成蹊大学理工学研究報告
J. Fac. Sci.Tech., Seikei Univ.
Vol.50 No.2 (2013)pp.33-36
（研究速報）

アクセス地点によりネットワーク遅延時間が異なるクラウドコンピューティング環境における最適資源割当て方式の基本検討

眞籠　祐太朗*1，栗林　伸一*2

Proposed resource allocation method for cloud computing environments with different network delay to users at multiple access points

Yutaro MAGOME*1 and Shin-ichi KURIBAYASHI*2

(Received Sep. 23, 2013)

1．はじめに

クラウドコンピューティング環境[1],[2]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は広域に分散し，各センタが提供するサービス品質は均一でなく異なる。

本研究室では，各要求に対して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に，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考慮した最適資源割り当て方式（以後，**方式4**）を提案した[3],[4]。しかし，方式4は，短い遅延時間を必要とする要求に対する品質確保に重点を置いていたため，短い遅延時間を必要とする要求が想定以上に発生した場合の対処が考慮されていない。さらに，方式4はアクセス遅延のみを考慮し，計算能力の資源属性である計算時間を含めたトータル処理時間を考慮したものではない。

このため，本稿では方式4に対して，i)短い遅延時間を必要とする要求が想定以上に発生した場合に他要求種別の品質劣化を防止する機能，ii)センタでの計算時間も含むトータル処理時間を考慮したセンタ選択機能，を追加した方式を提案し，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，文献[3],[4]と同様に，資源割当てにおける影響の大きい資源を着目資源とし，以下着目資源を前提に資源割当てアルゴリズムを説明する。さらに，今回は基本的な評価のみをまとめたものである。

*1：情報科学科４年生
*2：情報科学科教授（kuribayashi@st.seieki.ac.jp）

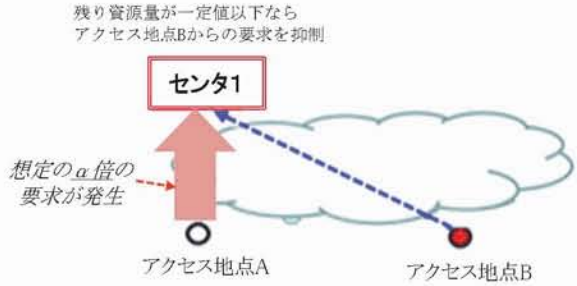
2．従来方式（方式4）の課題と対策

2．1　課題1

＜課題の内容＞

従来方式は，短い遅延時間を必要とする要求（要求種別1）と長い遅延時間を許容する要求（要求種別2）で同じセンタの資源を共用し，その空き資源が一定値以下になると要求種別2への資源割当てを抑制する。これにより，特定のセンタしか選択できない要求種別1の品質を確保しようとする。図1の例では，センタ1の空き資源が一定値以下になる地点Bからのアクセスを抑制し，地点Aから発生する要求種別1の要求向け資源を確保する。

しかし，図1において地点Aからの要求種別1の要求が想定以上に大量に発生すると（αは比率であり，想定から何倍に増加しているかたを示す），センタ1の資源をほとんど使用するため，地点Bから発生する要求種別2の要求をほとんど処理できなくなる。その結果，地点Bから発生する要求の要求棄却率が大幅に増加してしまう。



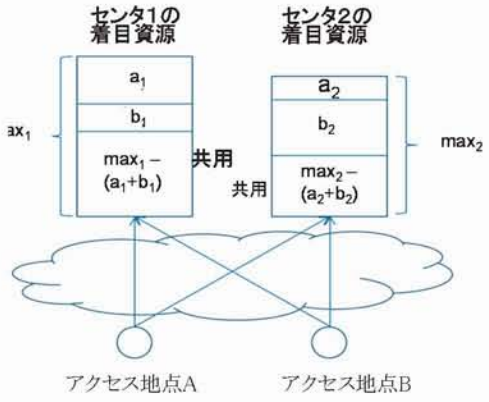
＊アクセス地点Aからの発生要求数が想定よりも大幅に増加すると、センタ1の資源はアクセス地点Bからの要求にはほとんど割当てられなくなってしまう。

図1　従来方式の課題1

＜対策＞

従来方式をベースに，以下の機能を追加する（以後，**方式 4 改**）。つまり，センタ毎に，アクセス地点対応に一定量の資源を確保し，残りを共用する。なお，アクセス地点対応に確保する資源量は，例えばアクセス地点ごとに必要品質を確保するための最小資源量とする。また，確保する資源量は実際の発生量に応じて動的に変化させる（共用資源が 0 というケースも想定）。

