessments Initiative Questionnaire (CAIQ) (that addresses the CCM)
Defined by this research
Metric 1.1
Metric 1.2
Metric 2.1
Metrics (that answer the transformed CAIQ questions)

Figure 2.4: Relation between GQM, CCM, and CAIQ

(Luna J., 2011) this paper presented the view on the importance and challenges of developing a security metrics framework for the Cloud, also taking into account ongoing research with organizations like the Cloud Security Alliance (CSA) and European projects like ABC4Trust, CoMiFin and INSPIRE. The authors also introduced the basic building blocks of a proposed security metrics framework for elements such as a CSP security assessment, taking into account the different service and deployment models of the Cloud.

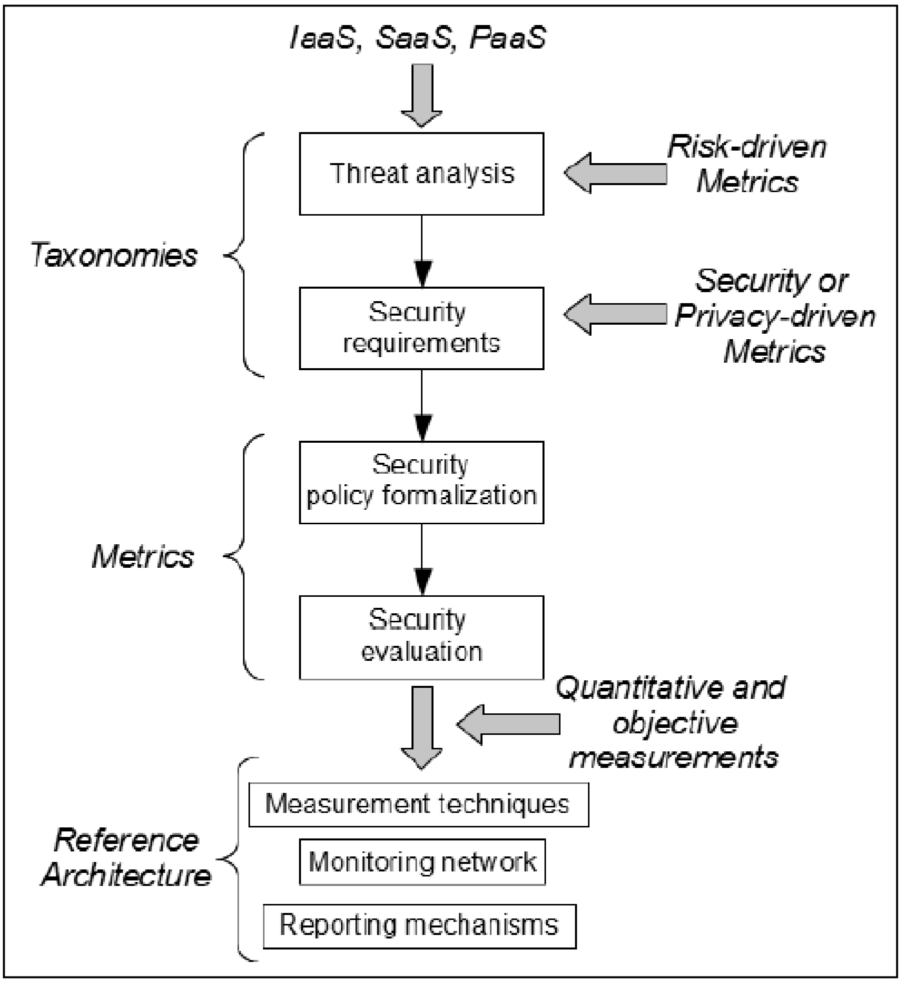


Figure 2.5: Basic building blocks of the proposed security metrics framework for the Cloud

2.3 Software Used in the Research

We used many tools in order to reach some necessary results, below is a brief description of those tools used in this research:

2.5.2 KAON TextToOntoTool

TextToOnto (Maedche A., 2001) is a tool suite built upon KAON in order to support the ontology engineering process by text mining techniques; providing a collection of independent tools for both automatic and semi-automatic ontology extraction (Figure2.2).

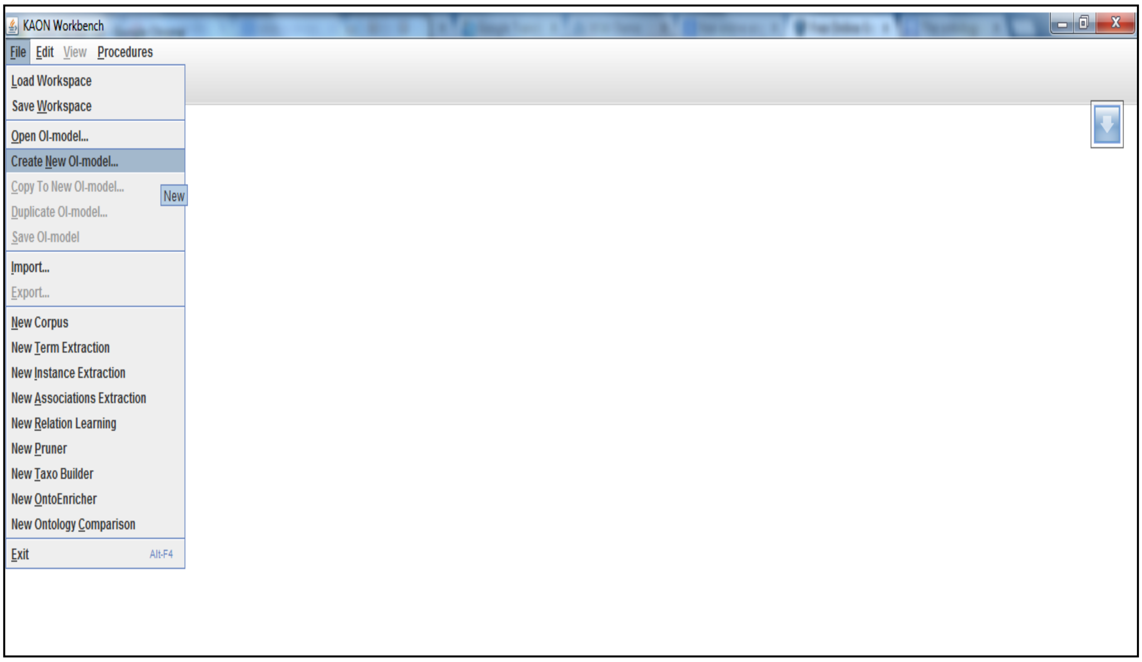
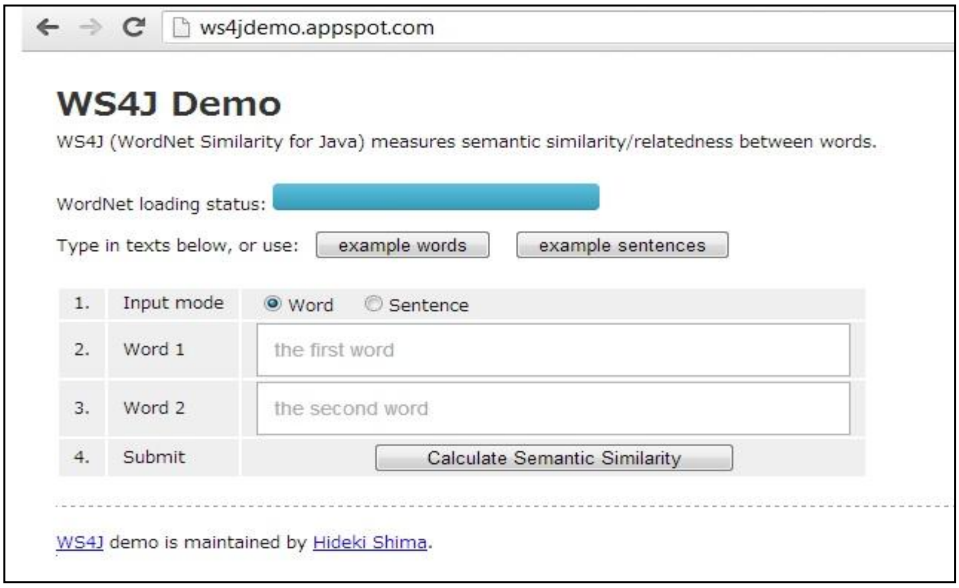


Figure 2.6: The front-end of the KAON TextToOnto Tool

2.5.3 WordNet Similarity for Java (WS4J)

WS4J provides (Figure 2.3) a pure Java API for several published Semantic Relatedness/Similarity algorithms[1].



WS4J Demo

WS4J (WordNet Similarity for Java) measures semantic similarity/relatedness between words.

WordNet loading status:

Type in texts below, or use:

1.	Input mode	● Word ○ Sentence
2.	Word 1	the first word
3.	Word 2	the second word
4.	Submit	Calculate Semantic Similarity

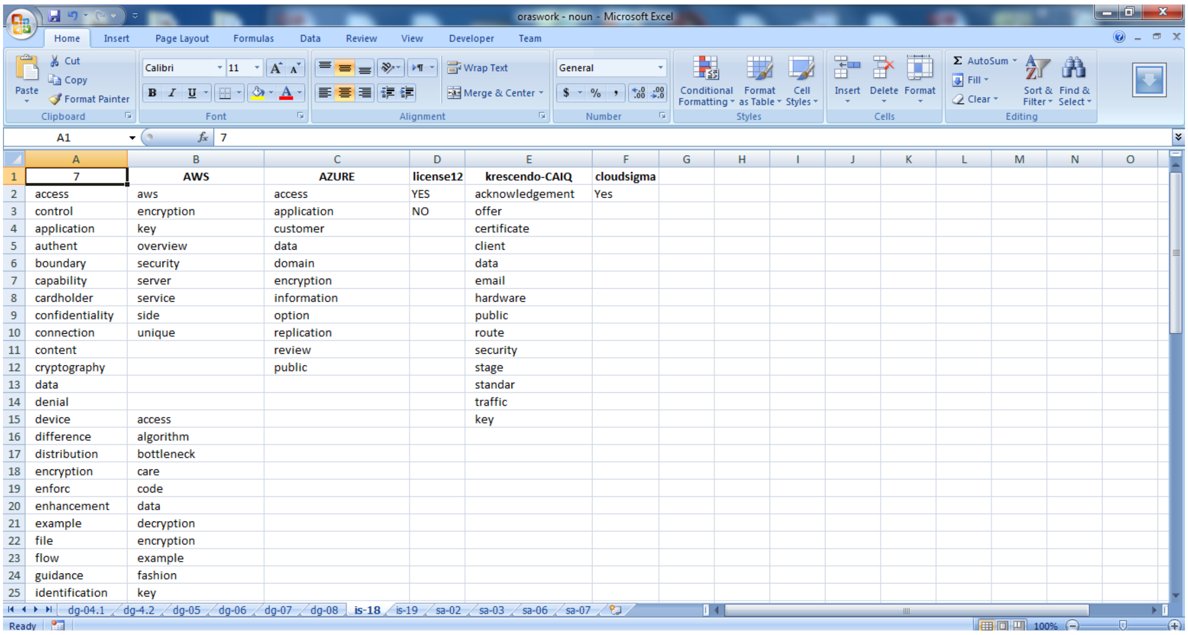
WS4J demo is maintained by Hideki Shima.

Figure 2.7: The front-end of the WS4J Tool

[1] http://ws4jdemo.appspot.com/

2.5.4 Microsoft Excel

Microsoft Excel (Figure 2.4) is a software program produced by Microsoft Corporation that allows users to organize, format and calculate data with formulas using a spreadsheet system.



