Evaluation of Selecting Cloud Services Approach for Data Storage using Secret Sharing Scheme

Shohei Ueno
Hosei University
Tokyo, Japan
shohei.ueno.4h@stu.hosei.ac.jp

Atsushi Kanai
Hosei University
Tokyo, Japan
yoikana@hosei.ac.jp

Shigeaki Tanimoto
Chiba Institute of Technology
Chiba, Japan
shigeaki.tanimoto@it-chiba.ac.jp

Hiroyuki Sato
The University of Tokyo
Tokyo, Japan
schuko@satolab.itc.u-tokyo.ac.jp

Abstract—**Cloud services have become more popular because of their decreasing cost. However, it is difficult to select the optimal cloud service because there are many services whose service levels are different. We evaluate our proposed method for dynamically selecting the optimal cloud services to store data in a heterogeneous multi-cloud environment. The evaluation used the SLAs of actual cloud services and the results indicate it is possible to select a combination of cloud services.**

Keywords-cloud computing; multi-cloud; hybrid cloud; secret sharing scheme; availability; confidentiality

I. INTRODUCTION

Cloud computing has recently become popular. Methods involving a combination of multiple cloud services have been proposed, [2]-[4], which provide users with more advantages (availability or confidentiality) than usage of single clouds.

These methods need to select the best combinations of cloud services. As there are many different types of cloud services with various service levels, a wide variety of service levels can be constructed in heterogeneous multi-cloud environments. Especially, multiple services are used at the same time.

We first describe the proposed method. Then, we are quantifying the evaluation using the developed prototype. Furthermore, we present a concrete case using actual cloud services.

The rest of this paper is organized as follows. We present the related work in Section 2. In Section 3, we describe the assumed environment and the proposed method. In Section 4, we show the overview of the evaluation system. In Section 5, we describe the evaluation using actual public cloud services implementing the prototype of this proposed method, and evaluate the communication speed. Finally, we conclude the paper in Section 6.

II. RELATED WORK

Approaches which use multiple cloud services have been proposed to improve availability and confidentiality, cost, performance, etc., when compared with single cloud services. For example, DepSky [3] improved the availability, integrity, and confidentiality of data stored in clouds. The high-availability and integrity layer (HAIL) [6], which accepts a set of servers to prove to clients that stored files are complete and recoverable, was developed on links between multiple cloud services.

Files that users want to manage in cloud storage have properties of various degrees of confidentiality and availability. Therefore, it is necessary to change the requirements per file. This means one has to reselect the best combination of cloud per file. Cardellini et al. demonstrated how to select the best services [10] in relation to the cost-effective use of such services. Tsai et al. proposed a cost-effective intelligent configuration model [11]. In addition, a file-distribution method using a secret sharing scheme was proposed and evaluated in a homogeneous multi-cloud environment [12]. A data management method in this environment was also proposed [13].

There are also security concerns about public clouds. Cloud security in terms of data management has also been discussed [15]-[18]. To solve one of these issues, a method in which a system automatically selects appropriate cloud services using a service-level agreement (SLA) written in extensible markup language (XML) has been proposed [19].

Currently, there is no way to select and evaluate optimal cloud services from many different clouds (heterogeneous multi-clouds) in using multiple clouds at the same time in the proposed environment [12][13].

ISBN: 978-1-61208-522-7

III. Proposed Method

A. Assumed Environment

We assumed a multi-cloud environment with many cloud-storage services in a secret sharing scheme, and all of these services had machine readable SLAs written in XML [12][13][22]. In this section, we introduce the proposed method [22] which is evaluated.

Figures 1 and 2 outline the proposed method. A user selects a set of cloud services using their SLAs depending on the required availability, confidentiality, and cost. Then, all combinations of cloud services are calculated by the user requirements, and it is determined to store in cloud services. When a file is stored in cloud services, it is distributed using a (k, L, n) secret sharing scheme [19]. However, the user requirement is different per file. The best combination of cloud services is selected by calculating when a user stores the secret information.

B. (k, L, n) secret sharing scheme

The (k, L, n) secret sharing scheme was devised by Yamamoto [20] and is an extension of the (k, n) secret sharing scheme presented by Shamir [21]. It can reduce the amount of distributed information compared to the (k, n) secret sharing scheme.

