International Journal of Engineering & Technology, 7 (2.33) (2018) 127-132

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A study on the trend of cloud service and security through text mining technique

Seong-Taek Park [1] *, Sung-Won Lee [2], Tae-Gu Kang [3]

[1] Department of MIS, Chungbuk National University, Chungdae-ro 1, Seowon-gu, Cheongju, Chungbuk, 28644, South Korea
[2] Department of Business Administration, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul, 02504, South Korea
[3] Department of Business Administration, Konyang Cyber University, 158Gwanjeo-dong, Seo-gu, Daejeon, 35365, South Korea
**Corresponding author E-mail: solpherd@cbnu.ac.kr*

Abstract

Background/Objectives: The purpose of this study is to recognize the significance of security of cloud service activation and success and to draw the direction and importance for the development of Korean cloud service. Cloud computing service makes it possible to conduct a test easily and quickly at a low cost, to be scalable with an inexpensive cost, and to curb unnecessary expenses.
Methods/Statistical analysis: This study collected news articles from NAVER with the use of crawling technique. The collected data were pre-processed (cleaned), and then 25 words with a high priority were analyzed in the category of the first half of 2015, the second half of 2015, and the first half of 2016.
Findings: To find the trend of cloud computing service security, this study collected news articles and analyzed them after data cleaning. The analysis revealed that there was a difference in importance of each period. In the first half of 2015, the 2nd half of 2015, and the 1st half of 2016, the common noun words extracted were cloud, service, security, and enterprise in order.
The difference was that the word 'environment' was found in the 1st half of 2015, the word 'management' in the 2nd half of 2015, and the words 'customer' and 'use' in the 1st half of 2016.
Improvements/Applications: According to the analysis, there was a difference in importance by year. It is considered that the study results will be able to serve as the guidelines for establishing a systematic plan of cloud service security in the firms and institutions providing cloud service.

Keywords: *Big Data; Cloud Service; Cloud Security; Text Mining, Crawling.*

1. Introduction

The conventional ICT environment tends to be changed quickly to the cloud virtualization environment that guarantees scalability, availability, agility, and cost efficiency in the virtualization of hardware resources. In the global market, Amazon's cloud already leads the cloud market-[1-2]. These days, Google, IBM, MS, Apple, and other firms put a lot of efforts to dominate the cloud market. Besides, the advanced governments design and support various policies to expand cloud service.
Since the Act on Promotion of Cloud Computing and User Protection (Cloud Act) was established in 2015 [3], the Korean government has made an effort to promote cloud. The cloud market is divided into the one for enterprises and the other for general purpose. Nevertheless, it has yet to be used actively.
Gartner, a market survey institution, predicted that 10% of the functions of the security products for enterprises would be supplied in the type of cloud by 2015. As such, the cloud service has shown the fastest growth globally as the ICT environment changed to the mobile environment. Generally, cloud computing makes it easier for users to produce, share, and consume contents.
Nevertheless, there are concerns over cloud service promotion, such as service stability, data security, and data confidentiality. Since cloud service was used by mobile devices, a lot of security threats, such as hacking, DDoS, attacks using virtualization vulnerabilities, information leakage by device loss, and stealing of a user account, have appeared[4].
In a keynote speech at 2011 Apple Worldwide Developers Conference (WWDC), there was an argument that the center of the digital era would move from PC environment to Cloud environment, and iCloud was introduced [five] in the past, cloud was considered to be the exclusive property of enterprises.
Now, with the emergence of iCloud that enables general consumers to access pictures, music, documents, and application programs regardless of time and space, the cloud era has emerged more quickly.
However, although the cloud service has many advantages, it still has security problems.
For instance, in iCloud known to be relatively safe against hacking, pictures of celebrities were hacked with their user accounts and then exposed outside in 2014. In cloud computing service, various types of data are saved into a server connected to the internet, and a user is able to access relevant data at any times and any places after accessing the server with its user account and password.
The cloud computing service is vulnerable to hacking. Therefore, to expand the cloud service, it is prerequisite to prepare for security threats.
In Dec. 2013, the UK government announced G-Cloud guidelines of security principles for introducing cloud products and service in public organizations [6]. In Oct. 2015, the US government announced the guidelines of security control for cloud computing service providers hosting the sensitive information of the government.

