

## 異なるネットワーク遅延時間を提供するクラウド環境における 最適資源割当てアルゴリズムの基礎評価

栗野 勇輝\*<sup>1</sup>, 栗林 伸一\*<sup>2</sup>

Proposed joint multiple resource allocation method for cloud computing services with  
different network delay

Yuuki AWANO\*<sup>1</sup> and Shin-ichi KURIBAYASHI\*<sup>2</sup>

**ABSTRACT** : In a cloud computing environment, it is necessary to simultaneously allocate both processing ability and network bandwidth needed to access it. The authors proposed the joint multiple resource allocation method in a cloud computing environment that consists of multiple data centers and each data center provides the different network delay. It is also highly likely that each data center would provide the different network delay to users at different locations.

This paper proposes to enhance the joint multiple resource allocation method in a cloud computing environment in order to handle the case where each data center provides the different network delay time to users at multiple locations. It is demonstrated by simulation evaluations that the proposed method can reduce the total request blocking probability up to 50%, compared with the conventional methods.

**Keywords** : Cloud computing , Joint multiple resource allocation, Different network delay

(Received Aug. 29, 2012)

### 1. まえがき

クラウドコンピューティング環境<sup>(1), (2)</sup>では、計算能力やストレージだけでなくそこにアクセスするための帯域も同時に確保する必要があり、要求ごとに貸し出すクラウド資源種別は1種類でなく複数存在する。また、クラウドコンピューティング環境を構成する個々のデータセンタが提供するサービス品質は均一でなく異なる可能性が高い。

本研究室では、計算能力や帯域など異なる資源種別の資源を同時に割当てることを前提とした最適資源割当て方式を提案した<sup>(3), (4)</sup>。さらに、計算資源を持つ各センタまでのネットワーク遅延時間の違いを考慮した最適資源割当て方式（以後、方式3）を明らかにした。さらに、文献<sup>(5)</sup>では、ネットワーク遅延時間だけでなく、複数の資源属性を考慮した最適資源割当て方式を明ら

かにした。しかし、それら方式は特定のアクセス地点だけを想定したものである。実際のシステムでは、例えば同じセンタであってもアクセス地点によりネットワーク遅延時間が異なることが想定される。

このため、本稿では同じデータセンタであってもアクセス地点によりネットワーク遅延時間が異なることを想定した複数資源同時割り当て方式（以後、方式4）を提案し、シミュレーション評価によりその有効性を明らかにする。なお、今回は基本評価を目的とするため、資源属性としてネットワーク遅延時間のみを考慮する。

### 2. クラウドコンピューティング環境における資源割当てモデル

クラウドコンピューティング環境における資源割当ての前提となるシステムモデルを図1に示す。これは文献<sup>(3)~(5)</sup>でも採用したものである。つまり、 $k$ 個のセンタが異なる場所に存在し、各センタは計算能力と帯域の2つの資源種別の資源を提供する。要求発生時に、 $k$ 個の

\*<sup>1</sup>:理工学部情報科学科4年生(us092008@cc.seikei.ac.jp)

\*<sup>2</sup>:理工学部情報科学科教授(kuribayashi@st.seikei.ac.jp)

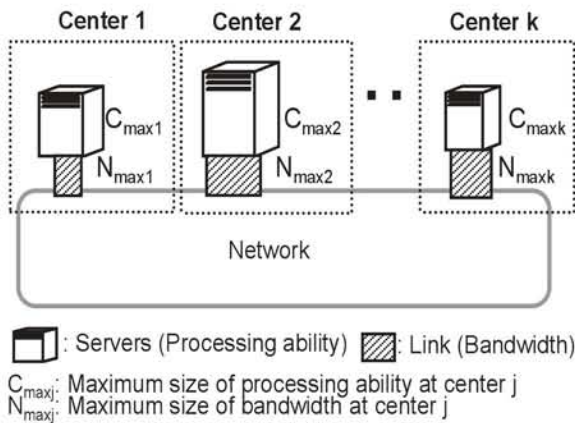
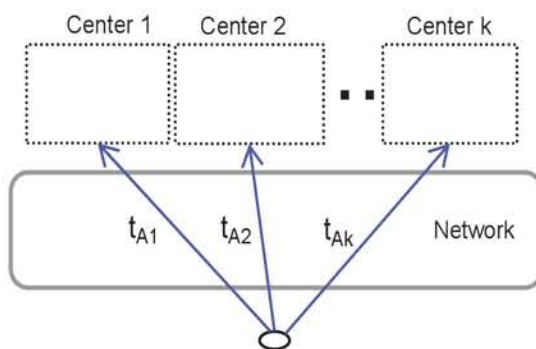