By applying the (k, L, n) secret sharing scheme to secret information x, n pieces of distribution information are obtained. The restoration of the information is performed by collecting k pieces. Additionally, the data size of the distributed information becomes 1/L times that of the secret information. It is possible to identify part of the secret information from many k-Ls that are less than the ks of distributed information. Fewer k-Ls provide safety with regard to information theory, so it is not possible to obtain any secret information.

C. Matching user requests with cloud service levels

In the proposed method [22], user requirements are defined using four indicators.

1) Cost: Required cost per amount data (to store 1 MB [yen/MB])
2) Confidentiality: Risk of secret data being identified from data stored to cloud services
3) Availability: Total operating rate [%] of multi-cloud
4) Transfer time: Upload time and download time [s/MB]

We assumed these indicators are written in SLA of cloud services. Therefore, we calculate and select the best combinations of cloud services using the user requests.

D. Formulas that correspond to user requests

In the proposed method, the best combination of cloud services is selected by calculating [22].

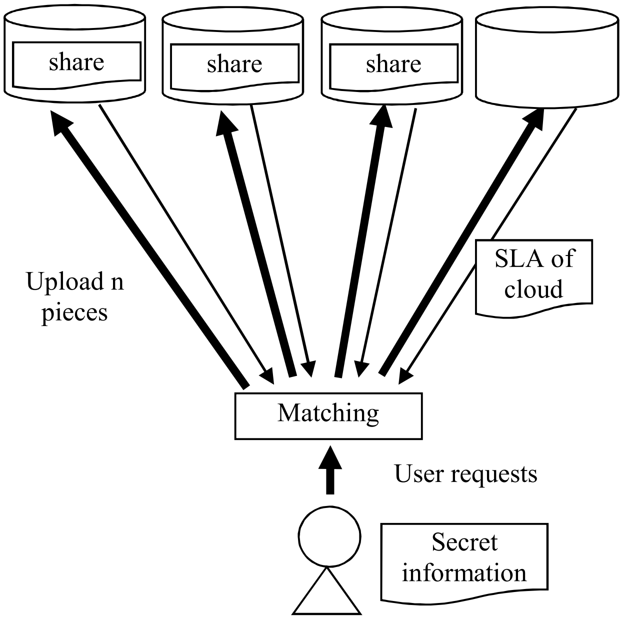

Figure 1. The image of uploading

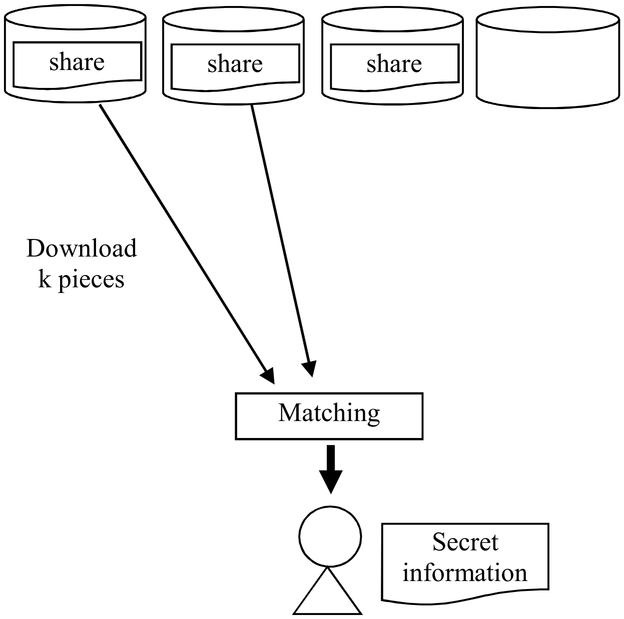

Figure 2. The image of downloading

a) Uploading

As it is necessary for users and cloud services to communicate during uploads, availability, transfer time, and costs are important as metrics.