Up to now, previous studies on cloud service security mostly focused on cloud service security requirements (functions), the trend of international standardization, and system security requirements. However, there is no research on the trend of cloud service security using text mining. Therefore, this study crawled and pre-processed (cleaned) news articles (NAVER) of cloud service security and analyzed them with R.

This study is comprised of as follows: chapter 2 describes the theoretical background of cloud, cloud service, cloud security, and text minding, and previous studies; chapter 3 presents the study method and procedure; chapter 4 shows the overall process of data collection and cleaning, and visualized final results; chapter 5 describes the conclusion of this study, suggestions, and research limitations.

2. Theoretical background

2.1. Cloud computing service

With the development of wireless devices and network technology, people are able to access internet at any times and any places so as to find proper information[7].

Cloud computing service is a type of service in which a user is able to hire ICT resources with no need of installing them in its own server. It means a computing environment where people can use ICT services, such as data saving, network use, contents use (apps, music, pictures, etc.).

In the cloud computing environment, a user is able to hire ICT resources timely from a cloud computing service provider without owning them. Therefore, the use of ICT resources can be improved and idle resources can be minimized. As a result, it is possible to save the ICT infrastructure introduction cost [8].

Cloud computing has some types of sub concepts, which are cloud infrastructure service (IaaS) providing computing system or network; cloud platform service (PaaS) providing a platform or a solution environment for user computing; and cloud application service (SaaS) providing a software application [9, 10, 11, 12].

Aside from the diversity of services, sharing of different cloud platform resources, cloud brokers for business, and cloud exchanges are emerged[7].

Cloud computing enables a user to hire ICT resources timely from a cloud computing service provider without owning them. Therefore, the use of ICT resources can be improved and idle resources can be minimized. As a result, it is possible to save the ICT infrastructure introduction cost.

2.2. Cloud security threat factors

2.2.1. Security issue

For cloud computing service, many different providers offer services to individuals and firms (through login with account and password). Since personal information is saved in the cloud environment, the issue of security arises [11].

In the cloud environment, there are various kinds of security threats. As shown in the case of iCloud, personal information can be leaked outside. Generally, cloud service uses a virtual machine so that it always includes eavesdropping, malignant code spread and infection, resource exhaustion attacks, and denial of service attacks [12].

Therefore, by expanding the concept of traditionally applied security, it is necessary to change it properly for cloud computing [13].

2.2.2 Security treats

The key security factors of cloud service drawn from the cases of the representative security threats suggested by NIST in the risk level of cloud customers are presented as follows [4].

They are virtualization vulnerability, the risk of information leakage by committed information, service failure by resource sharing and concentration, information leakage by the diversity of devices, difficult security application by distributed processing, and problems with regulations (Table 1).

Table1:. Security Threats

Security Threats factors	Feature
Virtualization vulnerability	Threats of malicious code infection and spread. Infringement on service availability.
The risk of information leakage by committed information	Information leak by separation between possession. Management, information leak by insiders.
Service failure by resource sharing and concentration	Service disruptions against all customers in case of system failure. Vulnerability to DDoS attacks, etc., in case of central control system exposure.
Information leakage by the diversity of devices	Information leak by loss of terminal unit, etc.
Difficult security application by distributed processing	Increased complexity in authentication/access control due to resource sharing and dynamic relocation of virtual machine. Difficulty in applying lump-sum authentication/access control to distributed computing system.
Problems with regulations	Unclear responsibility when information is leaked. Difficulty in audit trail due to resource sharing.

The nine security threats of the cloud environment announced by CSA were data leakage, data loss, stealing of account and service, unsafe API, service denial, malicious internal use, abuse of cloud service, poor understanding of cloud service, and vulnerability of sharing technology [14].

The security threat factors provided by CSA are mostly related to human rather than technical thing. This means that despite the presence of technical countermeasures against threats, intended threats by insiders are difficult to be prevented and so insider security awareness education is seen as a very important factor for cloud service security.

2.3. Texting

Unlike structured data, unstructured data has no data model defined previously. Generally, unstructured data are big and have different structures and patterns, including unstructured documents, images, and voices, and SNS news articles. The techniques of analyzing unstructured data and finding their patterns include text analysis and non-standard text analysis [15], [16].