A1 | 7

	A	B	C	D	E	F
1	7	AWS	AZURE	license12	krescendo-CAIQ	cloudsigma
2	access	aws	access	YES	acknowledgement	Yes
3	control	encryption	application	NO	offer	
4	application	key	customer		certificate	
5	authent	overview	data		client	
6	boundary	security	domain		data	
7	capability	server	encryption		email	
8	cardholder	service	information		hardware	
9	confidentiality	side	option		public	
10	connection	unique	replication		route	
11	content		review		security	
12	cryptography		public		stage	
13	data				standar	
14	denial				traffic	
15	device	access			key	
16	difference	algorithm				
17	distribution	bottleneck				
18	encryption	care				
19	enforc	code				
20	enhancement	data				
21	example	decryption				
22	file	encryption				
23	flow	example				
24	guidance	fashion				
25	identification	key				

Figure 2.8: The front-end of the Microsoft Excel

CHAPTER 3

Data Extraction Ontology Concepts

3.1 Introduction

In this chapter we will show in detail the steps of our approach to design a model of measuring the CC security semantically. In this study we will discuss the aspects related to the data breaches threat in CC to gathering the necessary data that covered its parameters to be measured. Also will discuss the preparation of documents to extracting the ontology concepts of the data breaches threat in order to match them with CSP response against that threat semantics. The research mentioned above consists of two stages as shown in Figure 3.1:

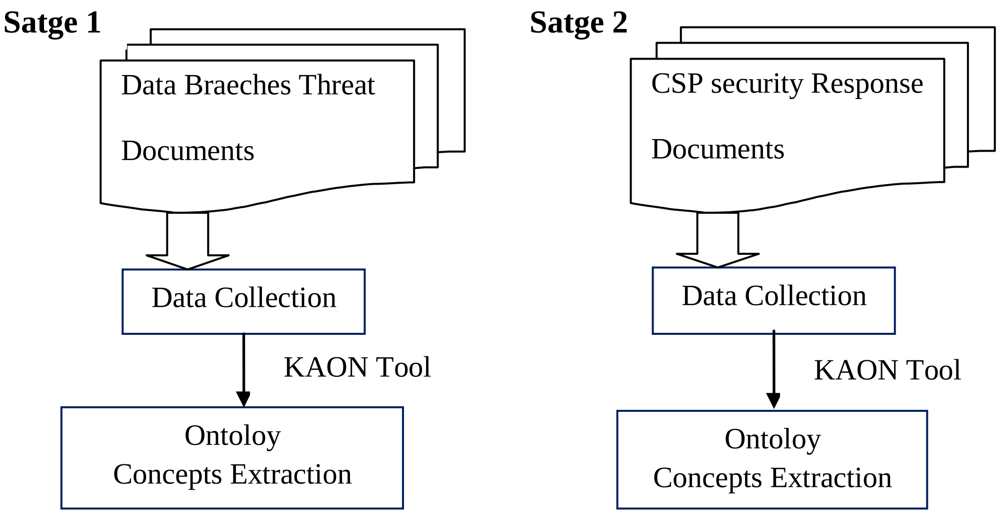

Satge 1
Data Braeches Threat Documents
Data Collection
KAON Tool
Ontoloy Concepts Extraction
Satge 2
CSP security Response Documents
Data Collection
KAON Tool
Ontoloy Concepts Extraction

Figure 3.1: Stages of preparing CC security measurement

Our approach aims to measure the CSP security response, by matching semantically his response with international standard compliance.

We have begun with The CSA CCM (Cloud Security Alliance, 2013); it’s specifically designed to provide fundamental security principles to CSP and to assist prospective cloud customers in assessing the overall security risk of a CSP. The CSA CCM provides a controls framework that gives a detailed understanding of security concepts and principles that are aligned with the CSA guidance in 13 domains. The foundations of the CSA CCM rest on its customized relationship with other industry-accepted security standards, regulations, and control frameworks such as the ISO 27001/27002, COBIT, PCI DSS… etc). The CSA documents are the foundation of our data collection.

3.2 Extracting Data Breaches Threat Ontology Concepts

This step to studying the aspects related to data breaches threat from the standpoint of major international standard (Ullah K. W., 2012) . Then we collected the documents that describe the target domain.

3.2.1 Classification of Data Breaches Threat

It is not easy to identify the classification of the security threats in CC. We searched and reading many documents (academic research, white papers, technical report... etc.) related to data breaches threat. Our start point was the latest CSA report called “A Security Guidance for Critical Areas of Focus in Cloud Computing V3.0”; this report shows us the outline of the beginning to identify the classification of the data breaches threat in CC. The data breaches in CC associated with the (11) parameters called Controls Area (CA) (Cloud Security Alliance, 2013) as shown in Table 3.1. The full specifications of each CA for data breach threat:

Table 3.1: The specification of the classification for data breaches threat

No.	Control Area (CA)	Control ID	Control Specification
1	Data Governance - **Retention Policy**	DG-04	Policies and procedures for data retention and storage shall be established and backup or redundancy mechanisms implemented to ensure compliance with regulatory, statutory, contractual or business requirements. Testing the recovery of backups must be implemented at planned intervals.
2	Data Governance - **Secure Disposal**	DG-05	Policies and procedures shall be established and mechanisms implemented for the secure disposal and complete removal of data from all storage media, ensuring data is not recoverable by any computer forensic means.
3	Data Governance - **Non-Production Data**	DG-06	Production data shall not be replicated or used in non-production environments.
4	Data Governance - **Information Leakage**	DG-07	Security mechanisms shall be implemented to prevent data leakage.
5	Data Governance - **Risk Assessments**	DG-08	Risk assessments associated with data governance requirements shall be conducted at planned intervals considering the following: • Awareness of where sensitive data is stored and transmitted across applications, databases, servers and network infrastructure • Compliance with defined retention periods and end-of-life disposal requirements • Data classification and protection from unauthorized use, access, loss, destruction, and falsification
6	Information Security - **Encryption**	IS-18	Policies and procedures shall be established and mechanisms implemented for encrypting sensitive data in storage (e.g., file servers, databases, and end-user workstations) and data in transmission (e.g., system interfaces, over public networks, and electronic messaging).
7	Information Security - **Encryption Key Management**	IS-19	Policies and procedures shall be established and mechanisms implemented for effective key management to support encryption of data in storage and in transmission.

Table 3.1: The specification of the classification for data breaches threat

No.	Control Area (CA)	Control ID	Control Specification
8	Security Architecture - **User ID Credentials**	SA-02	Implement and enforce (through automation) user credential and password controls for applications, databases and server and network infrastructure, requiring the following minimum standards: • User identity verification prior to password resets. • If password reset initiated by personnel other than user (i.e., administrator), password must be immediately changed by user upon first use. • Timely access revocation for terminated users. • Remove/disable inactive user accounts at least every 90 days. • Unique user IDs and disallow group, shared, or generic accounts and passwords. • Password expiration at least every 90 days. • Minimum password length of at least seven (7) characters. • Strong passwords containing both numeric and alphabetic characters. • Allow password re-use after the last four (4) passwords used. • User ID lockout after not more than six (6) attempts. • User ID lockout duration to a minimum of 30 minutes or until administrator enables the user ID. • Re-enter password to reactivate terminal after session idle time for more than 15 minutes. • Maintain user activity logs for privileged access or access to sensitive data.
9	Security Architecture - **Data Security / Integrity**	SA-03	Policies and procedures shall be established and mechanisms implemented to ensure security (e.g., encryption, access controls, and leakage prevention) and integrity of data exchanged between one or more system interfaces, jurisdictions, or with a third party shared services provider to prevent improper disclosure, alteration or destruction complying with legislative, regulatory, and contractual requirements.
10	Security Architecture - **Production / Non-Production Environments**	SA-06	Production and non-production environments shall be separated to prevent unauthorized access or changes to information assets.
11	Security Architecture - **Remote User Multi-Factor Authentication**	SA-07	Multi-factor authentication is required for all remote user access.

3.2.2 Data Collection

In this section we will talk about data collection of the compliance items issued from major international standard organizations. These items are associated with the classification of the data breaches threat (Cloud Security Alliance, 2013) as shown in Table 3.2.

Table 3.2: **Compliance Map for Classification of Data Breaches Threat**

Control Area	Control ID	Scope Applicability from International Standard Organizations				
		COBIT 4.1	**HIPAA / HITECH Act**	**ISO/IEC 27001-2005**	**NIST SP800-53 R3**	**PCI DSS v2.0**
Data Governance - Retention Policy	DG-04	DS 4.1 DS 4.2 DS 4.5 DS 4.9 DS 11.6	45 CFR 164.308 (a)(7)(ii)(A) 45 CFR 164.310 (d)(2)(iv) 45 CFR 164.308(a)(7)(ii)(D) 45 CFR 164.316(b)(2)(i)	Clause 4.3.3 A.10.5.1 A.10.7.3	CP-2 CP-6 CP-7 CP-8 CP-9 SI-12 AU-11	3.1 3.1.1 3.2 9.9.1 9.5 9.6 10.7
Data Governance - Secure Disposal	DG-05	DS 11.4	45 CFR 164.310 (d)(2)(i) 45 CFR 164.310 (d)(2)(ii)	A.9.2.6 A.10.7.2	MP-6 PE-1	3.1.1 9.10 9.10.1 9.10.2 3.1
Data Governance - Non-Production Data	DG-06		45 CFR 164.308(a)(4)(ii)(B)	A.7.1.3 A.10.1.4 A.12.4.2 A.12.5.1	SA-11 CM-04	6.4.3
Data Governance - Information Leakage	DG-07	DS 11.6		A.10.6.2 A.12.5.4	AC-2 AC-3 AC-4 AC-6 AC-11 AU-13 PE-19 SC-28 SA-8 SI-7	1.2 6.5.5 11.1 11.2 11.3 11.4 A.1
Data Governance - Risk Assessments	DG-08	PO 9.1 PO 9.2 PO 9.4 DS 5.7	45 CFR 164.308(a)(1)(ii)(A) 45 CFR 164.308(a)(8)	Clause 4.2.1 c) & g) Clause 4.2.3 d) Clause 4.3.1 & 4.3.3 Clause 7.2 & 7.3	CA-3 RA-2 RA-3 MP-8 PM-9 SI-12	12.1 12.1.2

Table 3.2: Compliance Map for Classification of Data Breaches Threat

Control Area	Control ID	Scope Applicability from International Standard Organizations				
		COBIT 4.1	HIPAA / HITECH Act	ISO/IEC 27001-2005	NIST SP800-53 R3	PCI DSS v2.0
				A.7.2 A.15.1.1 A.15.1.3 A.15.1.4		
Information Security - Encryption	IS-18	DS5.8 DS5.10 DS5.11	45 CFR 164.312 (a)(2)(iv) 45 CFR 164.312 (e)(1) 45 CFR 164.312 (e)(2)(ii)	A.10.6.1 A.10.8.3 A.10.8.4 A.10.9.2 A.10.9.3 A.12.3.1 A.15.1.3 A.15.1.4	AC-18 IA-3 IA-7 SC-7 SC-8 SC-9 SC-13 SC-16 SC-23 SI-8	2.1.1 3.4 3.4.1 4.1 4.1.1 4.2
Information Security - Encryption Key Management	IS-19	DS5.8	45 CFR 164.312 (a)(2)(iv) 45 CFR 164.312(e)(1)	Clause 4.3.3 A.10.7.3 A.12.3.2 A.15.1.6	SC-12 SC-13 SC-17 SC-28	3.4.1 3.5 3.5.1 3.5.2 3.6 3.6.1 3.6.2 3.6.3 3.6.4 3.6.5 3.6.6 3.6.7 3.6.8
Security Architecture - User ID Credentials	SA-02	DS5.3 DS5.4	45 CFR 164.308(a)(5)(ii)(c) 45 CFR 164.308 (a)(5)(ii)(D) 45 CFR 164.312 (a)(2)(i) 45 CFR 164.312 (a)(2)(iii) 45 CFR 164.312 (d)	A.8.3.3 A.11.1.1 A.11.2.1 A.11.2.3 A.11.2.4 A.11.5.5	AC-1 AC-2 AC-3 AC-11 AU-2 AU-11 IA-1 IA-2 IA-5 IA-6 IA-8 SC-10	8.1 8.2 8.3 8.4 8.5 10.1 12.2 12.3.8