a. Cost

Cost is the total expense of all cloud services and is expressed as

$$\mathrm{Cost} = \frac{1}{L}\sum_{i\in n} \mathrm{Cost\ of\ Cloud_i}. \qquad (1)$$

b. Availability

Because it must be able to communicate with all cloud services to store shared information, availability becomes:

$$\mathrm{Availability} = \prod_{i\in n} \mathrm{Operating\ rate\ of\ Cloud_i}. \qquad (2)$$

c. Transfer time

Transfer time is the total upload time to reach each service, and it becomes:

$$\mathrm{Transfer\ time} = \frac{1}{L}\sum_{i\in n}\frac{1}{Communication\ speed\ of\ \mathrm{Cloud_i}}. \qquad (3)$$

b) Storing

As it is not necessary to communicate with cloud services, confidentiality of the secret data is very important.

a. Confidentiality

Confidentiality is related to the probability of information leakage from each cloud and the total disclosure level of the information. Here, 0≦x≦n.

$$\mathrm{Confidentiality} = \sum_{x=k}^{n}\left(\sum_{i\in\{y|y\in P(n),\,|y|=x\}}\prod LP(i)\prod_{j\in n-y}\{1-LP(j)\}\right) * Specific\ Level(x). \qquad (4)$$

where P(n) is the power set of n, and LP(i) is the leakage probability of cloud i.

The disclosure level is represented by the following formula depending on the parameters of the (k, L, n) secret sharing scheme.

$$Specific\ Level(x) = \begin{cases} 0\ (x \le k-L) \\ 1-\frac{k-x}{L}\ (k-L < x < k). \\ 1\ (k \le x) \end{cases} \qquad (5)$$

c) Downloading

As it is necessary to communicate with clouds, the availability and transfer time is important.

a. Availability

The availability in a cloud service to upload distribution information is the probability that users can communicate with all the cloud services necessary to restore the shared data in all services that have stored shared data.

The A(i) in this equation is the operation ratio of cloud i.

$$\mathrm{Availability} = \sum_{x=k}^{n}\left(\sum_{i\in\{y|y\in P(n),\,|y|=x\}}\prod A(i)\prod_{j\in n-y}\{1-A(j)\}\right). \qquad (6)$$

b. Transfer time

Transfer time is the time to communicate with the cloud and restore information.

$$\mathrm{Tranfer\ time} = \frac{k}{Ln}\sum_{i\in n}\frac{1}{\mathrm{Communication\ Speed\ of\ Cloud_i}}. \qquad (7)$$

E. Relationship between the indicators and (k, L, n) secret sharing scheme

In the proposed method [22], the combination of cloud services is calculated by the user requirements, and secret information is uploaded for the selected cloud services using (k, L, n) secret sharing scheme. Table I summarizes the relationships between the indicators and the actions.

The availability in uploading is worse when the value of n is increasing. The total operating rate is worse because of increasing the number of distributions. The cost in uploading is better when the value of L is increasing, but it is worse when the value of n is increasing. The smaller size of data can be stored in cloud services inexpensively, but the total cost is increasing because of increasing the number of distributions. The transfer time in uploading is better when the value of L is increasing, but it is worse when the value of n is increasing. The smaller size of data can be stored in cloud services quickly, but the total transfer time is increasing because of increasing the number of distributions.

The confidentiality in storing is better when the value of k is increasing, but it is worse when the value of L and n are increasing because of equation (5).

The availability in downloading is better when the value of n is increasing, but it is worse when the value of k is increasing. It is necessary to collect k pieces of distribution

TABLE I. RELATIONSHIP BETWEEN PARAMETERS AND ACTIONS

	Uploading			Storing	Downloading	
	Availa-bility	*Cost*	*Transf-er time*	*Confiden-tiality*	*Availa-bility*	*Transf-er time*
k	-	-	-	Better	Worse	Worse
L	-	Better	Better	Worse	-	Better
n	Worse	Worse	Worse	Worse	Better	-

information for restoring the secret information. The transfer time in downloading is better when the value of L is increasing, but it is worse when the value of k is increasing. The smaller size of data can be downloaded quickly, but the total transfer time is increasing because of increasing the number of distributions for restoring.