In the big-data environment, almost over 80% are unstructured data. Therefore, data mining of big-data focuses on unstructured data. Of unstructured data mining methods, the typical text mining method applies natural language processing of texts to extract information from large documents, find correlations, and make classification and clustering to find the hidden meaning of data [17-19].

2.4. Related works

Han (2014) defined the security threats arising the mobile cloud environment and proposed a security measures in mobile apps and a corporate measures [4].Park et al. (2016) focused on the establishment of cloud storage system using Korean a file system that provides cloud storage service in the wireless internet environment 20.

Jung &Bae (2012) looked into the trend of security threat factors and security guidelines in cloud service and analyzed the issues of cloud service security and the cases of security accidents 21. Li & Li (2015) proposed a plan for optimal scheduling to implement hybrid cloud. By applying the market based hybrid cloud optimal scheduling algorithm, he proposed a plan with better performance than an existing one22.

Park et al. (2016) researched the influential factors of security threats on the continuous use intention of cloud service and conducted a comparative analysis with the use of Korean and Chinese users [1]. According to the analysis, the security threats recognized were different between Korean and Chinese users.

Kim et al. (2015) analyzed the effects of the security risk factors suitable to cloud computing paradigm on firms' acceptance of cloud service in order to find the factors to promote firms' cloud service introduction, classified security risks into compliance risk, information leakage risk, failure recovery risk, and service interruption risk, and analyzed them with TAM [11].

3. Research models and hypotheses

This study collected data through crawling to find the recognition of cloud service in Korea and analyzed the importance of cloud service security. In addition, it conducted an analysis and implemented visualization with the use of R, an open source tool for analyzing the data of cloud service security [17].

3.1. Problem definition

To expand cloud service, it is important to come up with measures for security and security threats. Therefore, this study tries to find the trend of cloud computing service security.

3.2. Necessary information

To do that, this study collected and crawled news articles of cloud security service from NAVER, a Korean representative portal site (news.naver.com).

3.3. Data needed to derive information

To draw the information necessary for problem finding, it is necessary to obtain the data for analysis. The first analysis data are the NAVER news articles collected. The raw data crawled from NAVER are the data for the systematical analysis in this study.

3.4Analysis technique to derive information

As shown in <Table 2>, to collect the news articles published from Jan. 1, 2015 to Jun. 30, 2016, Jsoup was used. TAXEDO and R were used as visualization tools. The collected data were cleaned for analysis. The data for final analysis were presented with the use of Excel, R, and TAXEDO.

Table 2: Data Collection of News

Category	Description	
Channel	NAVER News	
Conditions	Cloud Computing, Cloud Service, Cloud Security	
	Tenure	No. of Articles
Total News Articles Found	2015.01.01. ~ 2015.06.30	6269
	2015.07.01. ~ 2015.12.31	6827
	2016.01.01. ~ 2016.06.30	6634
Period	2015-01-01 00:00:00 ~ 2016-06-30 23:59:59	

4. Results

4.1 Data collection and pre processing

To collect the data for analysis, this researcher set a period and collected the news articles of cloud service security at news.naver.com. The collected data were cleaned and then converted into the data that can be analyzed. The data analysis process has four steps as illustrated in <Fig. 1>20.

Fig. 1: Data Analysis Procedures.

In the first step, issues are defined. In the second step, data are collected with the use of JSOUP. In the third step, data are cleaned. In the fourth step, the drawn results are visualized.

4.1.1. Collection of data

To collect the news articles data, this study used Eclipse. The overall survey data were the news articles published from Jan. 1, 2015 to Jun. 30, 2016. The keywords used for collection were cloud, cloud service, and cloud security. The news articles including the chosen words were extracted.
A total of news articles were collected. As mentioned earlier, only the NAVER news articles were used for this analysis. After the removal of duplicate news articles, a total of 19,730 cleaned NAVER news articles were selected and used for DB.

4.1.2. Extraction of data

Of the collected news articles, duplicate news ones were removed. With the use of R, the frequency of words was calculated. In R, the Korean natural language processing package “KoNLP” and “Sejong” dictionary were used.