図1 クラウド環境における資源割当てモデル



$t_{Aj}$ : Network delay time to access Center  $j$  at location  $A$

図2 センタアクセス時のネットワーク遅延時間  
(アクセス地点毎に異なる)

中から最適なセンタを1つ選択し、その選択したセンタから計算能力と帯域を同時にその要求に割当てる。計算能力と帯域を別々のセンタから割当てることはできない。サービス時間が終了すると、割当てた計算能力と帯域は同時に解放される。

また、図2に示すように、アクセス地点から各センタまでのネットワーク遅延時間は異なり、さらにその値はアクセス地点によっても異なるものとする。

### 3. 方式3の課題と対策案

#### 3.1 方式3の課題

アクセス地点毎にネットワーク遅延時間が異なることを想定しない方式3に基づき資源割り当てを実施すると、アクセス位置の違いによるサービス競合が発生する可能性がある。例えば、図3ではアクセス地点Aからみるとセンタ1は低遅延のアクセスを提供するため、遅延時間に対する条件が厳しくない要求に対してはセンタ2を積極的に選択する。これにより、今後発生する遅延時間に

対する条件が厳しい要求のためにセンタ1の資源をより多く確保することができる。ところがアクセス地点Bからみると逆にセンタ1を積極的に使用しようとするため、センタ1をできるだけ確保しようとする地点Aの対応と競合することになってしまう。

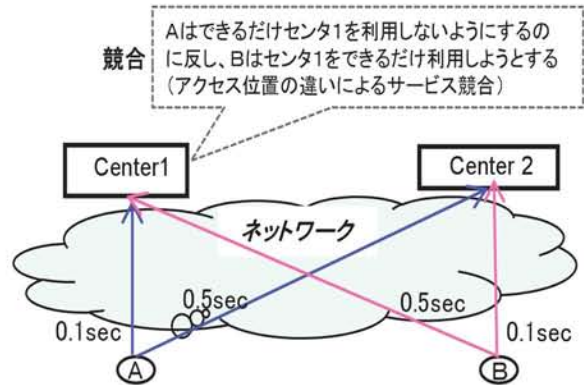


図3 アクセス位置の違いによるサービス競合の例

#### 3.2 対策案

3.1節で説明した「アクセス地点によるサービス競合」を緩和する方式(以後、方式4)を以下に提案する。

##### (1) 基本的な考え方

全てのアクセス地点、全ての要求品質の平均要求棄却率を最小化することを目的関数とする。

##### (2) 対策案

案1. センタの空き資源量が閾値(=最大資源量\*alevel)以下になった場合に、もっとも要求品質が厳しい要求だけを受け付ける。

案2. センタ毎にアクセス地点対応の資源を個別に確保し、残りは共用する。

公平性を考えると案2が有効な場合もあるが論理が複雑であることから、今回は案1を前提に評価を行う。

### 4. 方式4の有効性評価

#### 4.1 評価条件

1) C言語を用いたコンピュータシミュレーションにより評価を行う。

2) 基礎的な評価を行うため、今回はアクセス地点としてA, Bの2点を考慮する。また、資源属性としてネットワーク遅延時間だけを想定する。

3) センタは、センタ1, センタ2の2つ存在し、以下のネットワーク遅延時間を提供するものとする。

##### 【センタ1】

地点Aに対しては「優先品質」(ネットワーク遅延時間が短い)、地点Bに対しては「通常品質」(ネットワーク



遅延時間が長い), を提供。

【センタ2】

地点Aに対しては「通常品質」, 地点Bに対しては「優先品質」, を提供。

4) alevelはセンタ1に対してのみ有効とし, センタ2に対しては設定しない。

5) 要求発生比率

・要求発生比率

$$\text{地点A} : \text{地点B} = \text{prob\_a} : (1 - \text{prob\_a})$$

・地点Aにおける要求発生比率

「通常」要求: prob\_10, 「通常」要求: 1.0 - prob\_10

・地点Bにおける要求発生比率

「通常」要求: prob\_20, 「通常」要求: 1.0 - prob\_20

6) 要求が発生すると, 3章で提案した案1に従って最適なセンタを1つ選択し, 必要となる計算能力と帯域をそのセンタから同時に割当てる。なお, 今回は比較のため, 方式3も評価する。