Table 3.2: **Compliance Map for Classification of Data Breaches Threat**

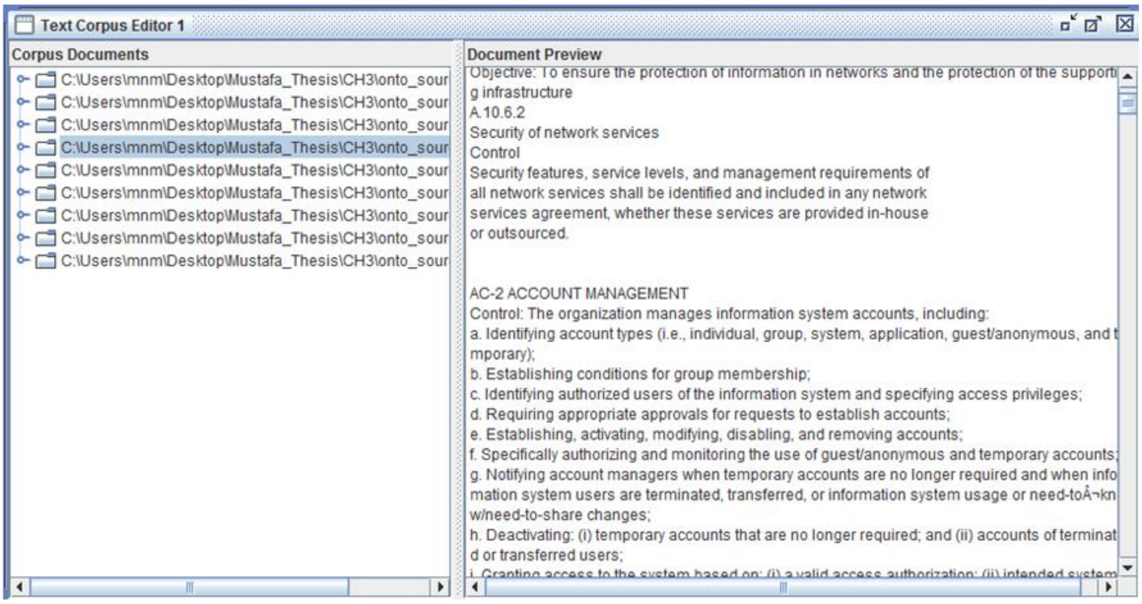
Control Area	Control ID	Scope Applicability from International Standard Organizations				
		COBIT 4.1	HIPAA / HITECH Act	ISO/IEC 27001-2005	NIST SP800-53 R3	PCI DSS v2.0
Security Architecture - Data Security / Integrity	SA-03	DS5.11		A.10.8.1 A.10.8.2 A.11.1.1 A.11.6.1 A.11.4.6 A.12.3.1 A.12.5.4 A.15.1.4	AC-1 AC-4 SC-1 SC-16	2.3 3.4.1 4.1 4.1.1 6.1 6.3.2a 6.5c 8.3 10.5.5 11.5
Security Architecture - Production / Non-Production Environments	SA-06	DS5.7		A.10.1.4 A.10.3.2 A.11.1.1 A.12.5.1 A.12.5.2 A.12.5.3	SC-2	6.4.1 6.4.2
Security Architecture - Remote User Multi-Factor Authentication	SA-07			A.11.1.1 A.11.4.1 A.11.4.2 A.11.4.6 A.11.7.1	AC-17 AC-20 IA-1 IA-2 MA-4	8.3

We have collected the details of each compliance item and related item in the above table into Text file for each CA separately. This means has collected (11) Text files in order to use them in the next section.

3.2.3 Extracting Ontology Concepts

These steps involved extracting ontology concepts; this word isn't easy and protracted procedure. We need to extract the ontology concepts for each CA of the data breaches threat; in order to do this task should be considered about the information knowledge. We has converted each of CA documents (as mentioned above) into a text file, we used KAON TextToOnto tool in order to extract the ontology concepts.

We are talking about the first CA called (Retention Policy) as an example; we added the prepared text corpus (from related documents) to the tool by using the new corpus function as showing in Figure 3.2.



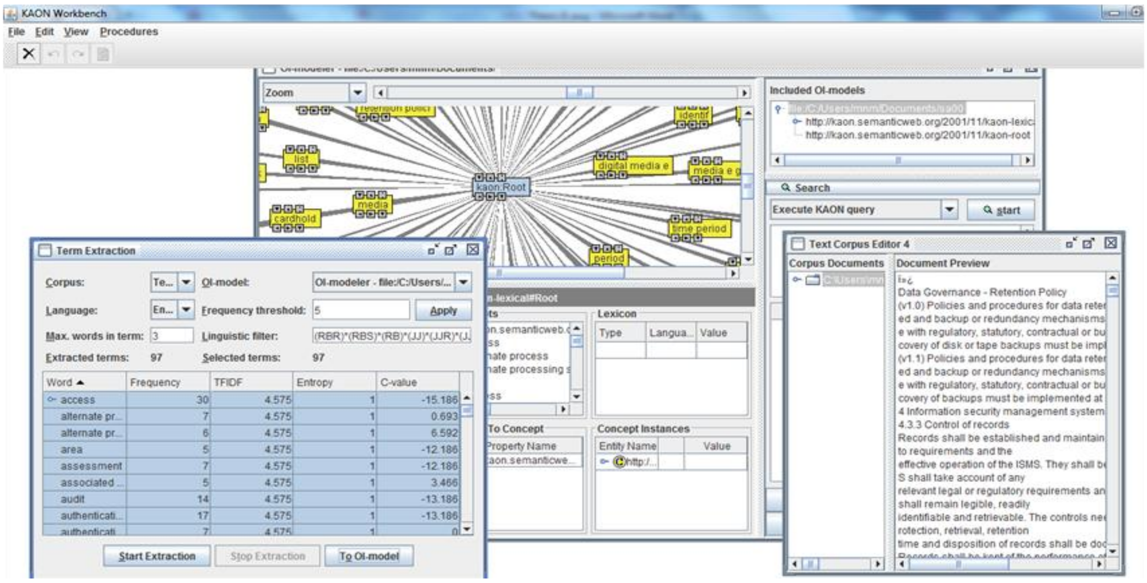
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour
C:\Users\mnm\Desktop\Mustafa_Thesis\CH3\onto_sour

Objective: To ensure the protection of information in networks and the protection of the supporti
g infrastructure
A.10.6.2
Security of network services
Control
Security features, service levels, and management requirements of
all network services shall be identified and included in any network
services agreement, whether these services are provided in-house
or outsourced.

AC-2 ACCOUNT MANAGEMENT
Control: The organization manages information system accounts, including:
a. Identifying account types (i.e., individual, group, system, application, guest/anonymous, and t
mporary);
b. Establishing conditions for group membership;
c. Identifying authorized users of the information system and specifying access privileges;
d. Requiring appropriate approvals for requests to establish accounts;
e. Establishing, activating, modifying, disabling, and removing accounts;
f. Specifically authorizing and monitoring the use of guest/anonymous and temporary accounts;
g. Notifying account managers when temporary accounts are no longer required and when info
mation system users are terminated, transferred, or information system usage or need-toÂ¬kn
w/need-to-share changes;
h. Deactivating: (i) temporary accounts that are no longer required; and (ii) accounts of terminat
d or transferred users;
i. Granting access to the system based on: (i) a valid access authorization; (ii) intended system

Figure 3.2: Create new corpus function using KAON TextToOnto Tools

Later we used the (New Term Extraction) function in order to extract concepts from the provided text corpus as showing in (Figure 3.3).



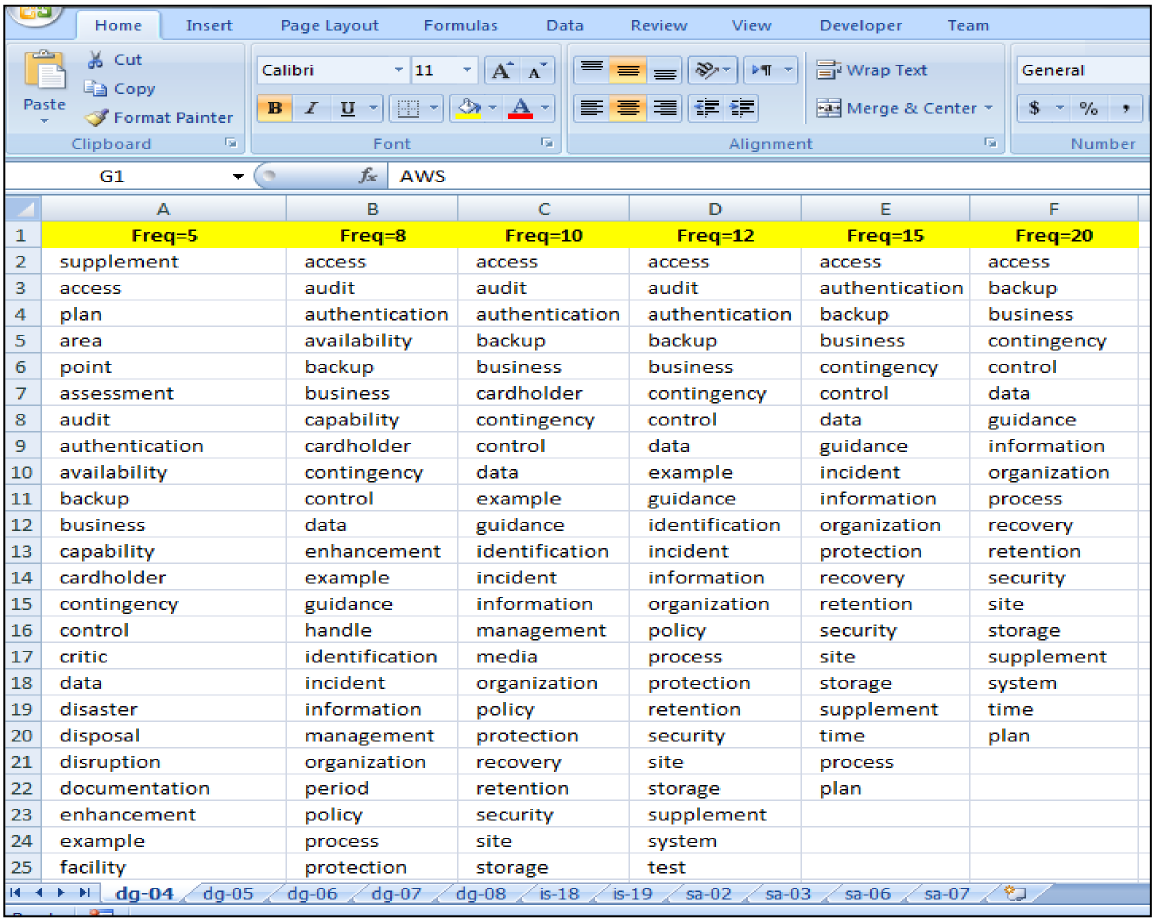
Corpus: Te...
OI-model: OI-modeler - file:/C:/Users/...
Language: En...
Frequency threshold: 5
Max. words in term: 3
Linguistic filter: (RBR)*(RBS)*(RB)*(JJ)*(JJR)*(J...
Extracted terms: 97
Selected terms: 97

Word	Frequency	TFIDF	Entropy	C-value
access	30	4.575	1	-15.186
alternate pr...	7	4.575	1	0.693
alternate pr...	6	4.575	1	6.592
area	5	4.575	1	-12.186
assessment	7	4.575	1	-12.186
associated ...	5	4.575	1	3.466
audit	14	4.575	1	-13.186
authenticati...	17	4.575	1	-13.186
authenticati	7	4.575	1	0

Data Governance - Retention Policy
(v1.0) Policies and procedures for data reter
ed and backup or redundancy mechanisms
e with regulatory, statutory, contractual or bu
covery of disk or tape backups must be impl
(v1.1) Policies and procedures for data reter
ed and backup or redundancy mechanisms
e with regulatory, statutory, contractual or bu
covery of backups must be implemented at
4 Information security management system
4.3.3 Control of records
Records shall be established and maintain
to requirements and the
effective operation of the ISMS. They shall be
S shall take account of any
relevant legal or regulatory requirements an
shall remain legible, readily
identifiable and retrievable. The controls nee
rotection, retrieval, retention
time and disposition of records shall be doc

Figure 3.3: New Term Extraction function using KAON TextToOnto Tool

This tool extracts concepts using parameters; we set the frequency threshold parameter to 5,8,10,12,15 and 20 (Figure 3.4), the number of words for retrieving concepts that on one unique word as a term. The results were 64,42,37,31,27 and 24 concepts respectively. We refined the results of extraction ontology concepts by applying an elimination process for stopping words and characters (it, c, g... etc.). Then the results of the concept after refining process are 57, 37,30,29,23 and 20 concepts respectively.