4.2. Data analysis

With the use of R, data were cleaned, and noun words were extracted. Based on the extracted noun words, visualization was performed. The note of shows the collected news articles and displays the process of drawing the frequency of keywords in the R package analysis [20].
Up to 30 words were extracted. Regardless of a period, the word 'cloud' was ranked the 1st, the word 'service' in the 2nd, the word 'security the 3rd, and the word 'enterprise' the 4th.
The reason for the result is that the words 'cloud computing service security' were entered for crawling. Irrelevant postposition words were drawn as well. Therefore, the words irrelevant to cloud computing service were removed. As a result, 20 noun words considered important were used.

4.3. Data analysis results

The analysis procedure of <Fig. 1> was performed. The analysis results are presented in <Table 3>. The larger and bolder the words, the higher the frequency of the noun words. The data analysis results are displayed in <Fig. 2>. The noun and proposition words irrelevant to the theme of this study were removed. The results are presented in <Table 4>.

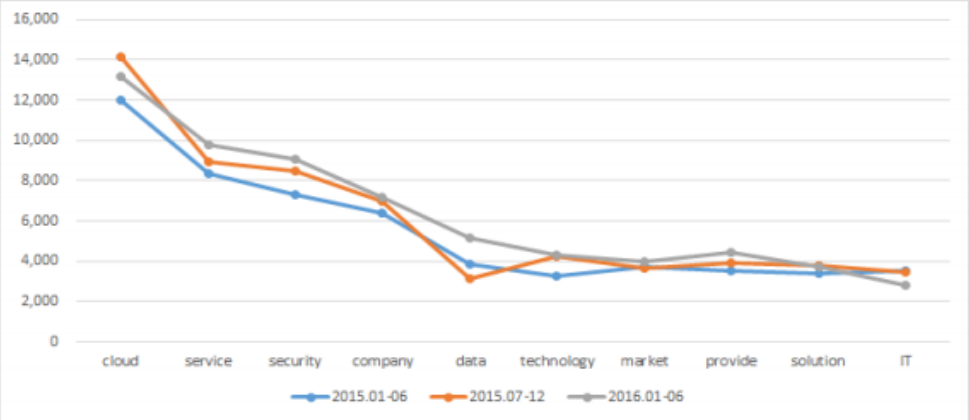

Fig. 2: Data Frequency (2015~2016).

Table 3: Data Cleaning Phase

	2015.01~06		2015.07~12		2016.01~06	
	rev	Freq	rev	Freq	rev	Freq
1	cloud	11969	cloud	14121	cloud	13190
2	service	8366	service	8924	service	9753
3	security	7336	security	8481	security	9044
4	company	6413	company	7006	company	7169
5	data	3870	technology	4239	data	5149
6	market	3751	provide	3893	provide	4445
7	IT	3527	solution	3756	technology	4326
8	provide	3524	market	3640	market	3995
9	solution	3412	IT	3483	domestic	3730
10	technology	3242	data	3139	solution	3721
11	business	3031	domestic	3100	business	3309
12	domestic	2839	business	3061	base	3049
13	base	2483	base	2830	IT	2777
14	system	2424	information	2544	development	2766
15	information	2317	utilize	2527	system	2625
16	construct	2270	support	2497	customer	2609
17	development	2267	construct	2482	do	2458
18	environment	2253	development	2425	support	2419
19	support	2204	do	2387	utilize	2396
20	use	2166	field	2346	management	2370
21	body	2154	industry	2328	field	2349
22	industry	2087	management	2317	information	2310
23	management	2074	system	2289	global	2261
24	relation	2050	body	2185	use	2259
25	do	2004	attack	2125	construct	2255
26	join	1988	network	2079	center	2254
27	news	1903	product	2063	enterprise	2181
28	platform	1889	use	2052	environment	2180
29	utilize	1816	introduction	2042	product	2175
30	product	1810	threat	2021	platform	2143

