法政大学大学院理工学・工学研究科紀要　Vol.56(2015 年 3 月）　法政大学

マルチクラウドにおける最適クラウド決定方式

AN APPROACH TO SELECTING OPTIMAL CLOUD SERVICES FOR DATA STORAGE IN MULTI-CLOUD ENVIRONMENT

梶浦悠生
Yuki KAJIURA
指導教員　金井敦

法政大学大学院理工学研究科応用情報工学専攻修士課程

Cloud services are becoming more popular and as their price has been decreasing, the opportunity to use them has been increasing. However, it has become difficult to select optimal cloud services from many services. This paper proposes an approach to dynamically selecting optimal cloud services to store data in heterogeneous-multi-cloud environments, which we evaluated by using a secret sharing scheme. The results indicated that it is possible to select the best services from the proposed model in homogeneous-multi-cloud environments and heterogeneous-multi-cloud environments.

***Key Words** : Clouds computing, Security, Multi-cloud,*

１．はじめに

近年，クラウドコンピューティングが発展し，様々な場面で利用されるようになっている．クラウドの提供形態にはプライベートクラウドやパブリッククラウドなどがあり，それぞれに特徴がある[1]．これらのサービスを一つのみ利用するより，複数のサービスを組み合わせて利用することでセキュリティやコストなどの面で利点がある．この複数サービスを併用する手法には，ハイブリッドクラウドやマルチクラウド，インタークラウドなどがある[3]-[4]-[5]．ハイブリッドクラウドは，プライベートクラウドやパブリッククラウドの利点を活かしながら，欠点を補い合うものである．マルチクラウドやインタークラウドと呼ばれる手法は，多数のクラウドサービスを組み合わせた状態でユーザに提供するものである．これらの手法は，組み合わせ方を変えることによって一つのクラウドサービスのみを用いた場合よりもスケーラブルにセキュリティやコストを最適化することが可能になる．特に，クラウドサービスを提供する組織が増えた現在においては，その提供されるサービスレベルも様々であり，より多種多様な環境が構築できるようになった．

しかし多数ある選択肢の中からユーザが自らの要求に見合うサービスを選び出すことは煩雑である．特に，技術に疎いユーザの場合にはサービスレベルの精査自体がそもそも困難であると言える．更に，マルチクラウド環境の場合には組み合わせ方によって全体として提供されるサービスレベルが変化するため，特に複雑である．そのため，自らの要求に見合う最適なサービスを選定することはより煩雑なものとなり，技術に精通したユーザであっても困難であると考えられる．また，一人のユーザがクラウドサービスに求めるものは，ファイルやプロジェクトごとなどのより小さな単位で異なることがある．異なる要求をすべて一緒くたに扱ってしまうことは，セキュリティ上の問題が発生したり，余計なコストがかかってしまったり，などの原因となる．

そこで本手法は，マルチクラウド環境にデータを分散して保管するストレージサービスを考える．そのうえで，ユーザの要求の定量化およびクラウドサービスの定量的な評価を行い，自動的にマッチングを行う手法を提案する[7]．更に，クラウドストレージに保存するデータごとに動的に分散方法を変化させることで，セキュリティやコストなどの最適化を目指す．

本稿の流れは以下の通りである．2 章において想定する環境や前提技術および，提案手法について説明する．3 章ではマルチクラウド環境の構成要素が特殊な場合における評価を行う．4 章では得られた知見および今後の課題について述べる．

２．提案手法

（１）想定する環境

本手法は，多数のクラウドストレージサービスが存在するマルチクラウド環境を想定している．また，それらのサービスは全て SLA を持っているものとする．各 SLA には，マッチングで必要となる項目が全て記載されていることを前提としている．そのため，それらの項目が一部でも欠けているクラウドサービスは利用できない．

それぞれのユーザは，上記の条件を満たす複数のクラウドサービスと契約済みであり，いつでも利用可能な状態であるものとする．マルチクラウド環境を利用する場合には，本手法で定めた要求を表す指標の全てを本手法に対して伝えるものとする．また，最適な保管先の決定はユーザの契約しているサービスの中から選択されるものとする．

