IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Towards an Ontology of Data Breaches Threat for Cloud Computing

Mohammed Noaman Murad Alhasan[1], Mustafa N. AL-Hassan[2]

Lecturer, Department of Computer Science, Cihan University, Erbil, Iraq[1]

Database Administrator, Manaf Hamza Engineering & Surveying Company (SPC), Manama, Kingdom of Bahrain[2]

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 than before. The data breaches threat for Cloud Computing is ranking No.1 and high risk level. This research proposed to build ontology of data breaches threat for Cloud Computing based on the concepts. This paper also shows how discover the aspects items which are related with our domain "Data Breaches Threat". We collected huge data and extracted it into "concepts" using KAON tool.

Keywords: Cloud Computing, Ontology, Data Breach, KAON.

I. INTRODUCTION

1.1. Cloud Computing

The most widely used definition of the Cloud Computing model is introduced by National Institute of Standards and Technology (NIST) 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."[1]

1.2. Data Breaches Threat

A Data Breach is the intentional or unintentional release of secure information in an un trusted environment. Other terms for this phenomenon include unintentional information disclosure, data leak and also data spill.[2] In Cloud Computing and according the Cloud Security. Alliance (CSA) report "Top 10 threats in Cloud Computing"; the data breaches threat is ranking No.5 in 2011 and No.1 in 2013[3]; and it's in the high risk level Risk Matrix, as shown in Fig. 1.andFig. 2.

Fig. 1Data Breaches Threat Top Ranking 2013

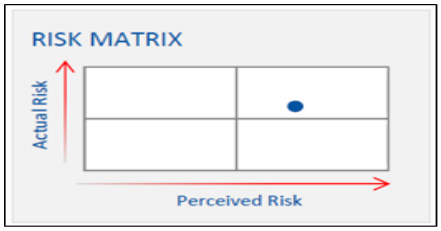

Fig. 2Data Breaches Threat Top Ranking 2013

1.2. Ontology

Ontology has its origin in the field of philosophy where it refers to the study of existence. In computer science ontologies define theories of what exist. There are several ambiguous and similar definitions that have been given to the word ontology in different field of study such as Artificial Intelligence, software engineering, information system, knowledge engineering etc.[4]In[5] Gruber defines Ontology as: "An ontology is an explicit specification of a conceptualization". This definition of ontology is said to create ambiguity mainly because of its brevity. The confusion of such definition might also be the fact that it use sterms that are already ambiguous and difficult to understand for someone new to the Ontology community. According to Gruber in [5] a conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge level agent is committed to some conceptualization, explicitly or implicitly. Specification refers to definitions of classes, relations, functions, and other objects which make the Ontology.

1.3. The motivation

Data Braches Threat in Cloud Computing is high risk level thus it is important to analyse its knowledge as ontology domain, share common understanding of the structure of information among people or software agents, enable reuse of domain knowledge, make domain assumptions explicit, separate domain knowledge from the operational knowledge.[9].

II. BUILDING THE ONTOLOGY

We have divided the methodologies for building Ontologies around three major stages of the ontology life cycle: Building, Manipulating, and Maintaining,as shown in Fig. 3.[6] These three stages are overlapped in our

DOI 10.17148/IJARCCE.2016.5501

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

research into two phases. Ellipses in Fig. 1 represent the inner steps for each stage. The Data Breaches Threat in Cloud Computing domain was selected for this research. The input is a set of documents. It is collected from several resources such as online reports, white papers, and academic research.

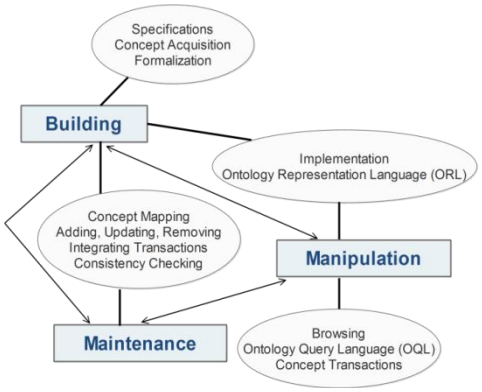

Fig. 3 Ontology Life Cycle

2.1 The First Phase: Data Collection

Cloud Security Alliance (CSA) is the main source for our data breaches threat for cloud computing ontology. CSA is the world's leading organization dedicated to defining and raising awareness of best practices to help ensure a secure cloud computing environment.[7]

CSA team has designed two main documents:

A. Cloud Security Alliance Cloud Controls Matrix (CCM) This matrix 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).[10].