Table 4: Data Cleaning after Implementation

	2015.01~06		2015.07~12		2016.01~06	
	rev	Freq	rev	Freq	rev	Freq
1	cloud	11969	cloud	14121	cloud	13190
2	service	8366	service	8924	service	9753
3	security	7336	security	8481	security	9044
4	company	6413	company	7006	company	7169
5	data	3870	technology	4239	data	5149
6	market	3751	provide	3893	provide	4445
7	IT	3527	solution	3756	technology	4326
8	provide	3524	market	3640	market	3995
9	solution	3412	IT	3483	solution	3721
10	technology	3242	data	3139	business	3309
11	business	3031	business	3061	base	3049
12	base	2483	base	2830	IT	2777
13	system	2424	information	2544	development	2766
14	information	2317	support	2497	system	2625
15	construct	2270	construct	2482	customer	2609
16	development	2267	development	2425	support	2419
17	environment	2253	field	2346	utilize	2396
18	support	2204	industry	2328	management	2370
19	industry	2087	management	2317	field	2349
20	platform	1889	system	2289	information	2310

In the period of Jan. to Jun. 2015, the importance order was cloud, service, security, enterprise, and data.
Generally, data (documents, pictures, music, etc.) are saved in cloud so that data was found to have the fifth most importance.
The words 'cloud solution' and 'cloud security solution' were found much. So, the solution was highly important. Since the system importance and platform became topics, the platform relevant to cloud has been emerged (Fig. 3).

Fig. 3: Data Visualization (2015 First Half).

In the period of Jul. to Dec. 2015, the importance order was cloud, service, security, enterprise and data. The cloud solution was more important in this period than in the first half of 2015 (Fig. 4). The reason was that cloud solutions (including security solutions) were released more in the 2nd half than in the first half.

Fig. 4: Data Visualization (2015 Second Half).

The ranking position of data was low. Its importance seems to be low because it is general to save most data in cloud. The difference between the 1st half of 2015 and the 2nd half is management. That is because the management of cloud system and system security has been more important as described early in the introduction.
In the period of Jan. to Jun. 2016, the importance order was cloud, service, security, enterprise, and data.
The difference between the 1st half of 2016, and the 1st and 2nd half of 2015 is customer and use (Fig. 5).

Fig. 5: Data Visualization (2016 First Half).

Although the Korean government established a relevant act for promoting cloud, the cloud service has yet to be promoted domestically. NAVER and DAUM, the representative portal sites in Korea, provided cloud service.
However, when they stopped the service in Dec. 2015, many customers experienced a lot of inconvenience. For the reason, customer was drawn as a critical factor. The factor 'use' was emerged in the period.
Generally, cloud service was simply used for data backup (pictures, music, documents, etc.). To promote the cloud market, it is necessary to design diverse plans for using cloud. As a result, it was drawn as a critical factor.
The most important thing to promote cloud service is security which should be taken into account both in the perspective of service providers and in the perspective of service users.

4.4. Discussion

These days, the cloud service tends to be expanded domestically. It requires login at web or on mobile. For the reason, it is greatly exposed to security threats. Unlike overseas cases of cloud service attacks, no such case was found domestically. However, the websites of banks, public organization, and open markets were attacked and their information was leaked frequently. In the circumstance, the cloud security is the most important. Therefore, it is necessary to apply security technology to a cloud solution in its development stage to prevent attacks. In particular, it is significant to apply the solutions relevant to target attacks and ransomware attacks.
To promote the cloud market, it is important to boost up the advantages of cloud service and to provide solutions to security vulnerabilities and plans for improved functional use.

5. Conclusion

Precedent studies have mostly focused on the security threat technology trend analysis and were mostly at the level of just introducing technology trends. Unlike the precedent studies, however, this study conducted an analysis by using crawling, text mining, and network analysis technique, which were big data analysis techniques, in order to identify the cloud service security trends.
To find the trend of cloud computing service security, this study collected news articles and analyzed them after data cleaning. The analysis revealed that there was a difference in importance of each period.
In the first half of 2015, the 2nd half of 2015, and the 1st half of 2016, the common noun words extracted were cloud, service, security, and enterprise in order.
The difference was that the word 'environment' was found in the 1st half of 2015, the word 'management' in the 2nd half of 2015, and the words 'customer' and 'use' in the 1st half of 2016.
This study is meaningful in the point that it drew keywords related to clouding computing with the use of text mining technique. Previous studies were based on objective data (numbers), whereas this study analyzed unstructured data of social media in given periods.
Nevertheless, there are limitations. First, this study collected news articles and extracted only key noun words relating to cloud service security from them. Secondly, it failed to establish a dictionary based on morphological analysis, because it extracted only key noun words.
There are no keyword network analysis and additional analysis on the extraction of compound noun words. In the future, it will be necessary to make network analysis, morphological analysis, and other additional analysis using various analysis tools.
Further studies need to investigate the method of objectifying various kinds of data such as the existing thesis, patent, and report.