IV. Overview of the Evaluation System

In this section, we explain the evaluation of a method proposed in a previous study [22] using some of the metrics in a heterogeneous cloud environment. In the previous study, we assumed the value of SLA for private and public cloud services, and selected some combinations using the proposed method.

For the result, some combinations were calculated for some situations; highest availability, lowest cost or highest confidentiality.

In the current study, we investigated some actual SLAs of public cloud services. Specifically, we investigated the SLAs of Google Drive, CloudN, KDDI, BOX, Dropbox, and One Drive. Table II lists the SLA metrics for these cloud services. However, these name of cloud services were expressed from P0 to P5 in Table II for consideration to the cloud services. In Table II, all cloud services did not provide leakage probability and communication speed. Therefore, we assumed the value of leakage probability based on the description of confidentiality. We decided whether the acquisition of security standards and the policy of security are written in SLA of each cloud services or not.

For communication speed, we did not evaluate the communication speed because we cannot estimate the value. However, it is necessary to evaluate the communication speed. Then we developed a prototype in this proposed method and evaluated the communication speed between the cloud services and user. Here, we use five cloud services: Box, Dropbox, Google Drive, and One Drive. For the implementation, we use these cloud services, which provide API. The results will be described later.

Then, Figure 3 shows the image of the evaluation model. We selected the combination of cloud services, and determined the parameter of a (k, L, n) secret sharing scheme using the proposed equation. In addition, we developed a prototype for uploading and downloading the distributed data using that parameter.

TABLE II. DESCRIPTION OF ACTUAL CLOUD SERVICES

Cloud	***Operating rate***	***Cost***	***Leakage probability***	***Communication speed***
P0	Written	Written	Not Written	Not Written
P1	Written	Written	Not Written	Not Written
P2	Written	Written	Not Written	Not Written
P3	Written	Written	Not Written	Not Written
P4	Written	Written	Not Written	Not Written
P5	Not Written	Written	Not Written	Not Written

V. Actual Cloud Evaluation

A. Actual SLAs description and setting

Table III lists the metrics of private and public cloud services that satisfy the actual SLAs. However, we assumed the same value as that of the private clouds in Table III because the actual value of private cloud is not written. Additionally, P0 is getting the ISO 27001[23] and written the policy of security in SLA, we assumed it is the better value of leakage probability than other public cloud services. On the other hand, P3 and P5 are written nothing about security. Then we assumed these are the worse value of leakage probability than other public cloud services.

Here, cost is defined as [yen/(month・GB)], and the user has already contracted for all the public cloud services.

$$\mathrm{Cost} = \textstyle\sum_{i \in n} \mathrm{Cost\ of\ Cloud_i}. \qquad (8)$$

Therefore, the costs of all combinations are fixed.

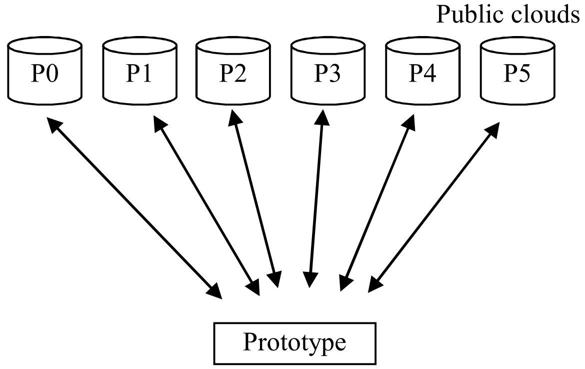

TABLE III. PARMETER SETTING FOR EVALUATION

Cloud	***Operating rate***	***Cost***	***Leakage probability***
Private Cloud	0.999	-	0.001
P0	0.999	16.6	0.01
P1	0.9999	8.6	0.1
P2	0.9999	30	0.1
P3	0.999	6.0	0.5
P4	0.999	0.54	0.1
P5	0.9999	0	0.5

All cloud services have sufficient communication speed; therefore, we did not evaluate transfer time. The L of all combinations was only one. Then, we calculated all combinations of cloud services, and Table IV lists the three unique combinations. As a result, the parameter of (k, L, n) secret sharing scheme is (k = 2, L = 1, n = 4) or (k = 4, L = 1, n = 4)

Combinations (p0, p1, p2, p4 in k = 2) and (p1, p2, p3, p4) have the best availability; Combination (p0, p1, p2, p4 in k = 2) has a lower cost than Combination (p1, p2, p3, p4). Combination (p0, p1, p2, p4 in k = 4) has the highest confidentiality, which is better than that of private clouds. However, the availability of Combination (p0, p1, p2, p4 in k=4) is the worst. This is caused by parameter k, that made availability in downloading worse.