クラウドサービスとの通信路および，ユーザがクラウドサービスに接続するために用いる端末は安全であるものと考える．

また，接続に用いる端末は一台のみを想定し，データの分散，復号処理はその端末上で行い，処理速度は十分であるものと考える．

また，すべてのストレージサービスは従量課金制であることを前提としている．

データを保存する際には，(k，L，n)秘密分散法を利用してデータを分散させる．次節では(k，L，n)秘密分散法をどのように適用するかについて述べる．

（２）(k, L, n)秘密分散法

(k，L，n)秘密分散法は山本によって考案され[2]，A.Shamir による(k，n)秘密分散法[6]を拡張した方式である．この方式は(k，n)秘密分散法と比較して各分散情報の容量を減少させることができるという特徴を持つ．

ある秘密情報 x に対して(k，L，n)秘密分散法を適用すると，n 個の分散情報が得られ，そのうち k 個を集めることで秘密情報の復元が可能になる．ここで，各分散情報のデータサイズは秘密情報の 1/L 倍となっている．このとき，k-L 個より多くかつ k 個未満の分散情報からは，秘密情報の一部を特定可能であり，それ以下の個数では情報理論的に安全であり，秘密情報に関する情報を一切得ることができない．

本手法では，この(k，L，n)秘密分散法をクラウドサービスに保存するデータを生成する際に利用する．ユーザがクラウドサービスに預けたいデータを秘密情報として(k，L，n)秘密分散法を適用する．得られた n 個の分散情報を，n 個の別々のクラウドサービスに保存しておき，元のデータが必要になった場合にはそのうちの k 個のサービスに接続し，分散情報を得ることで秘密情報が復元可能となる．各クラウドサービスで保管されるデータのサイズは元のデータの 1/L 倍である．

ここで，分散情報の保存先の数である n の最大値はユーザの契約しているクラウドサービスの数となる．

例として，あるデータに(2，1，3)秘密分散法を適用して分散保管を行う場合について考えてみる．

図 1 は(2，1，3)秘密分散法を適用して複数のクラウドサービスに分散して保管している状態を示している．

このとき，1 つのサービスに障害が発生してデータが取り出せない場合でも残りの 2 つから元のデータを復元することが可能であり，1 つのサービスに預けているデータが漏洩しても元のデータに関する情報は得られない．三つのパラメータ k，L，n を増減させることで可用性や機密性やコストなどの値も増減する．これにより，可用性や機密性，コストなどを最適化することが可能になると考えられる．

（３）ユーザの要求とクラウドサービスとのマッチング

ストレージサービスの特性として，常時通信が必要ではない，データを預けていない間には機密性は問題にならないなどクラウドやユーザの利用状況によって要求されるものが変わるため，段階分けをしてクラウドサービスの評価を行う．

この段階分けは 3 段階であり，アップロード時，データ保管時，ダウンロード時からなっている

以下にそれぞれの定義を示す．

1) アップロード時 … アップロードするデータの転送開始から転送完了まで

2) データ保管時 … マルチクラウド環境によってデータが管理されている間

3) ダウンロード … 復元に必要なデータの転送開始から完了まで

これら各段階におけるユーザの要求は以下の 4 つの指標で定義される．これらの指標は，関係のある段階でのみ評価される．

1) コスト … 預けるデータ 1MB あたりに必要な料金[円/MB]

2) 機密性 … クラウドサービスに預けたデータから元データが特定される危険度合い

3) 可用性 … 通信する必要のあるサービスと通信できる確率，つまり，マルチクラウド全体としての稼働率[%]

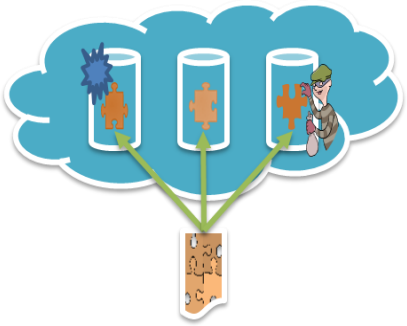

図 1 (2, 1, 3)秘密分散法の適用

4) 転送時間 … 預けるデータ 1MB あたりに対して転送が完了するまでに必要な時間，端的に言えばアップロードおよびダウンロード時間[s/MB]