References

[1] Park ST, Park EM, Seo JH, and Li G, Factors affecting the continuous use of cloud service: focused on security risks. Cluster Computing.2016, 19(1), pp.485-495.

[2] Noh KS, A study on the position of CDO for improving competitiveness based big data in cluster computing environment. Cluster Computing.2016, 19(3), pp.1659-1669.

[3] Korea Ministry of Government Legislation, Cloud Development Act, http://www.law.go.kr.

[4] Han, J.S.: Security Threats in the Mobile Cloud Service Environment. Journal of Digital Convergence.2014, 12(5), pp.263-269.

[5] WWDC, iCloud Storage Overview, https://developer.apple.com/videos/play/wwdc2011/501.

[6] GOV.UK. G-Cloud suppliers' guide, https://www.gov.uk/guidance/g-cloud-suppliers-guide.

[7] Park ST, Kim YR, Jeong SP, Hong CI, Kang TG, A Case Study on Effective Technique of Distributed Data Storage for Big Data Processing in the Wireless Internet Environment. Wireless Personal Communications.2016, 86(1), pp.239-253.

[8] Lee BS, Kim BS, Protection of Personal Information on Cloud Service Models. Journal of the Korea Institute of Information Security and Cryptology.2015, 25(5), pp.1245-1255.

[9] NIA. ICT new technology paradigm: Cloud computing strategy. CIO Report, 2009, 17, 1–40.

[10] O'Neal J, NetApp, storage infrastructure for the cloud, 2009. http://www.netapp.com/us/communities/tech-ontap/tot-cloud-storage-0509.aspx.

[11] Alqahtany S, Clarke N, Furnell S, Reich C, A forensic acquisition and analysis system for IaaS. Cluster Computing.2016, 19(1), pp.439-453.

[12] Carretero J, Blas JG, Introduction to cloud computing: platforms and solutions. Cluster Computing.2014, 17(4),pp.1225-1229.

[13] Kim DY, Li G, Park ST, Ko MH, A study on effects of security risks on acceptance of enterprise cloud service: moderating of employment and non-employment using PLS multiple group analysis. Journal of Computer Virology and Hacking Techniques.2016, 12(3), pp.151-161.

[14] Ju JH, Ma SY, Moon JS, Proposal of Security Requirements for Storage Virtualization System against Cloud Computing Security Threats. Journal of Security Engineering.2014, 11(6), pp.469-478.

[15] Noh KS, A Study on the Authentication and Security of Financial Settlement using the Finger Vein Technology in Wireless Internet Environment. Wireless Personal Communications.2016, 89(3), pp.761-775.

[16] CSA, Cloud Security Alliance. https://downloads.cloudsecurityalliance.org/assets/research/security-guidance/csaguide.v3.0.pdf.

[17] The Society of Digital Policy & Management. Big Data Analytics for Business. Kwangmoonkag, Seoul, 2015.

[18] The Society of Digital Policy & Management. Introduction to Big Data Analysis Kwangmoonkag, Seoul, 2016.

[19] Park EM, Seo JH, Ko MH, The effects of leadership by types of soccer instruction on big data analysis, Cluster Computing.2016, 19(3), pp.1647–1658.

[20] The Society of Digital Policy & Management. Big data analysis plan. Wowpass, Seoul, 2015.

[21] Jung SJ, Bae YM, Trend analysis of Threats and Technologies for Cloud Security. Journal of Security Engineering.2013, 10(2), pp.199-212.

[22] Li C, Li L, Efficient Market Strategy Based Optimal Scheduling in Hybrid Cloud Environments. Wireless Personal Communications.2015, 83(1), pp.581-602.