B. *Evaluation of communication speed*

Table V lists all the combinations of these cloud services. Here, the value of n in (k, L, n) secret sharing scheme is four. Combinations (p0, p3, p5), (p3, p4, p5), and (p0, p4, p5) have better availability than Combination (p0, p3, p4).

Then, we measured the communication speed for each cloud services (Table VI) and all the combinations (Table VII). Here, the data size is 10 [MB], upload time is the average of 10 measurements, and download time is the average of 3 measurements. In addition, the download time is measured for all combinations.

In Table VII, Combination (p0, p3, p4) has the best upload time, and the combination of cloud p0 and p4 has the best download time. Therefore, depending on the combination of clouds chosen, it is possible to have a better communication speed than using only one cloud service. However, all of the combinations are worse upload time than only each cloud, and cloud p3 is also worse download time. Thus, the communication speed of some combinations is worse than using each cloud service.

TABLE IV. THREE BEST COMBINATIONS EXTRACTED FROM RESULTS

Combination	*k*	*L*	*n*	*Leakage probability*	*Availability when downloading*
p0,p1,p2,p4	2	1	4	304.3	0.99999999978
p1,p2,p3,p4,	2	1	4	1495	0.99999999978
p0,p1,p2,p4	4	1	4	0.1	0.99780141

TABLE V. ESTIMATED VALUE

Combination	*k*	*L*	*n*	*Leakage probability*	*Availability in downloading*
p0,p3,p5	2	1	3	0.255	0.9999988
p3,p4,p5	2	1	3	0.3	0.9999988
p0,p3,p4	2	1	3	0.055	0.999997002
p0,p4,p5	2	1	3	0.055	0.9999988

TABLE VI. COMMUNICATION SPEED BETWEEN CLOUD SERVICES AND USERS FOR SINGLE SERVICE

Cloud	*Upload time [ms]*	*Download time [ms]*
P0	4118	6858
P3	3680	2452
P4	4744	4642
P5	8815	7451

TABLE VII. COMMUNICATION SPEED FOR ALL COMBINATIONS

Combi-nation	*Upload time [ms]*	*Download time [ms]*	*Download Clouds*
p0,p3,p5	10070	7499	p0,p5
		12856	p3,p5
		8986	p0,p3
p3,p4,p5	10829	10182	p4,p5
		8652	p3,p4
		12094	p3,p5
p0,p3,p4	6820	5445	p0,p4
		11487	p3,p4
		8594	p0,p3
p0,p4,p5	10616	8372	p4,p5
		4390	p0,p4
		8601	p0,p5

Additionally, Combination (p0, p3, p4) has the worst availability in Table V. Combination (p0, p4, p5) does not have a good upload time but has a good average download time compared to other combinations. We need the communication speed of the SLA to evaluate cloud services not only operating rate, and find the best cloud services.

However, this evaluation is one example. In the actual situation, the best combinations can be selected by the calculation taking into consideration the user requirements in this proposed method.

VI. CONCLUSION

We evaluated a method using multiple cloud storage services in a heterogeneous cloud environment by using concrete values of three metrics. We found that some combinations of cloud services were more useful compared to only one private cloud service. All combinations had both advantages and disadvantages. Therefore, we found that the communication speed is necessary for the new evaluation value. However, we only implemented the prototype. In the future, we need to implement the actual system.

ACKNOWLEDGEMENT

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 15H02783.

REFERENCES

[1] NIST [Online] Avalilable from: http://www.nist.gov/itl/cloud/ 2016.01.11

[2] M. Vukolic, "The Byzantine empire in the intercloud," ACM SIGACT News, 41, 2010, pp. 105–111.