また，これらの指標に対してマッチングに利用される，各サービスの SLA に記述されているべき項目は以下の 4 つである．

1) コスト　… 保存するデータ 1MB あたりにかかる費用 [円 / MB]

2) 漏洩確率 … 1 年あたりに起きるセキュリティインシデントの数 [インシデント数/年]

3) 稼働率　… システムが稼働している確率[%]

4) 通信速度 … 1MB のデータを通信するのにかかる時間 [s/MB]

ユーザの要求と SLA に記述された項目は一対一で対応しており，この対応をまとめたものを表 1 に示す．

前述の(k，L，n)秘密分散法のパラメータおよび，マルチクラウド環境を構成する個々のクラウドサービスの性質によって，マルチクラウド全体としての性質が決定される．

a）ユーザ要求に対応する式

1．アップロード時

アップロード時には，データを預けるためのサービスとの通信が発生するため，可用性および転送時間が関係する．また，ここで保存する保存するデータのサイズも決定するため，コストも必要となる．それぞれの式は以下の通りである．

a．コスト

全てのクラウドサービスに掛かる費用の総和であり，以下の式で表される．

$$\frac{1}{L}\sum_{i\in n}\text{Cost of Cloud}_i \tag{1}$$

b．可用性

分散情報を保存するすべてのクラウドサービスと通信可能である必要があるため，以下の式となる．

$$\prod_{i\in n}\text{Operating rate of Cloud}_i \tag{2}$$

c．転送時間

表 1 SLA 項目とユーザ要求の対応

SLA 項目	ユーザ要求
稼働率	可用性
コスト	コスト
通信速度	転送時間
漏洩確率	危険度合い

各サービスへのアップロードに掛かる時間の総和であり，以下で示される．

$$\frac{1}{L}\sum_{i\in n}\frac{1}{\textit{Communication speed of Cloud}_i} \tag{3}$$

2．データ保管時

データ保管時には通信が発生せず，すでに保存するデータサイズも決定しているため，可用性，転送時間およびコストは関係しない．唯一関係するのは元データの機密性だけである．

a．機密性

機密性は危険度合いという指標によって表されるが，これは，漏洩個数$x(0 \le x \le n)$を用いて x 個漏洩する確率とそのときの特定度合いの積の総和によって表され以下の式となる．

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

ここで，P(n)は n の冪集合であり，LP(i)は，クラウド i の漏洩確率とする．

また，特定度合いは(k, L, n)秘密分散法のパラメータから以下のように求まる．

$$\textit{Specific Level}(x)=\begin{cases}0\ (x\le k-L)\\ 1-\dfrac{k-x}{L}\ (k-L<x<k)\\ 1\ (k\le x)\end{cases} \tag{5}$$

3．ダウンロード時

ダウンロード時には，データを取り出す通信のために可用性および転送時間が関係する．

a．可用性

分散情報をアップロードしたクラウドサービスのうち元のデータを復元するのに必要なサービスと通信できる確率を以下の式によって求める．

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

ここで，A(i)はクラウド i の稼働率とする．

b．転送時間

元のデータを復元するのに必要なサービスと通信し，それらからデータを取り出すのに掛かる時間は以下で

ある．

$$\frac{k}{Ln}\sum_{i\in n}\frac{1}{\text{Communication Speed of Cloud}_i} \tag{7}$$

秘密分散法の各パラメータが増加した際にこれらの各指標がどう変化するかを表 2 に示す．

３．評価

ここで，マルチクラウドを構成するクラウドサービスの性質の違いによって 2 つに分ける．一つは構成するクラウドサービスの性質が全て同じであるホモジニアスな環境である．これを構成するクラウドサービスはコスト，漏洩確率，稼働率，通信速度の全ての項目が同じ値を持つクラウドサービス群である．同じプロバイダの同種のサービスを複数利用するような場合はこれに当て嵌まる．現実的にはホモジニアスなマルチクラウド環境を利用する機会は少ないと考えられるが，単純化および明確に区別するために導入する．

もう一つは，違う性質を持つものを許容するヘテロジニアスな環境である．これは，ホモジニアスな環境に属するもの以外すべてを包含する．様々なプロバイダのサービスを併用する場合はこちらに該当する．ホモジニアスな環境より一般的であると考えられる．