7) 要求毎に必要な計算能力と帯域の大きさは全要求種別に共通で, それぞれ平均C, Nのガウス分布(分散は5)に従う。

8) 要求の発生間隔ならびには資源割当て時間は全要求

に共通で, それぞれ平均rの指数分布, H(一定)に従うものとする。割当てられた資源は資源割当て時間経過後に同時に解放する。

9) 要求発生パターンは,

$\{C=a_1, N=b_1; C=a_2, N=b_2; \dots; C=a_w, N=b_w\}$ の繰返しとする。ここで, wは1繰返し内の発生要求数,  $a_u$  ( $u=1 \sim w$ )はu番目の要求のCの大きさ,  $b_u$  ( $u=1 \sim w$ )はu番目の要求のNの大きさ, をそれぞれ示す。

4.2 シミュレーション結果と評価

シミュレーション結果を図4, 図5に示す。なお,  $C_{max1}=N_{max1}=30$ ,  $C_{max2}=N_{max2}=30$ ,  $C=N=7$ ,  $\text{prob\_a}=0.6$ とした。また, 方式4として提案した制御(案1)はセンタ1だけで実施する。

図4では全要求の60%が地点Aから発生し, その90%が「優先」を要求, 10%が「通常」(prob\_10=0.1)を要求する。地点Bから発生する要求の10%が「優先」を要求, 90%が「通常」(prob\_20=0.9)を要求する。地点Aの「優先」要求はセンタ1のみ選択可能であるが, 「通常」要求はセンタ2ならびにセンタ1を選択可能である。一方, 地点Bの「通常」要求はセンタ1をまず選択し, 空き資源がない時にはセンタ2を選択する。横軸