	A	B	C	D	E	F
1	**Freq=5**	**Freq=8**	**Freq=10**	**Freq=12**	**Freq=15**	**Freq=20**
2	supplement	access	access	access	access	access
3	access	audit	audit	audit	authentication	backup
4	plan	authentication	authentication	authentication	backup	business
5	area	availability	backup	backup	business	contingency
6	point	backup	business	business	contingency	control
7	assessment	business	cardholder	contingency	control	data
8	audit	capability	contingency	control	data	guidance
9	authentication	cardholder	control	data	guidance	information
10	availability	contingency	data	example	incident	organization
11	backup	control	example	guidance	information	process
12	business	data	guidance	identification	organization	recovery
13	capability	enhancement	identification	incident	protection	retention
14	cardholder	example	incident	information	recovery	security
15	contingency	guidance	information	organization	retention	site
16	control	handle	management	policy	security	storage
17	critic	identification	media	process	site	supplement
18	data	incident	organization	protection	storage	system
19	disaster	information	policy	retention	supplement	time
20	disposal	management	protection	security	time	plan
21	disruption	organization	recovery	site	process	
22	documentation	period	retention	storage	plan	
23	enhancement	policy	security	supplement		
24	example	process	site	system		
25	facility	protection	storage	test		

Figure 3.4: Concepts extraction for multi frequency

We present the results of extracting concepts to experts person in the field of (Retention Policy), and accepted for the terms in frequency threshold (5) because it has a set of important terms in the field (like: disaster, documentation, integrity ... etc) (Table 3.1)

Table 3.3: The 57 concepts after refined for DG-04 “Retention Policy”

access	critic	handle	organization	recovery	storage
area	data	identification	paper	response	store
assessment	disaster	impact	part	restoration	supplement
audit	disposal	incident	period	resumption	support
availability	disruption	information	plan	retention	system
backup	documentation	infrastructure	point	risk	test
business	example	integrity	policy	security	time
capability	facility	list	process	service	
contingency	framework	loss	protection	site	
control	guidance	management	record	software	

We repeated the preceding steps for all data breaches threat classification (CAs), Table 3.4 present the ontology concepts of them. The extracted concept number differs from one CA to another depending on the collected document size.

Table 3.4: The Ontology Concepts of the Data Breaches Threat’s Classification

DG-05 - Secure Disposal				**No. of Concepts: 23**	
business	management	supplement			
cardholder	organization	system			
code	policy	use			
control	process				
data	protection				
disposal	retention				
enhancement	risk				
example	sanitization				
guidance	storage				
information	strategy				
DG-06 - Non-Production Data				**No. of Concepts: 32**	
access	effectiveness	plan	system		
analysis	environment	process	use		
assessment	evaluation	production			
organization	extent	remediation			
author	flaw	report			
control	guidance	requirement			
cycle	information	security			
monitoring	management	test			

data	objective	software			
document	part	supplementation			
DG-07 - Information Leakage				**No. of Concepts: 46**	
access	example	management	rest	time	
application	file	network	risk	traffic	
authorization	flow	object	role	transfer	
basis	guidance	operation	rule	type	
control	information	organization	security	usage	
data	integrity	policy	service	user	
destination	intrusion	privilege	session		
enforcement	leakage	process	source		
engineering	level	protection	system		
environment	lock	provider	test		
DG-08 - Risk Assessments				**No. of Concepts: 61**	
access	capability	device	guidance	part	service
account	categorization	documentation	identification	plan	source
application	classification	domain	impact	policy	storage
assessment	component	effect	information	process	strategy
authentication	configuration	enforcement	interconnection	protection	supplement
author	connection	enhancement	interface	protocol	system
authorization	content	example	inventory	review	traffic
basis	control	file	level	risk	transfer
boundary	data	flow	methodology	security	treatment
business	destination	framework	organization	selection	type
					use
IS-18 - Encryption				**No. of Concepts: 56**	
access	data	guidance	network	risk	transfer
application	denial	identification	object	security	transmission
boundary	device	implementation	organization	service	type
capability	difference	incident	packet	session	use
cardholder	distribution	information	pan	source	user
confidentiality	encryption	integrity	policy	supplement	web
connection	enhancement	internet	process	support	
content	example	level	protection	system	
control	file	management	protocol	time	
cryptography	flow	mechanism	response	traffic	
IS-19 - Encryption Key Management				**No. of Concepts: 46**	
access	destruction	information	organization	system	
account	disclosure	integrity	protection	text	
author	disk	isms	provider	time	
cardholder	distribution	key	public	transmission	
certificate	encryption	knowledge	removal	use	
class	end	management	rest	user	
confidentiality	example	manual	security		
control	file	mechanism	service		
cryptography	generation	misuse	storage		
data	guidance	note	supplement		
SA-02 - User ID Credentials				**No. of Concepts: 52**	

access	data	implementation	network	risk	use
activity	default	inactivity	number	security	user
administration	device	information	organization	session	
application	enhancement	level	password	strategy	
audit	example	list	period	supplement	
author	factor	lock	policy	system	
behalf	group	log	process	termination	
connection	guidance	management	registration	time	
content	identification	minimum	remote	type	
control	identity	need	retention	usage	
SA-03 - Data Security/Integrity				**No. of Concepts: 59**	
access	control	exchange	internet	provider	support
application	data	factor	list	remote	system
author	denial	failure	management	removal	traffic
boundary	device	file	monitoring	risk	transfer
cardholder	difference	flow	network	security	transmission
cloud	domain	function	organization	service	type
code	encryption	guidance	policy	software	use
computer	enforcement	information	process	source	user
connection	enhancement	integrity	protection	strategy	web
content	example	interface	protocol	supplement	
SA-06 - Production/Non-Production Environments				**No. of Concepts: 25**	
acceptance	guidance	software			
access	information	system			
business	interface	technology			
control	management	use			
development	network	user			
elasticity	operation				
enhancement	policy				
environment	production				
example	security				
functionality	separation				
SA-07 - Remote User Multi-Factor Authentication				**No. of Concepts: 39**	
access	example	line	role		
appropriation	exterior	mechanism	security		
authority	factor	network	storage		
confidentiality	guidance	object	subject		
connection	impact	organization	supplement		
control	implementation	performance	system		
cryptography	information	policy	trust		
duality	integrity	process	use		
encryption	key	regard	user		
enhancement	level	relevance			

3.3 Cloud Service Provider Security Issue

This section describes the data collection for CSP data breaches threat security action and countermeasures, then extracting the ontology concepts.

3.3.1 Data Collection

We visited different CSP websites and read what they published (documents, sites, white papers.. etc) regarding their security issues and countermeasures. These documents need to read and analysis carefully because some technician points are contained as indirect answers. And in the other hand some CSPs published a few lines regarding their security issues and countermeasures. It leads us to find more clear information. Our research will include five providers: Amazon Web Service (AWS), Windows Azure, Krescendo, CloudSigma and License12. These providers are chosen based on the availability level of data describes security issues, different service, size (small, medium and large), also AWS and Windows CSPs are certified from different international organizations and included the top 100 CSPs ranking (Cloud Times and TalkinCloud); the last three CSPs are out on those ranking web sites. During collection data procedure we associate any terms or items necessary to be defined by more information in the Text file as possible, as showing in Figure 3.4.

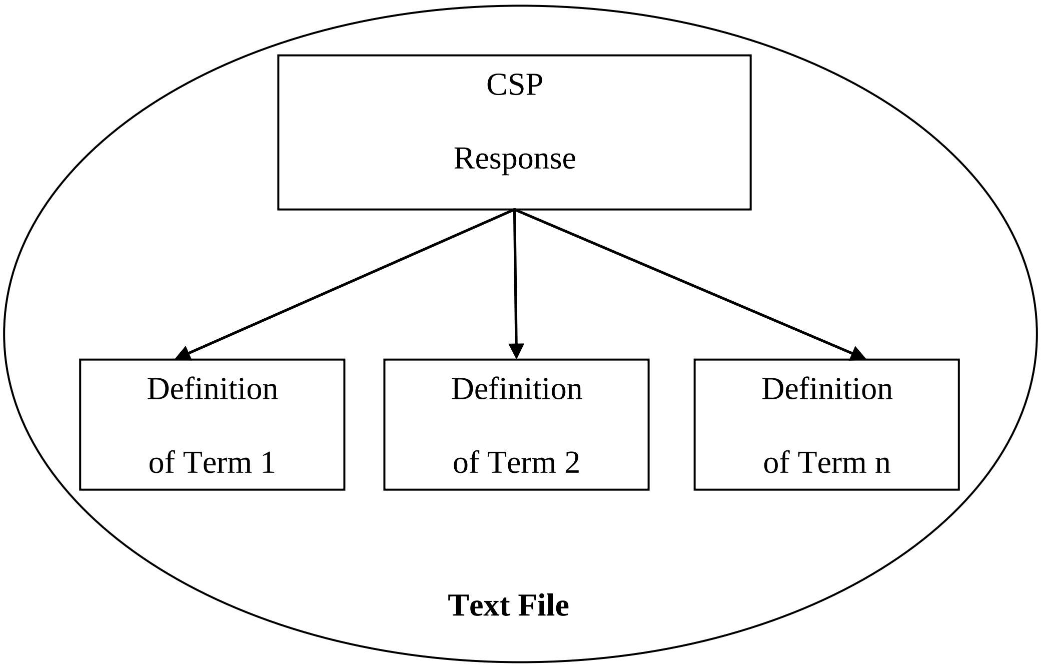


Figure 3.5: Data Collection Procedure of CSP

Here we took AWS and collected their response documents against (Retention Policy); according the above figure we collected the text file. The AWS response it was[2]:

"AWS provide customers with the ability to delete their data. However, AWS customers retain control and ownership of their data so it is the customer's responsibility to manage data retention to their own requirements. Refer to AWS Overview of Security Processes Whitepaper for additional details - available at http://aws.amazon.com/security".

3.3.1 Extracting Ontology Concepts

In this section we extracted ontology concept for a CSP response for each CA. Practically, we extracted ontology concept for the AWS response (as mentioned above) by using the TextToOnto KAON tool for this task. We got (11) concepts as Table 3.3 with frequency threshold (1):

[2] http://media.amazonwebservices.com/AWS_Risk_and_Compliance_Whitepaper.pdf

Table 3.5 : Results of extracting ontology concepts for AWS

Threshold Frequency=1

AWS Concepts	AWS Concepts
availability	privacy
backup	relation
data	retention
database	storage
law	time
option	

It’s necessary to get more accurate concepts when associates the above text (AWS response) with more information (Figure 3.4) regarding the (Retention Policy) from the AWS website as mentioned in their “http://aws.amazon.com/security” response. Then we got (18) concepts as Table 3.4 with frequency threshold (2):

Table 3.6 : Results of extracting ontology concepts for AWS

Threshold Frequency=2

AWS Concepts	AWS Concepts
availability	production
backup	recovery
business	relation
capacity	retention
data	service
database	storage
enforcement	time
instance	
law	
option	
period	

We repeated the preceding steps for all CSP with their response for each CA, as shows in Table 3.6. The extracted concept number differs from one CSP to another depending on the collected document size.

Table 3.7 : Ontology Concepts of CSPs for each CAs

1) DG-04 - Retention Policy

No.	AWS	Azure	License12	Krescendo	CloudSigma
1	availability	backup	addition	access	completion
2	backup	business	data	agency	distribution
3	business	center	internality	business	enhancement
4	capacity	customer	regularity	client	industry
5	data	data	standard	compliance	management
6	database	disaster	time	data	material
7	enforcement	domain		government	notification
8	instance	event		information	
9	law	example		request	
10	option	fault		retention	
11	period	help		service	
12	production	information		solution	
13	recovery	infrastructure		specification	
14	relation	loss		system	
15	retention	machine		view	
16	service	platform			
17	storage	program			
18	time	recovery			
19		redundancy			
20		replication			
21		restoration			
22		retention			
23		review			
24		service			
25		state			
26		storage			
27		tolerance			
28		validation			

2) DG-05 - Secure Disposal

No.	AWS	Azure	License12	Krescendo	CloudSigma
1	contract	asset	classification	archive	archive
2	control	contract	data	basis	basis
3	data	data	device	client	client
4	device	disposal	distribution	data	destruction
5	disposition	distribution	information	destruction	organization
6	distribution	guidance	material	exit	practice
7	enhancement	information	notification	hardware	security
8	example	management	operation	policy	
9	information	method	proposal	removal	
10	method	organization	sanitization	security	
11	operation	paper	schedule	specific	
12	proposal	practice	section		
13	purpose	process	standard		
14	retention	protection	storage		
15	risk	sanitization	success		
16	security	storage	termination		
17	standard	strategy	time		
18	storage	system			
19	success	use			
20	system				
21	time				

3) DG-06 - Non-Production Data

No.	AWS	Azure	License12	Krescendo	CloudSigma
1	ability	access	ability	data	customer
2	administration	customer	administration	development	data

3	assessment	data	author	moment	department
4	author	department	customer	production	development
5	customer	development	data	site	direction
6	data	direction	information	testing	information
7	information	duty	ownership	time	operation
8	organization	environment	round		policy
9	ownership	information	session		principle
10	production	investigation			review
11	requirement	movement			segregation
12	responsibility	operation			service
13	round	policy			
14	session	principle			
15	system	production			
16	use	protection			
17		review			
18		segregation			
19		service			
20		system			
21		test			