本稿では，仮想的なホモジニアスな環境を想定して評価を行う．また，評価に用いる仮想クラウドサービスの持つ値を表 3 に示す．

（１）ホモジニアスな環境における一般式

提案手法で示した，ユーザ要求に対応する式をホモジニアスな環境に適用する場合，より単純化することができる．

以下にそれぞれの段階で単純化された式を示す．

1．アップロード時

a．コスト

$$\frac{n}{L}\,クラウドのコスト \tag{9}$$

b．可用性

$$(クラウドの稼働率)^n \tag{10}$$

c．転送時間

$$\frac{n}{L*(クラウドの通信速度)} \tag{11}$$

2．データ保管時

a．機密性

$$\sum_{x=1}^{n}\binom{n}{x}(1-クラウドの機密性)^{n-x}(クラウドの機密性)^{x} * 特定度合い(x) \tag{12}$$

3．ダウンロード時

a．可用性

$$\sum_{x=1}^{n}\binom{n}{x}(1-クラウドの稼働率)^{n-x}(クラウドの稼働率)^{x} \tag{13}$$

b．転送時間

$$\frac{k}{L*(クラウドの通信速度)} \tag{14}$$

（２）ダウンロード時の可用性，機密性およびコストによる分析

本稿においては，表 2 でトレードオフの関係として見ることができるダウンロード時の可用性と機密性の関係性および，そのときのコストを用いて分析を行う．また，分かりやすさのために以降のグラフでは可用性の値は元の値を 1 から引いたもので示す．これにより，すべての軸は値が小さいほど有用なものとして見ることができる．

まず，ユーザが最大 7 個のクラウドを利用できる場合について，k,L,n を取り得る範囲内で変化させたグラフを図 2，図 3 に示す．これらのグラフでは，同じ k の値を持つ点同士が結ばれており，それぞれの点上に(k，L，n)の値が示されている．また，図 2，図 3 ともに x 軸は機密性，y 軸はダウンロード時の可用性であり，図 2 の z 軸はコストとなっている．ここで(1,1,1)のパラメータを持つ点つまり，単一のクラウドサービスのみを利用した際の点に着目すると，他の点と比較して低コストである．しかし，可用性と機密性の値について，両方の値が高いものは 6 個しか存在せず，それ以外の点はどちらかの指

表 2 秘密分散法のパラメータとユーザの要求指標の関係

	アップロード時			保管時	ダウンロード時	
	可用性	コスト	転送時間	機密性	可用性	転送時間
k	-	-	-	良化	悪化	悪化
L	-	良化	良化	悪化	-	良化
n	悪化	悪化	悪化	悪化	良化	-

表 3 仮想クラウドの性質

稼働率	コスト	通信速度	漏洩確率
0.99	1	1	0.001

標において，単一クラウドサービスを利用した場合よりも優位であると言える．

また，ホモジニアスな環境であるにも関わらず，3 つの指標がすべて同じ値を持つものは存在していない．このことからホモジニアスな環境においても多様なニーズに答えることが可能であると言える．

また，図 3 において可用性と機密性が共に最大となる点は存在せず，表 2 のとおりトレードオフの関係になっていることが見てとれる．そのなかで(4, 1, 7)のパラメータを持つ点は，高コストではあるが，可用性と機密性の指標の中間に位置しており，両方の指標が重要な場合には選択肢の候補となりうる．

ここで，このグラフスケールを維持したまま，最大 8 個のサービスを用いた場合のダウンロード時の可用性と機密性の関係性を示した図 4 に示す．すると，(4，1，7)のパラメータを持つ点の最近傍に，その他の点と比較して似た性質を持つ(5，2，8)のパラメータを持つ点が表れている．この二つの点のコストを比較すると(4，1，7)のパラメータを持つ点が 7[円/MB]，(5，2，8)のパラメータを持つ点が 4[円/MB]であり，(4，1，7)のパラメータを持つ点の方が 1.75 倍のコストがかかる．よって，コストを抑えたい場合には(4，1，7)のパラメータを持つ点よりも(5，2，8)のパラメータを持つ点を選ぶ方が良い．