B. 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 forCloud Computing. The data breachesthreatfor CC associated with the (11) parameters called Controls Area (CA) .[11],are shown in Table 1.

TABLE 1: THE CLASSIFICATION FOR DATA BREACHES THREAT

No.	Control Area (CA)	Control ID
1	Data Governance - Retention Policy	DG-04
2	Data Governance - Secure Disposal	DG-05
3	Data Governance - Non-Production Data	DG-06
4	Data Governance - Information Leakage	DG-07
5	Data Governance - Risk Assessments	DG-08
6	Information Security - Encryption	IS-18
7	Information Security - Encryption Key Management	IS-19
8	Security Architecture - User ID Credentials	SA-02
9	Security Architecture - Data Security / Integrity	SA-03
10	Security Architecture - Production / Non-Production Environments	SA-06
11	Security Architecture - Remote User Multi-Factor Authentication	SA-07

Each (CA) are complies group of items issued from major international standard organizations. The Compliance Map for Classification of Data Breaches Threat for Cloud Computing is shown in Table 2.

TABLE 2: COMPLIANCE MAP FOR CLASSIFICATION OF DATA BREACHES THREAT

Control Area (CA)	Control ID	Scope Applicability from International Standard Organizations	
		COBIT 4.1	ISO/IEC 27001-2005
Data Governance - Retention Policy	DG-04	DS 4.1 DS 4.2 DS 4.5 DS 4.9 DS 11.6	Clause 4.3.3 A.10.5.1 A.10.7.3
Data Governance - Secure Disposal	DG-05	DS 11.4	A.9.2.6 A.10.7.2
Data Governance - Non-Production Data	DG-06		A.7.1.3 A.10.1.4 A.12.4.2 A.12.5.1
Data Governance - Information Leakage	DG-07	DS 11.6	A.10.6.2 A.12.5.4
Data Governance - Risk Assessments	DG-08	PO 9.1 PO 9.2 PO 9.4 DS 5.7	Clause 4.2.1 c) & g) Clause 4.2.3 d) Clause 4.3.1 & 4.3.3 Clause 7.2 & 7.3 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	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
Information Security - Encryption Key Management	IS-19	DS5.8	Clause 4.3.3 A.10.7.3 A.12.3.2 A.15.1.6
Security Architecture - User ID Credentials	SA-02	DS5.3 DS5.4	A.8.3.3 A.11.1.1 A.11.2.1 A.11.2.3 A.11.2.4 A.11.5.5
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
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
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

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 phase.

2.2The Second Phase: Extracting Ontology Concepts and their Relationship

This phase involved extracting ontology concepts and builds relationship between them. 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

DOI 10.17148/IJARCCE.2016.5501

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

KAON Text To Onto tool in order to extract the ontology concepts. Text To Onto [8] 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.

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 shown in Fig. 4.

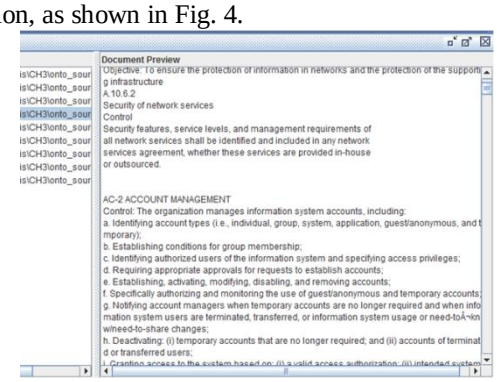

Fig. 4.Create new corpus function using KAON TextToOntoTools

Later we used the (New Term Extraction) function in order to extract concepts from the provided text, as shown in Fig. 5.