4) DG-07 - Information Leakage

No.	AWS	Azure	License12	Krescendo	CloudSigma
1	access	administration	solution	access	assessment
2	architecture	assessment	stand	accident	climate
3	assessment	climate		addition	domain
4	authorization	depth		building	environs
5	basis	domain		card	greatness
6	content	environs		data	information
7	contentment	foundation		electron	
8	control	greatness		employee	
9	customer	information		exterior	
10	domai	order		fobs	
11	environment	security		key	
12	facility	system		leakage	
13	leakage	transfer		office	
14	management	web		part	
15	order			risk	
16	process			security	
17	rest			setup	
18	risk			site	
19	security			standard	
20	system			storage	
21	theft			unit	
22	traffic			weekend	
23	transfer				
24	unit				
25	version				

5) DG-08 - Risk Assessments

No.	AWS	Azure	License12	Krescendo	CloudSigma
1	account	assessment		access	assessment
2	authorization	availability		aide	availability
3	auditor	classification		application	confidentiality
4	consideration	clause		change	information
5	organization	data		compliance	integrity
6	information	domain		detection	software
7	guidance	evaluation		environment	
8	integrity	hardware		extension	
9	software	impact		integrity	
10	data	information		intrusion	
11	network	review		management	
12	group	risk		network	
13	use	software		office	
14	system	author		operation	

15	access	management		package	
16		use		penetration	
17				place	
18				policy	
19				support	
20				system	
21				vulnerability	

6) IS-18 - Encryption

No.	AWS	Azure	License12	Krescendo	CloudSigma
1	access	access	access	acknowledgement	data
2	care	application	code	certificate	email
3	code	care	data	client	hardware
4	connection	connection	domain	data	key
5	control	customer	example	email	public
6	data	data	key	hardware	route
7	example	domain	operation	key	stage
8	key	encryption	option	offer	
9	operation	information	part	public	
10	part	option	review	route	
11	security	part	security	security	
12	service	replication	theory	stage	
13	side	review		traffic	
14	store	security			
15	system	side			
16	theory				
17	transfer				

7) IS-19 - Encryption Key Management

No.	AWS	AZURE	License12	Krescendo	CloudSigma
1	alignment	data	alignment	client	alignment
2	certificate	domain	certification	computer	computer
3	certification	effect	data	data	data
4	control	encryption	domain	encryption	encryption
5	data	information	effect	environment	environment
6	domain	key	information	information	information
7	information	management	management	offering	offering
8	key	review	option	production	production
9	management	service	party	rest	server
10	option	storage	provider	standard	
11	organization	support	review		
12	party	transmission	security		
13	provider		server		
14	security		service		
15	server		side		
16	service		storage		
17	side				
18	storage				
19	system				
20	text				
21	transmission				
22	user				

8) SA-02 - User ID Credentials

No.	AWS	AZURE	License12	Krescendo	CloudSigma
1	access	activity	corporation	access	data
2	administration	corporation	custom	client	encryption
3	corporation	domain	domain	corporation	password
4	custom	entropy	infrastructure	data	traffic
5	factor	expiry	management	domain	user
6	management	guess	option	encryption	
7	option	information	review	entropy	
8	section	infrastructure	security	guess	
9	security	management	service	integration	
10	service	minimum	user	password	

11	system	organization		place	
12	time	policy		review	
13	use	review		traffic	
14	user	section		user	
15		security		way	
16		strength			
17		user			

9) SA-03 - Data Security/Integrity

No.	AWS	AZURE	License12	Krescendo	CloudSigma
1	architecture	access	architecture	client	access
2	author	author	author	compliance	architecture
3	certification	certification	certification	context	client
4	cloud	cloud	data	data	compliance
5	control	contractor	domain	security	contractor
6	data	data	industry		data
7	domain	exchange	information		industry
8	factor	factor	review		information
9	industry	information			review
10	information	production			
11	security	review			
12	standard	security			
13	transfer	staff			
14		standard			

10) SA-06 - Production/Non-Production Environments

No.	AWS	AZURE	License12	Krescendo	CloudSigma
1	ability	business	ability	interior	storage
2	access	development	access	production	source
3	business	domain	energy	testing	information
4	demand	environment	environment		flexibility
5	development	exchange	flexibility		energy
6	education	information	guidance		access
7	energy	production	information		
8	enhancement	relevance	source		
9	environment	review	storage		
10	guidance	separation			
11	information	stage			
12	line				
13	model				
14	operation				
15	power				
16	production				
17	responsibility				
18	security				
19	separation				
20	storage				
21	system				
22	time				
23	use				

11) SA-07 - Remote User Multi-Factor

No.	AWS	AZURE	License12	Krescendo	CloudSigma
1	access	access	corporation	access	
2	account	corporation	information	client	
3	authentication	customer	key	corporation	
4	authority	direction	review	data	
5	connection	domain	setting	domain	
6	control	factor	staff	encryption	
7	customer	information	terminal	entropy	
8	environment	key	user	guess	
9	factor	network		integration	
10	feature	policy		password	
11	hardware	review		place	
12	identity	setting		review	

13	increase	staff		traffic	
14	individual	support		user	
15	information	terminal		way	
16	key	use			
17	level	user			
18	management				
19	network				
20	password				
21	policy				
22	regard				
23	second				
24	security				
25	support				
26	system				
27	use				
28	user				

The details of the final step, matching process between concepts which we present how calculate the matching percentage is mentioned in the next chapter.

Chapter 4

Proposed Model and Matching Process

4.1 Overview

This chapter explains, in detail, the use of the proposed model. This will be done through the design and implementation of A Reference Model to Measure the Effectiveness of CC Security. This model measures the security compliance of the CSP semantics and matching their response to major international compliance guidance.

4.2 Proposed Model

Our proposed model consists of three phases:

1) Extracting ontology concepts of CC threat.
2) Extracting ontology concepts of CSP.
3) Matching Process between both ontology concepts.

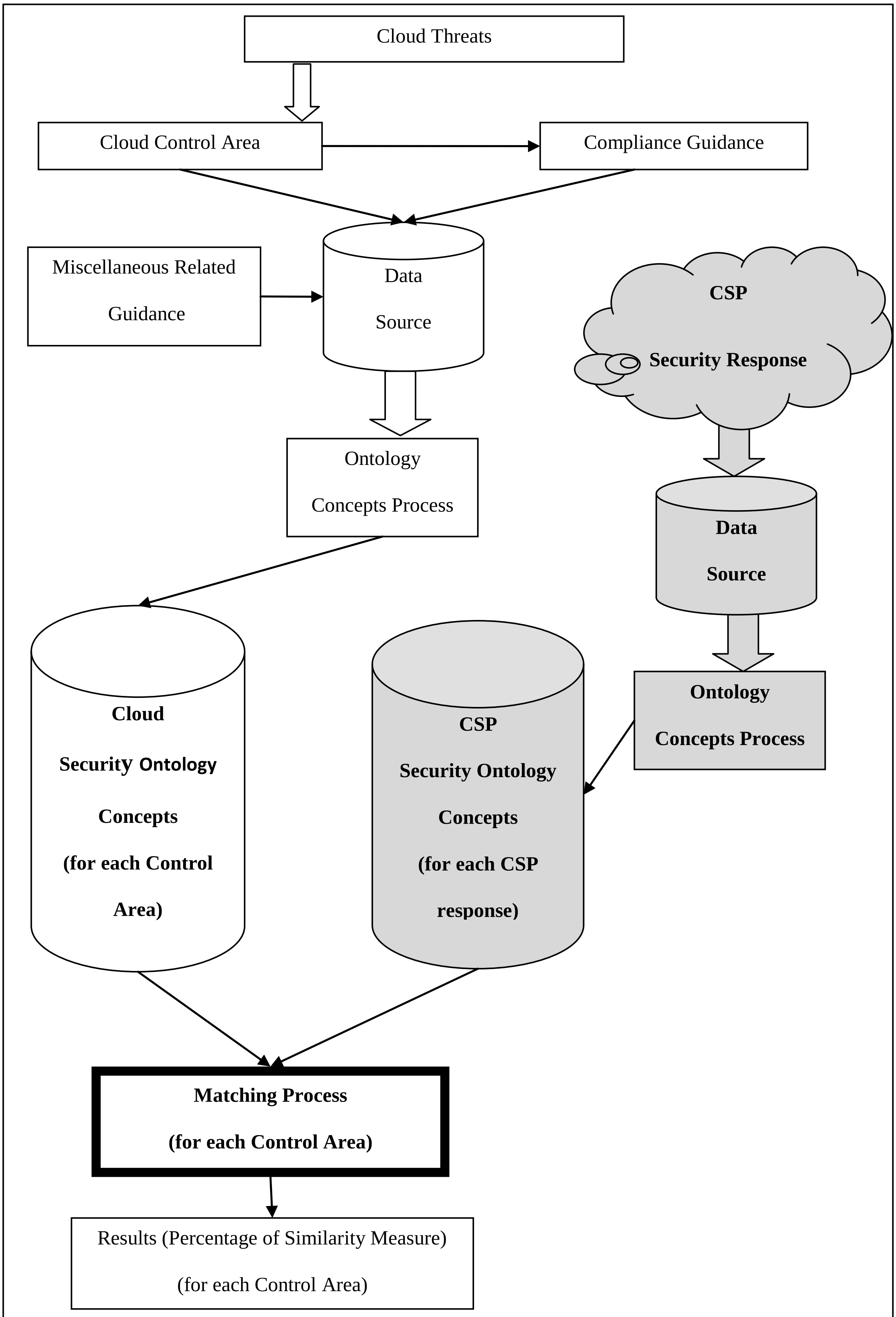

Cloud Threats
Cloud Control Area
Compliance Guidance
Miscellaneous Related Guidance
Data Source
CSP Security Response
Data Source
Ontology Concepts Process
Ontology Concepts Process
Cloud Security Ontology Concepts (for each Control Area)
CSP Security Ontology Concepts (for each CSP response)
Matching Process (for each Control Area)
Results (Percentage of Similarity Measure) (for each Control Area)

Figure 4.1: Proposed Model to Measure Effectiveness of CC Security

4.2.1 Phase One

The goal of this part is to get the ontology concepts extractions for each CA. We have taken 11 CA in our study. These CAs describe their specifications in CCM (Appendix B)

4.2.1.1 Cloud Threats

In this work have addressed which CC threat can be measured. As mentioned earlier we studied the (Data Breaches Threat), CSA reported "TOP 10 CC Threats in 2013", and (Data Breaches Threat) was the 1st threat should be considered from CSPs and consumers. In this report can be defined the main parameters (CA) that cover this threat.

4.2.1.2 Cloud Control Area

This section addressed by two parts:

1- **Cloud Control Matrix (CCM):** This is designed to provide fundamental security principles to guide CSP and to assist prospective cloud customers in assessing the overall security risk of a CSP. The CCM provides a controls framework that gives a detailed understanding of security concepts and principles that are aligned to the Cloud Security Alliance guidance in 13 domains" (Cloud Security Alliance, 2013).

2- **Consensus Assessments Initiative Questionnaire (CAIQ):** This effort is focused on providing industry-accepted ways to document what security controls exist in IaaS, PaaS, and SaaS offerings, providing security control transparency, and it's associated with CCM for more descriptions about CA with questions (Cloud Security Alliance, 2011).

4.2.1.3 Compliance Guidance

This part represents a several international standard compliance guidance items that associated with each CA, complete items presented in Appendix B.

4.2.1.4 Miscellaneous Related Guidance

This part of several efforts has already taken place to offer guidance for cloud security. These include (Winkler, V. J. 2011):

1- **Security Guidance for Critical Areas of Focus in CC V3.0:** Published in 2011 presented security guidance for a number of areas in cloud computing; these include architecture, governance, traditional security, and virtualization. (Cloud Security Alliance, 2011).

2- **Domain 12: Guidance for Identity & Access Management V2.1:** Published in April 2010 discusses the major identity management functions as they relate to cloud computing. This work forms a cornerstone of the CSA’s Trusted Cloud Initiative (Kumaraswamy, S. , 2010).

3- **Cloud Computing: Information Assurance Framework:** Published in November 2009. Presents a set of assurance criteria that address the risk of adopting cloud computing (Catteddu, D., 2009).

4- **The Federal CIO Council’s** Proposed Security Assessment and Authorization for U.S. Government Cloud Computing. The core importance of this document is that it adopts the NIST 800-53R3 security controls for cloud computing in low- and moderate-risk systems (Council I. A., 2012).

4.2.1.5 Data Source

It contains a set of the text files which are collected from the above documents.

4.2.1.6 Cloud Security Ontology Concepts (for each Control Area):

It contains the ontology concept extraction (by KAON TextToOnto Tool) for each CA separately, in our work we were extracted ontology concepts for 11 CA (Appendix C).