更に，このグラフスケールを維持したまま最大 15 個利用できる場合の関係性を示した図 5 を見てみると，(4，1，7)より可用性と機密性を両立できている点が増えている．特に(8,1,15)のパラメータを持つ点は，このグラフスケール内で可用性と機密性を最大化できる点であると言える．このことから，利用できるクラウドサービスの最大数を増やすことができれば，ダウンロード時の可用性と機密性を両立できるようになっていくことが分かる．

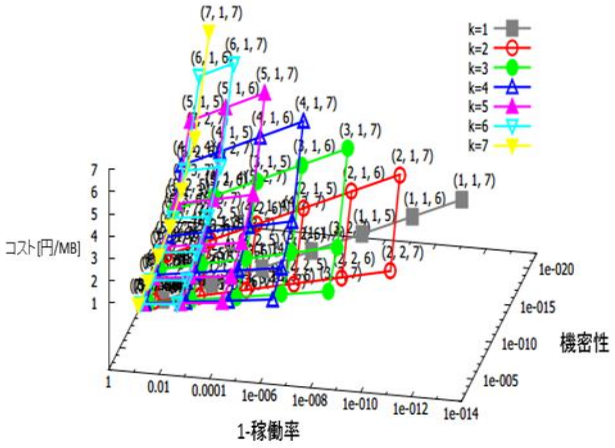

図 2 ダウンロード時の可用性，機密性，コスト

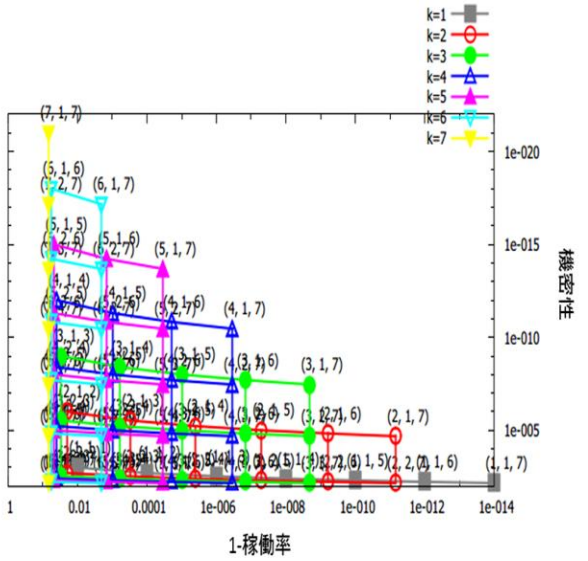

図 3 最大 7 個のクラウドを利用した場合のダウンロード時の可用性と機密性

また，(4,1,7)のパラメータを持つ点と(8，2，14)のパラメータを持つ点はどちらも同じコストであるが，(8，2，14)のパラメータを持つ点の方が可用性と機密性の値は低く有用である．加えて，可用性と機密性について(8,2,14)のパラメータを持つ点の近傍にある(9,3,15)のパラメータを持つ点は，コストが(8,2,14)のパラメータを持つ点よりも安価である．この結果から，ユーザが利用できるクラウドの最大数が 7 個のときよりも最大数が 15 個の時の方が，可用性と機密性を両立したうえで，更にコストも抑えることが可能になると言える．つまり，ユーザはできるかぎり多くのクラウドと契約しておくことでより高い水準の可用性と機密性を安価に得ることができる．