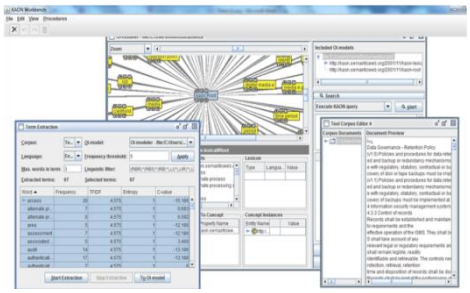

Fig. 5.New Term Extraction function using KAON TextToOntoTool

KAON TextToOntoTool extracts concepts using parameters; we set the frequency threshold parameter to 5,8,10,12,15 and 20), 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. KAON TextToOnto Tool was used to extract relationships between the extracted concepts. The tool was provided with the text corpus which was prepared previously and with the concepts which want to study the relationships between them. Around 145concepts have been extracted.

The following is just a sample: disposal, protection, system, risk , storage, encryption, analysis, identification , access, control, policy, retention, service, network, interface, information, transmission, capability, enforcement, session, integrity, use, cardholder, enhancement, source, factor,, guidance, confidentiality, data, lock, device, key, encryption , code, level, etc.

III. CONCLUSION

The building new Ontology domain from scratch is not easy work because it needs valued data reflect the related aspects for the domain. We study the use of existing resources (data and tools)to build ontology of Data Breaches Threat for Cloud Computing. For refining the concepts and relation between them depends on expert in field with more effort. For future work when follow our proposed method can be build an ontology for other domain such as Data Loss in Cloud Computing or Service Traffic Hijacking in Cloud Computing.

REFERENCES

[1]. The NIST definition of cloud computing, US Department Of Commerce, 800 (145), 7, 2011.
[2]. Johnson L., Computer Incident Response and Forensics Team Management: Conducting a Successful Incident Response, 1st Ed. Waltham, USA: Syngress, 2014.
[3]. Cloud Security Alliance (2013). Top 10 threats in Cloud Computing[Online]. Available:
[4]. https://downloads.cloudsecurityalliance.org/initiatives/top_threats/
[5]. The_Notorious_NineCloud_Computing_Top_Threats_in_2013.pdf
[6]. Ntieche, K. S., "Towards an ontology for System Administration. Case Study: Backup Operation", UNIVERSITY OF OSLO, Oslo, 2007.
[7]. Thomas R. Gruber. A Translation Approach to Portable Ontology Specifications. September 1992, Revised April 1993
[8]. [Ahmad Kayed and Robert Colomb. Extracting ontological concepts for tendering conceptual structures. Data and Knowledge Engineering, 40(1):71-89, 2002.
[9]. Cloud Security Alliance. (2016) homepage . [Online]. Available: https://cloudsecurityalliance.org.
[10]. Maedche, A., &Volz, R, The ontology extraction & maintenance framework Text-To-Onto. In Proc. Workshop on Integrating Data Mining and Knowledge Management, USA, 1-12, 2001.
[11]. Natalya F. Noy and Deborah L. McGuinness (2002) Ontology Development 101: A Guide to Creating Your First Ontology, homepage [Online].Available: http:// protege. stanford.edu/ publications/ontology development/ontology101-noy-mcguinness. html
[12]. Cloud Security Alliance (2014). Cloud Control Matrix v3.0.1.,[Online]. Available: https:// cloud security alliance.org/ research/ccm
[13]. Cloud Security Alliance (2011). Security guidance for critical areas of focus in cloud computing v3. 0. Cloud Security Alliance, ([Online]. Available: https:// cloud security alliance.org /guidance /csaguide.v3.0.pdf

BIOGRAPHY

Mr. Mohammed Noaman Murad Alhasan, MS.Sc. Computer Science Lecturer, Department of Computer Science, Cihan University, Erbil, Iraq.

Mr. Mustafa N. AL-Hassan, MS.Sc. Computer Information Systems, Database Administrator, Manaf Hamza Engineering & Surveying Company (SPC), Manama, Kingdom of Bahrain.