4.2.2 Phase Two

The goal of this part is to get the ontology concepts extractions for each CSPs response. We have taken 5 CSPs in our study.

4.2.1.1 CSP Security Response

It contains a different security response or actions of the CSPs for their compliances for each CA. We have taken 5 CSPs and searches to find 55 responses.

4.2.1.2 Data Source

It contains a set of the text files which are collected from the above section.

4.2.1.3 Cloud Security Ontology Concepts (for each Control Area):

It contains the ontology concept extraction (by KAON TextToOnto Tool) for each CSPs has a response for 11 CA separately, in our work we were extracted 55 state of ontology concepts totally. (Appendix C).

4.2.3 Phase Three

This is the last component, it presents the matching process between the each output of the ontology concepts in part1 with the each output of the ontology concepts in part2 (Figure 4.2), then present the results.

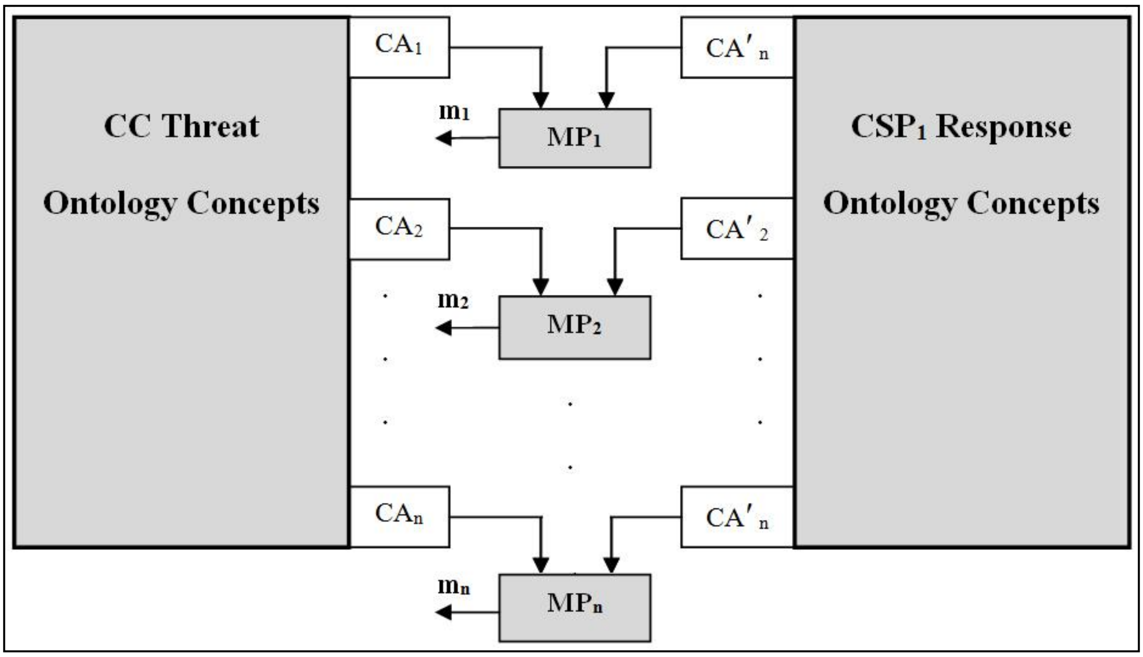


Figure 4.2: Matching Concepts Process

The symbolic description of the (Figure 4.2) is :

CA_1 …. CA_n	:	Ontology concepts of CC Threat “Control Area CA”
CA'_1 …. CA'_n	:	Ontology concepts of CSP response for each CA
MP_1 …. MP_n	:	Semantic similarity measure between concepts of CA and CA′
m_1 …. m_n	:	Results (Total Measure Ratio)
n	:	Number of CA

4.2.3.1 Matching Process

As mentioned before about the matching concept extraction. We have used:

<u>First:</u>

Exact matching between the concepts extractions as below; we took for ontology concepts (e.g., CA_1 and CA′1):

Table 4.1: Ontology Concepts for CA_1 “Retention Policy”

No.	Concepts	No.	Concepts	No.	Concepts	No.	Concepts
1	access	16	documentation	31	organization	46	risk
2	area	17	example	32	paper	47	security
3	assessment	18	facility	33	part	**48**	**service**
4	audit	19	framework	**34**	**period**	49	site
5	**availability**	20	guidance	35	plan	50	software
6	**backup**	21	handle	36	point	**51**	**storage**
7	**business**	22	identification	37	policy	52	store
8	**capability**	23	impact	38	process	53	supplement
9	contingency	24	incident	**39**	**protection**	54	support
10	control	25	information	40	record	55	system
11	critic	26	infrastructure	**41**	**recovery**	56	test
12	**data**	27	integrity	42	response	**57**	**time**
13	disaster	28	list	43	restoration		
14	disposal	29	loss	44	resumption		
15	disruption	30	management	**45**	**retention**		

Table 4.2: Ontology Concepts for AWS response (CA'_1)
“Retention Policy”

No.	Concepts	No.	Concepts
1	**availability**	**16**	**service**
2	**backup**	**17**	**storage**
3	**business**	**18**	**time**
4	**capacity**		
5	**data**		
6	database		
7	enforcement		
8	instance		
9	law		
10	option		
11	**period**		
12	production		
13	**recovery**		
14	relation		
15	**retention**		

When we see for the above two tables and matching measure between the concepts, we find (11 from 18) concepts are exact matching (in the gray highlight).

In order to compute the matching ration:

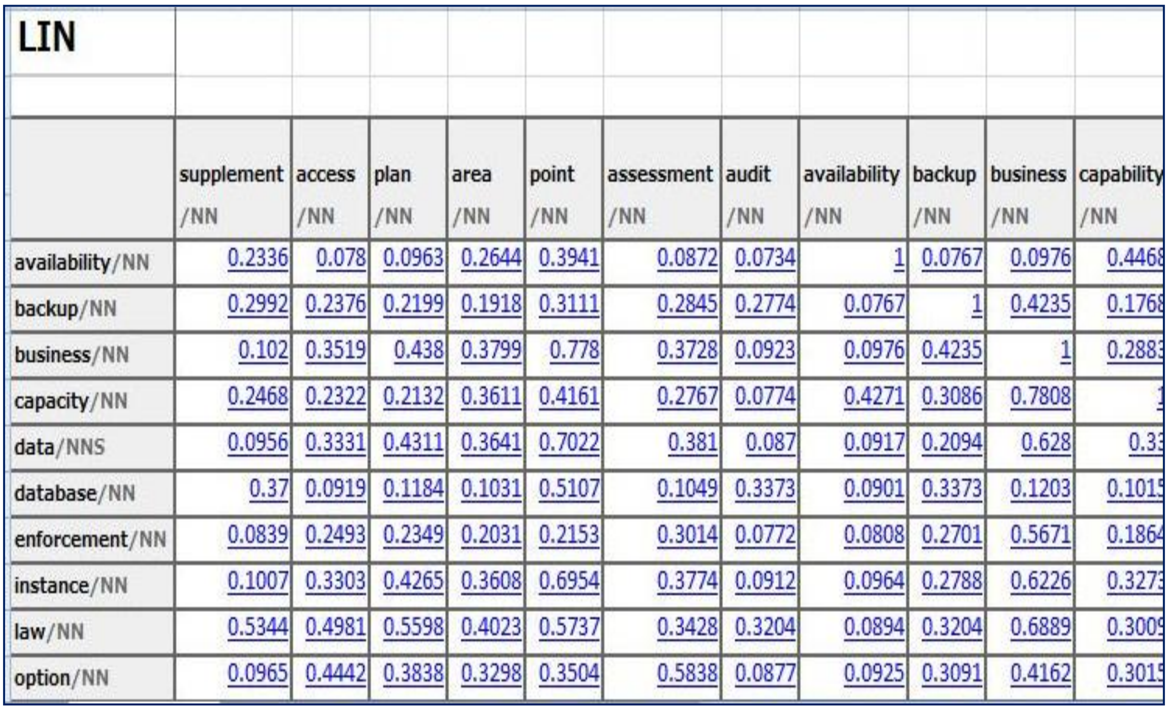
$$\text{Exact Matching Ratio} = \frac{\text{No. of Concepts by Exact Matching}}{\text{No. of Concepts of CA}'1} \qquad \text{....... (1)}$$

$$\text{Exact Matching Ratio} = \frac{11}{18}$$

$$\text{Exact Matching Ratio} = 0.611$$

Second:

We applied Semantic Similarity Measure (SSM) for matching between concepts by using WS4J Tool (Figure 4.3). We have used similarity measure developed by Lin (Lin, D., 1998) and it is intended to be useful in nearly any environment (Warin, M., 2004) with accepted correction value is (0.834) and it is range (0-1) (Jarmasz M., 2012).



LIN

	supplement /NN	access /NN	plan /NN	area /NN	point /NN	assessment /NN	audit /NN	availability /NN	backup /NN	business /NN	capability /NN
availability/NN	0.2336	0.078	0.0963	0.2644	0.3941	0.0872	0.0734	1	0.0767	0.0976	0.4468
backup/NN	0.2992	0.2376	0.2199	0.1918	0.3111	0.2845	0.2774	0.0767	1	0.4235	0.1768
business/NN	0.102	0.3519	0.438	0.3799	0.778	0.3728	0.0923	0.0976	0.4235	1	0.2883
capacity/NN	0.2468	0.2322	0.2132	0.3611	0.4161	0.2767	0.0774	0.4271	0.3086	0.7808	1
data/NNS	0.0956	0.3331	0.4311	0.3641	0.7022	0.381	0.087	0.0917	0.2094	0.628	0.33
database/NN	0.37	0.0919	0.1184	0.1031	0.5107	0.1049	0.3373	0.0901	0.3373	0.1203	0.1015
enforcement/NN	0.0839	0.2493	0.2349	0.2031	0.2153	0.3014	0.0772	0.0808	0.2701	0.5671	0.1864
instance/NN	0.1007	0.3303	0.4265	0.3608	0.6954	0.3774	0.0912	0.0964	0.2788	0.6226	0.3273
law/NN	0.5344	0.4981	0.5598	0.4023	0.5737	0.3428	0.3204	0.0894	0.3204	0.6889	0.3009
option/NN	0.0965	0.4442	0.3838	0.3298	0.3504	0.5838	0.0877	0.0925	0.3091	0.4162	0.3015

Figure 4.3: Sample for using Lin Similarity Measure for AWS by WS4J Tool

After applied the Lin similarity measure we found (4) new concepts are similar with variant similar correction value (Table 4.3):

Table 4.3: Correction value for new (4) concepts

AWS Concepts **(CA′$_{1)}$**	**Ontology Concepts for** **CA$_1$**	**Lin** **Similarity Measure** **(Correction Value)**
database	list	1
instance	example	1
production	business	0.9209
relation	part	0.9182

We used the below equation In order to compute the total measure ratio ($m_{1}..m_{11}$):

$$\text{Total Measure } = \frac{\text{No. of Concepts by Lin Similarity Mesure}}{\text{No. of Concepts of CA}'1} \qquad \text{...... (2)}$$

$$\text{Total Measure } = \frac{15}{18}$$

$$\text{Total Measure} = 0.833$$

We applied the above equations for all CSP, and compare it with human judges (professors, doctors, and practitioners we asked them during our study) (Table 4.4 – Table 4.8)

Table 4.4: AWS Measure Ratio

No.	CA	Human Ratio	Total Measure Ratio ($m_1..m_{11}$)	Exact Matching Ratio	No. of Concepts by SSM	No. of Concepts
1	DG-04	0.950	0.833	0.611	15	18
2	DG-05	0.900	0.857	0.380	18	21
3	DG-06	0.850	0.750	0.562	12	16
4	DG-07	0.850	0.760	0.600	19	25
5	DG-08	0.950	0.800	0.533	12	15
6	IS-18	0.900	0.764	0.529	13	17
7	IS-19	0.900	0.772	0.727	17	22
8	SA-02	0.850	0.785	0.642	11	14
9	SA-03	0.900	0.846	0.692	11	13
10	SA-06	0.950	0.913	0.565	21	23
11	SA-07	0.950	0.750	0.535	21	28
	Average	**0.905**	**0.803**	**0.580**		