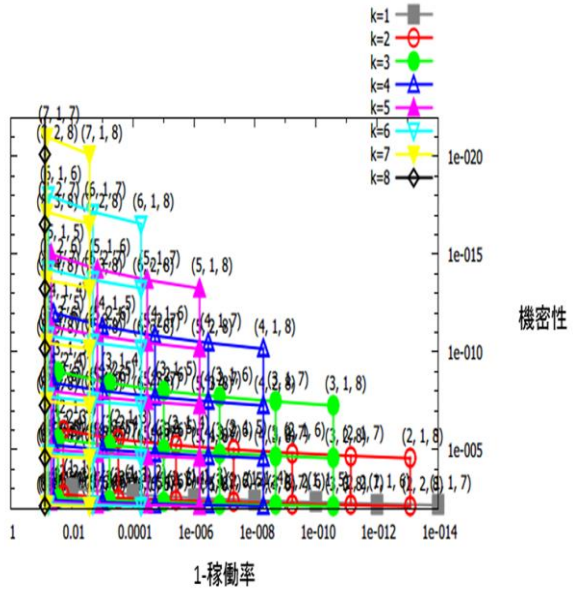

図 4 最大 8 個のクラウドを利用した場合のダウンロード時の可用性と機密性

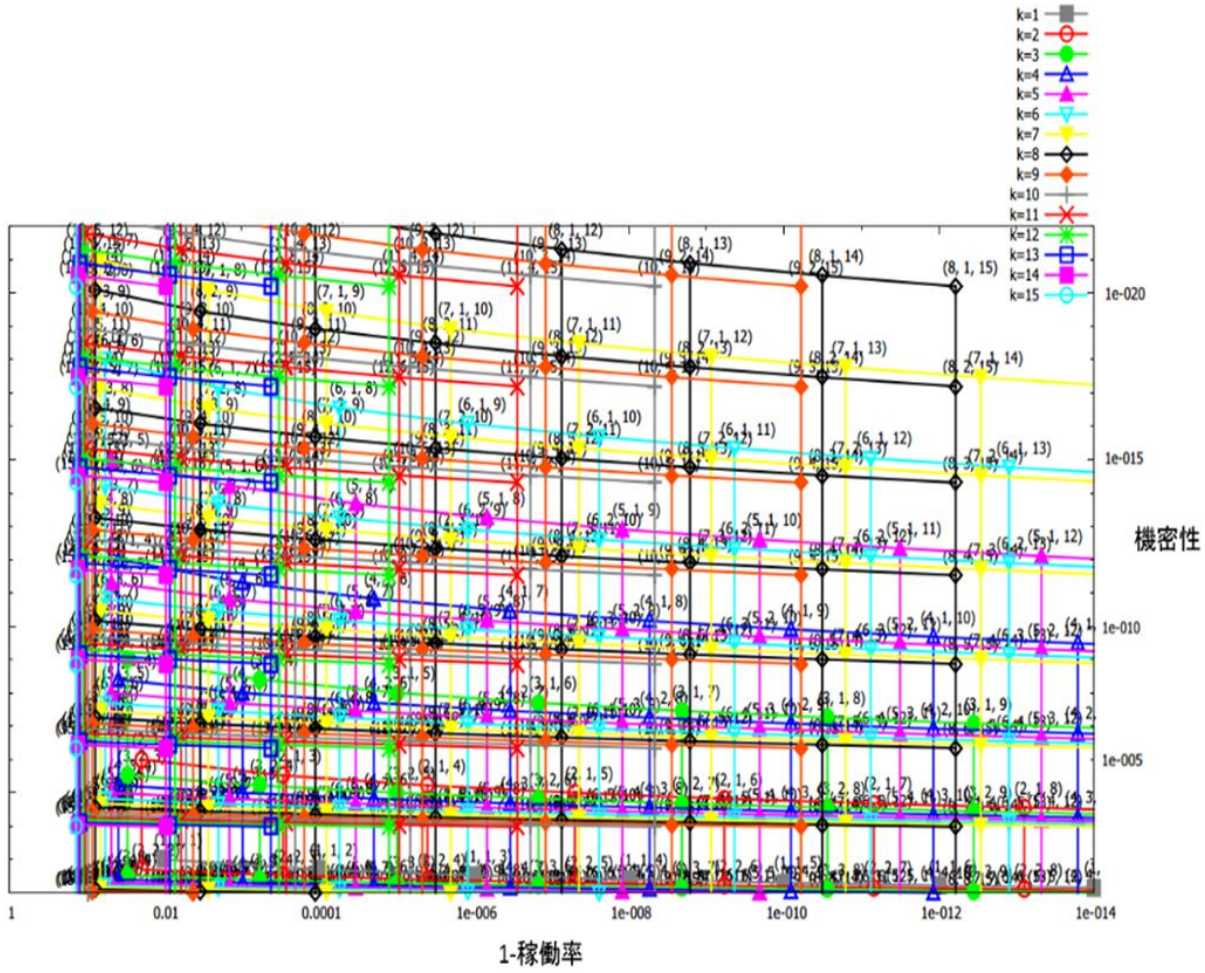

図 5 最大 15 個のクラウドを利用した場合のダウンロード時の可用性と機密性

（1）ヘテロジニアスな環境における評価

ヘテロジニアスな環境の評価は単一のプライベートクラウドと，パブリッククラウドのみによって構成されたマルチクラウド環境との比較によって行う．この評価で利用するプライベートクラウドとパブリッククラウドのリストを以下の表 4 に示す.