[3] A. Bessani, M. Correia, B. Quaresma, F. André, and P. Sousa, "DEPSKY: dependable and secure storage in a cloud-of-clouds," EuroSys'11: Proc. of 6th Conf. on Computer Systems, 2011, pp. 31–46.

[4] H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon, "RACS: A case for cloud storage diversity," SoCC'10: Proc. of 1st ACM Symposium on Cloud Computing,2010, pp. 229–240.

[5] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A high-availability and integrity layer for cloud storage," CCS'09: Proc. of 16th ACM Conf. on Computer and Communications Security, 2009, pp. 187–198.

[6] H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon, "RACS: A case for cloud storage diversity," SoCC'10: Proc. of 1st ACM Symposium on Cloud Computing, 2010, pp. 229–240.

[7] NRI Secure Technologies. [Online]. Available from: http://www.nri-secure.co.jp/service/global/gss.html 2016.01.11

[8] T. Matsumoto, T. Seito, A. Kamoshita, T. Shingai, and A. Sato, “High-Speed Secret Sharing System for Secure Data Storage Service,” SCIS2012. The 29th Symposium on Cryptography and Information Security, 2012.

[9] V. Cardellini, V. Valerio, V. Grassi, S. Iannucci, and F. Presti, “A New Approach to QoS Driven Service Selection in Service Oriented Architectures,” SOSE ,2011, pp. 102–113.

[10] W. Tsai, G. Qi, and Y. Chen, “A Cost-Effective Intelligent Configuration Model in Cloud Computing,”ICDCSW,2012,pp.400-408

[11] Y. Kajiura, A. Kanai, S. Tanimoto, and H. Sato, “A File-distribution Approach to Achieve High Availability and Confidentiality for Data Storage on Multi-cloud,” SAPSE2013, 2013.

[12] A. Kanai, S. Tanimoto, and H. Sato,” Data Management Approach for Multiple Clouds Using Secret Sharing Scheme,” 8th International Workshop on Advanced Distributed and Parallel Network Applications (ADPNA-2014), 2014,pp. 432–437.

[13] Cloud Security Alliance, “Cloud Control Matrix Version 3.0”, 2013.

[14] H. Sato, A. Kanai, and S. Tanimoto, "A Cloud Trust Model in a Security Aware Cloud," Proc. of IEEE/IPSJ International Symposium on Applications and the Internet (SAINT2010), 2010,pp. 121–124.

[15] S. Tanimoto, S. Matsui, H Sato, A Kanai, and et al,” A Study of Risk Management in Hybrid Cloud Configuration,” Computer Information Science, Springer, vol. 493, 2013,pp. 247–257.

[16] S. Tanimoto, H. Sato, A. Kanai, and et al,” A Study of Data Management in Hybrid Cloud Configuration,” 14th IEEE/ACIS, SNPD2013, 2013,pp. 381–386.

[17] H. Sato, A. Kanai, and S. Tanimoto, “Building a Security Aware Cloud by Extending Internal Control to Cloud,” Proc. of 10th Int'l Symposium on Autonomous Decentralized Systems (ISADS 2011), 2011,pp. 323–326..

[18] H. Sato, S. Tanimoto, and A. Kanai, “A Policy Consumption Architecture that Enables Dynamic and Fine Policy Management,” Proc. of 3rd ASE International Conf. on Cybersecurity,2014, pp. 1–11.

[19] H. Yamamoto, “Secret Sharing System Using (k, L, n) Threshold Scheme,” Electron. Commun. Jpn. (Part I: Commun.), 1986,vol. 69, no. 9, pp. 46–54.

[20] A. Shamir, "How to share a secret," Communications of the ACM, 22(11), 1979,pp. 612–613.

[22] Y. Kajiura, S. Ueno, A. Kanai, S.Tanimoto, and H. Sato,” Approach to Selecting Cloud Services for Data Storage in Heterogeneous Multi-cloud Environment with High Availability and Confidentiality,” The First International Workshop on Service Assurance in System Wide Information Management (SASWIM2015), 2015, pp. 205-210.

[23] ISO/IEC 27001:2005, “Information technology – Secutiry techniques –Information security management systems – Requirements,“ 2013