Table 4.5: Azure Measure Ratio

No.	CA	Human Ratio	Total Measure Ratio ($m_1..m_{11}$)	Exact Matching Ratio	No. of Concepts by SSM	No. of Concepts
1	DG-04	0.800	0.714	0.392	20	28
2	DG-05	0.950	0.894	0.684	17	19
3	DG-06	0.800	0.727	0.304	16	22
4	DG-07	0.850	0.714	0.357	10	14
5	DG-08	0.900	0.812	0.687	13	16
6	IS-18	0.850	0.733	0.400	11	15
7	IS-19	0.800	0.706	0.583	12	17
8	SA-02	0.800	0.750	0.411	15	20
9	SA-03	0.850	0.714	0.500	10	14
10	SA-06	0.800	0.733	0.500	11	15
11	SA-07	0.900	0.705	0.471	12	17
Average		**0.845**	**0.746**	**0.481**		

Table 4.6: License12 Measure Ratio

No.	CA	Human Ratio	Total Measure Ratio ($m_1..m_{11}$)	Exact Matching Ratio	No. of Concepts by SSM	No. of Concepts
1	DG04	0.500	0.500	0.333	3	6
2	DG-05	0.500	0.470	0.235	8	17
3	DG-06	0.650	0.555	0.333	5	9
4	DG-07	-	-	-	-	-
5	DG-08	-	-	-	-	-
6	IS-18	0.500	0.500	0.333	6	12
7	IS-19	0.500	0.562	0.500	9	16
8	SA-02	0.500	0.400	0.300	4	10
9	SA-03	0.700	0.625	0.500	5	8
10	SA-06	0.650	0.555	0.444	5	9
11	SA-07	-	-	-	-	-
Average		**0.563**	**0.521**	**0.372**		

Table 4.7: Krescendo Measure Ratio

No.	CA	Human Ratio	Total Measure Ratio ($m_1..m_{11}$)	Exact Matching Ratio	No. of Concepts by SSM	No. of Concepts
1	DG-04	0.550	0.466	0.467	7	15
2	DG-05	0.500	0.363	0.182	4	11
3	DG-06	0.600	0.714	0.286	5	7
4	DG-07	0.550	0.454	0.182	10	22
5	DG-08	0.600	0.523	0.095	11	21
6	IS-18	0.500	0.538	0.231	7	13
7	IS-19	0.600	0.500	0.300	5	10
8	SA-02	0.550	0.400	0.200	6	15
9	SA-03	0.700	0.800	0.400	4	5
10	SA-06	0.400	0.333	0.333	1	3
11	SA-07	0.550	0.400	0.200	6	15
Average		**0.555**	**0.499**	**0.261**		

Table 4.8: CloudSigma Measure Ratio

No.	CA	Human Ratio	Total Measure Ratio ($m_1..m_{11}$)	Exact Matching Ratio	No. of Concepts by SSM	No. of Concepts
1	DG-04	0.500	0.571	0.29	4	7
2	DG-05	0.300	0.428	0.14	3	7
3	DG-06	0.400	0.583	0.17	7	12
4	DG-07	0.550	0.428	0.29	3	7
5	DG-08	0.100	0.333	0.33	2	6
6	IS-18	0.300	0.285	0.14	2	7
7	IS-19	0.300	0.222	0.22	2	9
8	SA-02	0.550	0.400	0.40	2	5
9	SA-03	0.400	0.444	0.22	4	9
10	SA-06	0.600	0.500	0.03	3	6
11	SA-07	0.550	0.375	0.38	3	8
Average		**0.414**	**0.415**	**0.237**		

4.3 Experimental Results and Discussion

1- We have computed the total error among human judge's ratio and total measure ratio (m_1 .. m_{11}) by using Mean Square Error (MSE):

$$\text{MES} = \frac{\sum_n^1 (\text{Human Ratio} - \text{Total Measure Ratio})^2}{\text{n}} \times 100$$

Table 4.9 and Figure 4.4 presents the error percentage using MES for all CSPs of (11 CAs)

Table 4.9: Average Error for MSE (11 CAs)

Threat Domain	Error Percentage (%)				
	AWS	Azure	License12	Krescendo	CloudSigma
Data Breaches	1.268	1.164	0.480	1.135	1.776
Average Error	**1.165 %**				

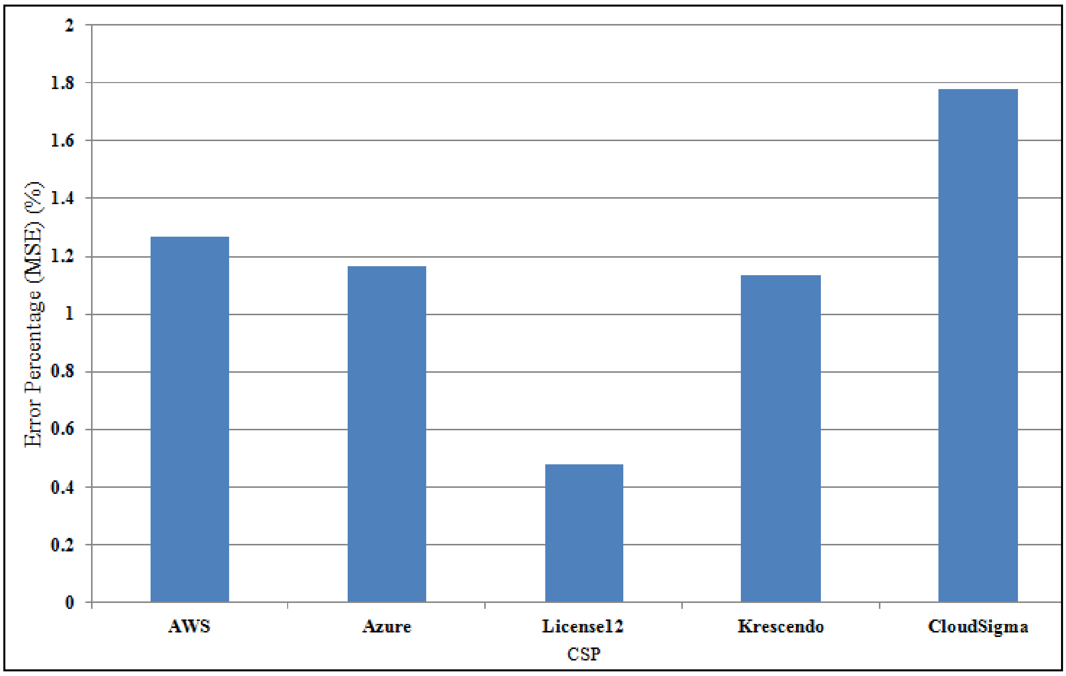


Figure 4.4: Average Error for MSE (11 CAs)

We note that the error percentage of (License12) is (0.480%) about the half percentage from other CSPs due its security compliance's for (8) CAs. Therefore, we have computed the error percentage for the (8) participates CAs.

2- Table 4.10 and Figure 4.5 present the error percentage using MES for all CSPs of (8 CAs).

Table 4.10: Average Error for MSE (8 CAs)

Threat Domain	Error Percentage (%)				
	AWS	Azure	License12	Krescendo	CloudSigma
Data Breaches	0.861	0.797	0.480	1.090	1.195
Average Error	**0.885 %**				

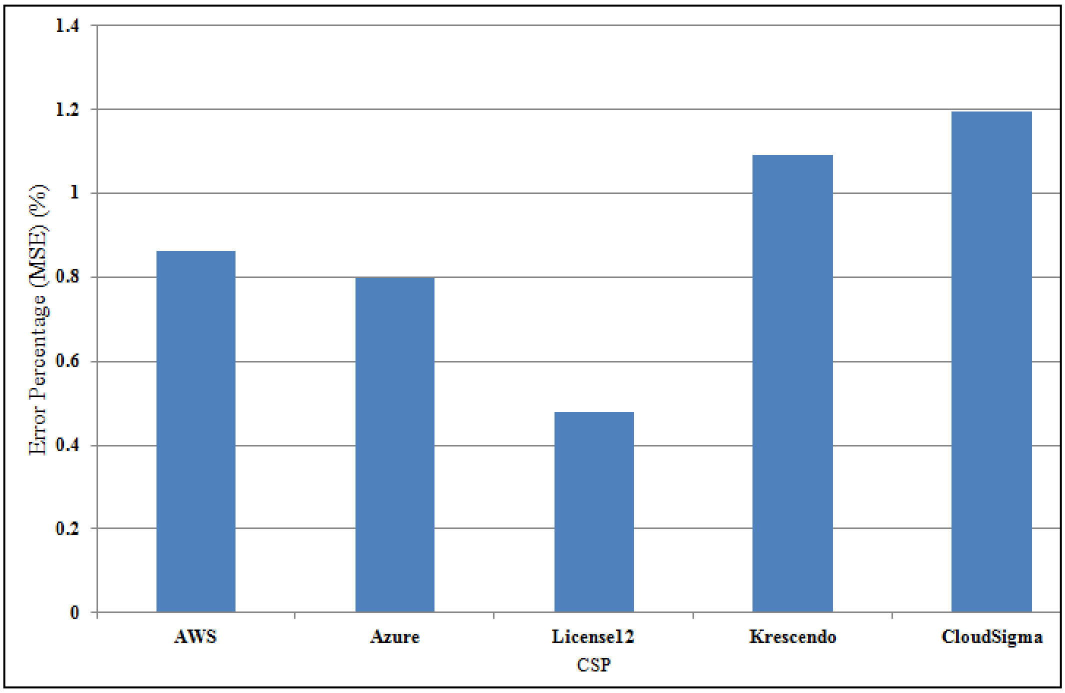


Figure 4.5: Average Error for MSE (8 CAs)

The error percentage of (AWS) and (Azure) have decreased, and the last two CSP remained almost the same percentage. We believe due availability of the published security compliance documents.

3- Table 4.11 and Figure 4.6 presents the measurement comparison between human and our measure (Total Measure).

Table 4.11: Average Error for MSE (8 CAs)

Threat Domain	Average (%)									
	AWS		Azure		License12		Krescendo		CloudSigma	
	Human	SSM	Human	SSM	Human	SSM	Human	SSM	Human	SSM
Data Breaches	**90.5**	**80.3**	**84.5**	**74.6**	**56.3**	**52.1**	**55.5**	**49.9**	**41.4**	**41.5**

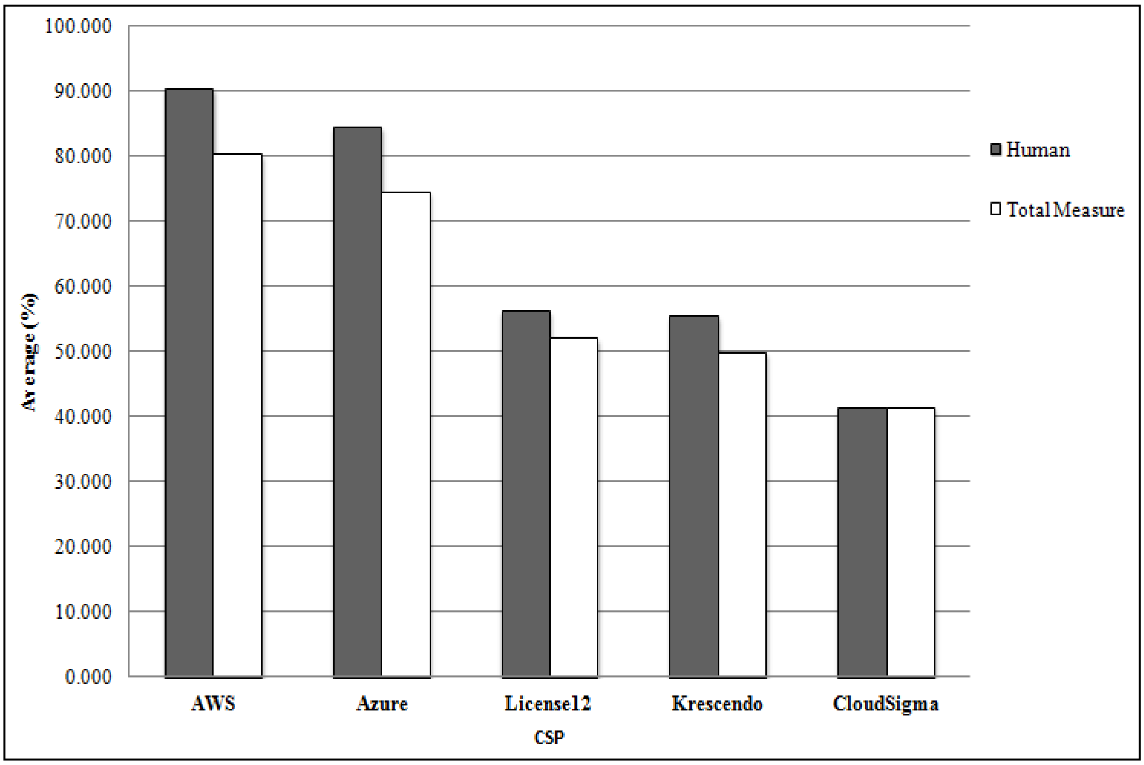


Figure 4.6: Average Error for MSE (11 CAs)

The highest score of the security compliance is for (AWS) and (Azure) due they are certified from international standard organizations that support our measure.