実際に評価値を算出し，コスト，ダウンロード時の稼働率，危険度合いの全ての項目でプライベートクラウド単体を上回る値を持つヘテロジニアスクラウド環境のうちいくつかの数値で他より良い値を持つものを表 5 に示す.

これらの結果をまとめると以下の通りである.

コンビネーション 0 :高可用性

コンビネーション 8 :低コスト

コンビネーション 14:3 指標全てが平均的

コンビネーション 18:高機密性

ユーザはこれら 4 つの候補から保管したいデータに要求されるものによって利用するものを決定する．これにより，プライベートクラウドを単体で利用する場合よりも稼働率，コスト，機密性の点で有用性があり，細かいニーズに対応したヘテロジニアスクラウド環境を提供することができた.

４． 結論

ユーザの要求とクラウドサービスのマッチングを行う方法を示し，ホモジニアスな環境における評価を行い，ホモジニアスな環境においても多様な値を作り出すことができることが分かった．また，利用できるクラウド数が少ない場合と比較して，利用できるクラウド数が多いほど，ダウンロード時の可用性と機密性を両立し　たうえで低コスト化も見込めることが分かった.

今後はその他の指標を用いた評価および，ヘテロジニアスな環境における評価を行う.

参考文献

1) (NIST), http://www.nist.gov/itl/cloud/

2) H. Yamamoto, “Secret Sharing System using (k, L, n) threshold scheme,” Electron. Commun. Jpn. (Part I: Commun.), vol. 69, no. 9, pp.46-54, 1986.

3) M. Vukolic,"The Byzantine empire in the intercloud", ACM SIGACT News, 41,2010, pp. 105-111.

4) A. Bessani, M. Correia, B. Quaresma, F. André and P. Sousa, "DEPSKY: dependable and secure storage in a cloud-of-clouds", EuroSys'11:Proc. 6thConf. on Computer systems, 2011, pp. 31-46.

5) H. Abu-Libdeh, L. Princehouse and H.Weatherspoon, "RACS: a case for cloud storage diversity", SoCC'10:Proc. 1st ACM symposium on Cloud computing, 2010, pp. 229-240.

6) A. Shamir, "How to share a secret", Communications of the ACM, 22(11), 1979, pp.612-613.

7) Yuuki Kajiura, Atsushi Kanai, Shigeaki Tanimoto, Hiroyuki Sato, "A File-distribution Approach to Achieve High Availability and Confidentiality for Data Storage on Multi-cloud", SAPSE2013, 2013

表 4 ヘテロジニアス環境に利用するクラウド群

	可用性	*コスト*	*機密性*	*名前*
Private Cloud	0.999	1	0.001	P
Public Cloud0	0.9999	0	1.0	p0
Public Cloud1	0.999	0.1	0.5	p1
Public Cloud2	0.995	0.2	0.1	p2
Public Cloud3	0.99	0.3	0.05	p3
Public Cloud4	0.90	0.4	0.01	p4
Public Cloud5	0.9999	0.8	0.01	p5
Public Cloud6	0.999	0.3	0.05	p6
Public Cloud7	0.995	0.2	0.1	p7
Public Cloud8	0.99	0.1	0.5	p8
Public Cloud9	0.90	0	1.0	p9
Public Cloud10	0.999	0.8	0.01	p10

表 5 最適ヘテロクラウド環境の候補

組み合わせ	*k*	*l*	*n*	*コスト*	*機密性 (*10^4)*	*可用性*	*クラウド*
Combination 0	4	2	5	0.9	87275	0.999854	p2,p3,p5,p6,p7
Combination 8	5	3	6	0.87	68847.92	0.999833	p2,p3,p5,p6,p7,p10
Combination 14	6	3	7	0.9	35362.54	0.999811	p1,p2,p3,p5,p6,p7,p10
Combination 18	7	3	8	0.93	18155.9	0.999584	p1,p2,p3,p5,p6,p7,p8,p10