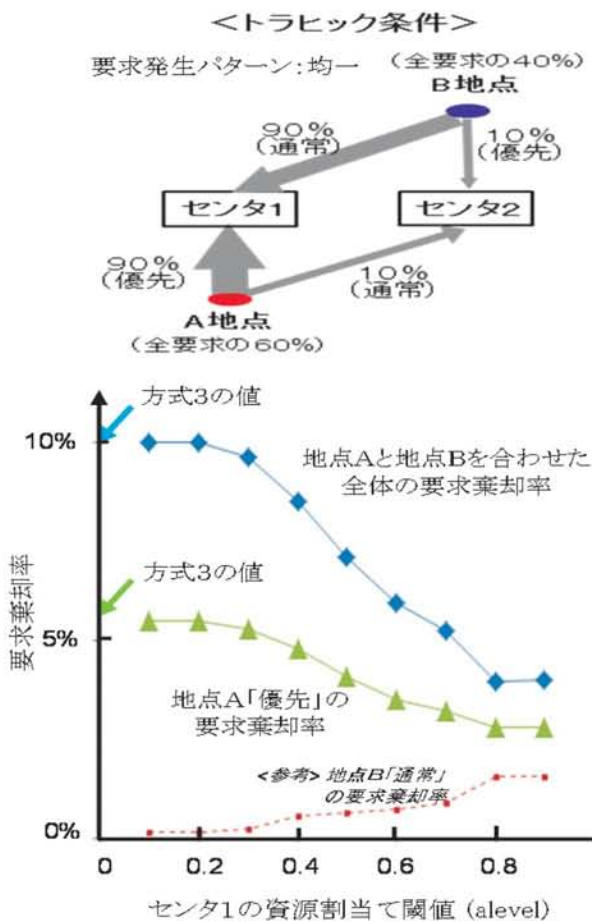


図4 方式4の評価結果例1

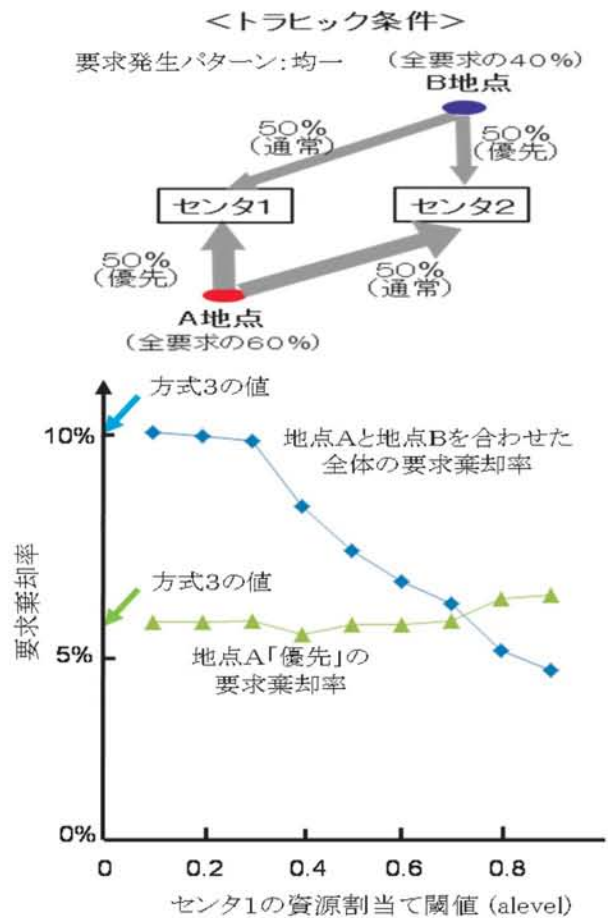


図5 方式4の評価結果例2

は、センタ1における閾値係数 $\alpha$ の値である。

図5では、地点Aと地点Bから発生する「優先」要求と「通常」要求の比率をそれぞれ50%とした場合 ( $\text{prob}_{10}=\text{prob}_{20}=0.5$ ) の評価結果である。

図4および図5より、方式3に比べ方式4は全体の要求棄却率を削減できることがわかる。特に、図4のように、地点Aからの「優先」要求に対する資源に余裕がない場合にその削減効果が大きくなる (図4の例では約半分)。

これは以下の理由による。

(1) 方式3では、地点Aから発生する「優先」要求はセンタ1しか選択できないにも関わらず、地点Bから発生する「通常」要求が (センタ2も利用できるにも関わらず) センタ1資源を優先的に使用する。このため、地点Aから発生する「優先」要求の棄却率が大きくなってしまふ。地点Bから発生する「通常」要求はセンタ1が無理なら資源に余裕のあるセンタ2も選択できる。

一方、方式4はセンタ1において「優先」要求に対する資源割り当てを優先することで (地点Bの「通常」要求アクセス制限) その棄却率を削減し、結果として全体の要求棄却率も少なくすることができる。

(2) 今回評価結果は示していないが、地点Aの「優先」要求が多くなり過ぎると地点Bの「通常」要求アクセス制限効果が薄れてくることも確認している。

## 5. むすび

本稿は、広域に分散した複数のデータセンタから構成されるクラウドコンピューティング環境を前提に、同一センタであってもアクセス地点によりネットワーク遅延時間が異なることを想定した複数資源同時割り当て方式 (方式4) を提案した。シミュレーション評価により、提案方式は従来方式 (方式3) に比べ全体の要求棄却率を大幅に削減できることを示した。

今回は、基本評価を行うため、資源属性はネットワーク遅延時間のみを考慮し、アクセス地点数、センタ数などを制限した評価のみ行った。今後は、それらの数を増加させた場合の評価を行い、提案方式の有効性とその条件を明確にしていく予定である。

## 参考文献

[1] G.Reese: “Cloud Application Architecture”, O’Reilly& Associates, Inc., Apr. 2009.

[2] J.W.Rittinghouse and J.F.Ransone: “Cloud Computing: Imprementation, Management, and Security”, CRC Press LLC, Aug. 2009.

[3] 津村, 栗林: “コンピュータネットワークにおける複数資源同時割当て方式の提案”, 成蹊大学理工学研究報告 VOL.43, No.2, pp.107-120 (Dec. 2006)

[4] S.Kuribayashi, “Optimal Joint Multiple Resource Allocation Method for Cloud Computing Environments”, International Journal of Research and Reviews in Computer Science (IJRRCS), Vol. 2, No.1, Feb. 2011.

[5] Yuuki Awano and Shin-ichi Kuribayashi, “Reducing Power Consumption and Improving Quality of Service in Cloud Computing Environments”, Proceeding of the 15-th International Conference on Network-Based Information Systems (NBIS-2012), Sep. 2012.