4.3 Validation

We have validated and compared our approach measurement with other approach. This approach presented in paper (Bhensook N., 2012) by using Goal Question Metric Approach (GQM). The authors used CCM (Cloud Security Alliance, 2013); each domain as goal, used some metrics by questions then make score for each answer. Authors have scored two domains for example in their experimental approach on AW (G-compliance = 20%.) and (G-05 Information Security = 40%.), these scores for example to use their approach. The reliability of assessment results depend on the security information that is provided by cloud providers and experts questions.

CHAPTER FIVE

Conclusions and Future work

5.1 Conclusions

1- The reliability of the semantic measure result depends on the security information that is provided from CSP.

2- CSPs do not disclose more about his security issues.

3- The highest score to security compliance is for CSP (AWS and Azure), due they has certified from the major international standard organizations (Table 4.11).

4- Number of ontology concepts assigns the level of the security compliance.

5- Uncertified CSP has a limitation for their security response.

6- Human Judge Percentage score some time less than our semantic similarity measure in uncertified CSP due limitation for their security information (Table 4.7).

7- Some concepts matching have full matching by using Lin similarity measure (Table 4.3) that enhances our measure.

5.2 Summary

1- In our research we studied the CC, its security, ontology concepts and the semantic similarity measure. Also presented the importance of the CC security.

2- We defined data breaches threat in CC and classified its parameters. These parameters are (11) called (Control Area) are covered data breaches threat; each (Control Area) is a part with domain of (Cloud Control Matrix).

3- In order to extracting ontology concepts, we collected data (documents, reports, white papers... etc) for both (Control Areas) and CSP then prepared text corpora from them to be used in a tool to extract them. In our study we have taken (5) CSP. Then converted all documents into Text file.

4- Extraction ontology concepts for the (11) CA and CSPs security action for each CA. that is mean we worked (11) ontology concepts for (CA) and (55) ontology concepts for CSPs. Then we refined them from stopping words.

5- We have matched the concepts first with exact matching then by semantic similarity measure (by Lin similarity measure) among the concepts of (Control Area) and CSPs. Then we checked the enhancement among them.

6- We presented the results for each CSPs and CA, the total error is (0.885 %) Mean Square Error.

5.3 Future Work

Through conducting this research, many ideas and issues were unfolded but not accomplished yet because of time, resources, and other constraints. We would like to suggest a few ideas for future study:

1- Possibility to use our approach to measure other CC threats (like Account Hijacking, Data Loss ... etc) semantically.

2- Build full otology domain for each CA as security requirements.

3- Providing full coverage for all ontology domains, to let the measure be more accurate and reliable.

4- Develop a graphics user interface to present the results.

References

Almorsy M., Grundy J., & Müller, I. (2010). An analysis of the cloud computing security problem. In Proceedings of APSEC 2010 Cloud Workshop, Sydney, Australia, 30 th Nov 2010, (On Line), available:
http://www.cs.auckland.ac.nz/~john-g/papers/cloud2010_1.pdf

Baker, W., Hutton, A., Hylender, C. D., Pamula, J., Porter, C., & Spitler, M. (2011). Data breach investigations report. Verizon RISK Team. [Online].Available:
www.verizonbusiness.com/resources/reports/rp_databreach-investigations-report-2011_en_xg. pdf

Bhensook, N., & Senivongse, T. (2012). An assessment of security requirements compliance of cloud providers. In Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference 520-525, Doi: 10.1109/CloudCom.2012.6427484

Budanitsky, A., & Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. Computational Linguistics, 32 (1), 13-47.

Catteddu, D., & Hogben, G. (2009). Cloud computing information assurance framework. European Network and Information Security Agency (ENISA).

Che, J., Duan Y., Zhang T., & Fan, J. (2011). Study on the security models and strategies of cloud computing. Procedia Engineering. 23, 586-593, [Online].Available:
http://www.ccbc.ir/files_site/files/r_16_130217100828.pdf.

Chen, Z., & Yoon, J. (2010). It auditing to assure a secure cloud computing. In Services (SERVICES-1), 2010 6th World Congress, 253-259, Doi: 10.1109/SERVICES.2010.118.

Cloud Security Alliance (2011). Security guidance for critical areas of focus in cloud computing v3. 0. Cloud Security Alliance, (On-Line), available: https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf

Cloud Security Alliance (2011). Top 10 threats in Cloud Computing, (On-Line), available:
https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf

Cloud Security Alliance (2013). Cloud Control Matrix v1.4., (On-Line), available: https://cloudsecurityalliance.org/research/ccm

Cloud Security Alliance (2013). Consensus Assessments Initiative Questionnaire V1.1, (On-Line), available:
https://cloudsecurityalliance.org/research/cai/

Cloud Security Alliance (2013). Top 10 threats in Cloud Computing, (On-Line), available: https://downloads.cloudsecurityalliance.org/initiatives/top_threats/The_Notorious_Nine_Cloud_Computing_Top_Threats_in_2013.pdf

Council, I. A. (2012). Federal risk and authorization management program (FedRAMP).

Curran, k., Carlin, S., & Adams, M. (2011). Security issues in cloud computing. Elixir Network Engg, 38, 4069-4072, Doi: 10.4018/978-1-4666-0957-0.ch014.

Dai Yuefa, W. B., Yaqiang, G., Quan, Z., & Chaojing, T. (2009). Data security model for cloud computing. In Proceedings of the 2009 International Workshop on Information Security and Application (IWISA 2009) Qingdao, China, 141-144.

El-Khameesy, N., & Rahman, H. A. (2012). A proposed model for enhancing data storage security in cloud computing systems. Journal of Emerging Trends in Computing and Information Sciences, 3 (6), 970-974.

Fenz, S. (2010). Ontology-based generation of it-security metrics. Proceedings of the 2010 ACM Symposium on Applied Computing, 1833-1839.

HIPAA., 1996. Health Insurance Portability and Accountability Act, Pub. L.No.104-191, 110 Stat. (codified as amended in scattered sections of 42 U.S.C. and 29 U.S.C.).

Jarmasz, M. (2003). Roget's thesaurus as a lexical resource for natural language processing. arXiv preprint arXiv:1204.0140.

Jensen, M., Schwenk, J., Gruschka, N., & Iacono, L. L. (2009). On technical security issues in cloud computing. In IEEE International Conference on Cloud Computing. 109-116, Doi: 10.1109/CLOUD.2009.60

Jiang, R. (2013). From ontology to semantic similarity: calculation of ontology-based semantic similarity. The Scientific World Journal, 2013.

Jing, X., & Jian-jun, Z. (2010). A brief survey on the security model of cloud computing. In Distributed Computing and Applications to Business Engineering and Science (DCABES), 2010 Ninth International Symposium, 475-478, Doi: 10.1109/DCABES.2010.103

Julisch, K. (2009). Security compliance: the next frontier in security research, In Proceedings of the 2008 workshop on New security paradigms.
71-74, doi:10.1145/1595676.1595687.

Kandukuri, B. R., Paturi, V. R., & Rakshit, A. (2009). Cloud security issues. In Services Computing, 2009. SCC'09. IEEE International Conference, 517-520, Doi:10.1109/SCC.2009.84

Kumar, V., Swetha, M., Muneshwara, M. S., & Prakash, S. (2012). Cloud computing: towards case study of data security mechanism. Int. J. Adv. Technol. Eng. Res, 2 (4).

Kumaraswamy, S., Lakshminarayanan, S., Stein, M. R. J., & Wilson, Y. (2010). Domain 12: Guidance for Identity & Access management V2. 1. Cloud Security Alliance, (On-Line), available:
http://www. cloudsecurityalliance. org/guidance/csaguide-dom12-v2, 10.

Luna, J., Ghani, H., Germanus, D., & Suri, N. (2011). A security metrics framework for the cloud. In Proc. of the INSTICC Internation Conference on Security and Cryptography, 245-250.

Maedche, A., & Volz, R. (2001). The ontology extraction & maintenance framework Text-To-Onto. In Proc. Workshop on Integrating Data Mining and Knowledge Management, USA, 1-12.

Malik, A., & Nazir, M. M. (2012). Security framework for cloud computing environment: a review. Journal of Emerging Trends in Computing and Information Sciences, 3 (3).

Mell, P., & Grance, T. (2011). The nist definition of cloud computing, english, special publication, Us Department Of Commerce, 800 (145), 7.

Michelizzi, J. (2005). Semantic relatedness applied to all words sense disambiguation, (Doctoral dissertation), University of Minnesota, Twin Cities, U.S., (On-Line), available:
http://www.d.umn.edu/~tpederse/Pubs/jason-thesis.pdf

Mohamed, E. M., Abdelkader, H. S., & El-Etriby, S. (2012). Enhanced data security model for cloud computing. In Informatics and Systems (INFOS), 2012 8th International Conference on CC-12 - CC-1

Nagwani, N., & Verma, S. (2011). A frequent term and semantic similarity based single document text summarization algorithm, International Journal of Computer Applications. 17 (2), 36-40, Doi: 10.1109/ICTKE.2012.6152388

Pal, D. G. (2012). A aovel open security framework for cloud computing, International Journal of Cloud Computing and Services Science (IJ-CLOSER). 1 (2), 45-52.

Parekh, M. D. H., & Sridaran, R. (2013). An analysis of security challenges in cloud computing, (IJACSA) International Journal of Advanced Computer Science and Applications. 4 (1), 38-46.

Putri, N. R., & Mganga, M. C. (2011). Enhancing information security in cloud computing services using SLA based metrics (Doctoral dissertation). Master's thesis: Blekinge Institute of Technology.

Ristov, S., Gusev, M., & Kostoska, M. (2012). Cloud computing security in business information systems, International Journal of Network Security & Its Applications (IJNSA). 4 (2), 75-93.

Shuanglin, R. (2012), Data security policy in the cloud computing. In Computer Science & Education (ICCSE), 2012 7th International Conference, 222-225, Doi: 10.1109/ICCSE.2012.6295062

Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing , Journal of Network and Computer Applications. 34 (1), 1-11.

Sullivan, K., Clarke, J., & Mulcahy, B. P. (2010). Trust-terms ontology for defining security requirements and metrics. In Proceedings of the Fourth European Conference on Software Architecture,
175-180, doi.10.1145/1842752.1842789.

Talib, A. M., Atan, R., Abdullah, R., & Murad, M. A. A. (2012). Security ontology driven multi agent system architecture for cloud data storage security: ontology development. JCSNS, 12 (5), 63.

Tariq, M. I. (2012). Towards information security metrics framework for cloud computing. International Journal of Cloud Computing and Services Science (IJ-CLOSER), 1 (4), 209-217.

Townsend, M. (2009). Managing a security program in a cloud computing environment. In 2009 Information Security Curriculum Development Conference, 128-133, Doi: 10.1145/1940976.1941001

Ullah, K. W. (2012). Automated security compliance tool for the cloud, (Doctoral dissertation). Norwegian university of science and technology, Norwegian University, Trondheim.

Warin, M. (2004). Using WordNet and Semantic Similarity to Disambiguate an Ontology. Retrieved January, 25, 2008, (On-Line), available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.9729&rep=rep1&type=pdf

Winkler, V. J. (2011). Securing the cloud: cloud computer security techniques and tactics. Elsevier.

Yildiz, M., Abawajy, J., Ercan, T., & Bernoth, A. (2009). A Layered security approach for cloud computing infrastructure. In Pervasive Systems, Algorithms, and Networks (ISPAN), 10th International Symposium ,763-767, Doi: 10.1109/I-SPAN.2009.157

Yuefa, D., Bo W. , Yaqiang, G., Quan, Z., & Chaojing, T. (2009).” Data security model for cloud computing”. Proceedings of the 2009 International Workshop on Information Security and Application (IWISA 2009), (On-Line), available:
http://www.academypublisher.com/proc/iwisa09/papers/iwisa09p141.pdf

Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues, Future Generation Computer Systems, 28 (3), 583-592.

Websites References

1- www.cloudsecurityalliance.org

2- www.talkincloud.com

3- www.wikipedia.com

4- www.cloudtimes.com

5- www.aws.amazon.com

6- www.windowsazure.com

7- www.krescendo.com

8- www.cloudsigma.com