センタ数 2 の場合の方式 4 改の概要を図 2 に示す。なお，現実のアクセス地点数を考慮すると，エリア毎，または要求種別毎にアクセス地点を複数グループに集約し，それらグループごとにセンタ選択を実施する必要がある。



a_1,a_2：アクセス地点Aからの要求専用に確保する資源量
b_1,b_2：アクセス地点Bからの要求専用に確保する資源量
max_1、max_2：センタ1、センタ2の最大資源量

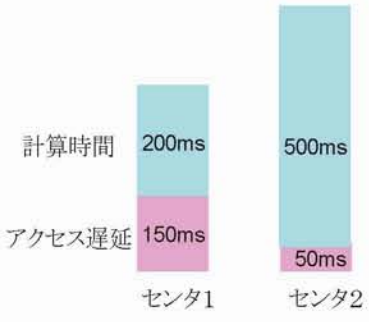
図 2　提案する方式 4 改の概要

2. 2　課題 2

＜課題の内容＞

従来方式は，計算能力の資源属性である計算時間を考慮していない。そのため，例えば図 3 のように，トータル処理時間が長いセンタ 2 を本来優先的に選択すべきであるが，遅延時間だけ考慮する従来方式はセンタ 1 を優先的に選択してしまう。

＜対策＞



従来方式はアクセス遅延しか考慮していないため、アクセス遅延にセンタでの計算時間を加えたトータル時間が短いセンタ(この図ではセンタ1)を優先的に選択してしまう可能性がある。

図 3　従来方式の課題 2

遅延時間だけでなく，遅延時間（上りと下りの合算）と計算時間のトータル処理時間が短くなることを必要とする要求を要求種別 1，トータル処理時間が長くなることを許容する要求を要求種別 2，と置き換え，方式 4 改と同じアルゴリズムを適用する（以後，**方式 5**）。

3．提案方式の有効性評価

3. 1　シミュレーションモデル

基本的に，文献[3],[4]と同じモデルを前提とする。なお，本稿で新たに想定する条件は 3.2 節，3.3 節の中で個別に説明する。

3. 2　方式 4 改の評価

＜シミュレーション条件＞

・センタ数：2（センタ 1，センタ 2）

・max_1=160，max_2=20；a_1=20，a_2=60，a_2=0，b_2=20

・アクセス地点数：2（地点A，地点B）

・地点Aからの発生要求数と地点Aからの発生要求数の比率を 6:4 と想定。

・地点Aからセンタ 1 ならびにセンタ 2 へのアクセス遅延はそれぞれ 50ms，150msと想定。

・地点Bからセンタ 2 ならびにセンタ 1 へのアクセス遅延はそれぞれ 50ms，150msと想定。

・地点Aから発生する要求はセンタ 1 のみ選択可能，地点Bから発生する要求はセンタ 1 とセンタ 2 の両方を選択可能と想定。

・地点Aからの要求数が想定を大きく上回って発生した場合，方式 4 改ではアクセス地点毎の想定発生要求数に比例し，かつ$a_1+a_2=max_1$になるようにa_1,b_1を再設定する（共用部分は 0 とする）。一方，方式 4 では地点Aから発生する要求数によらず閾値は固定とする。

＜シミュレーション結果と考察＞

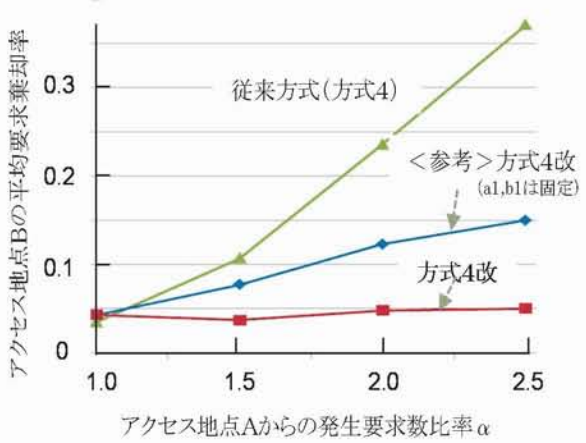


図 4　アクセス地点Ｂの平均要求棄却率

図 4 にシミュレーション結果を示す。図 4 の横軸は比率 α を示す。これは，地点Aからの要求が想定の何倍発生したかを示すもので，1.0 が想定通りの要求数が発生した場合である。縦軸は，地点Bから発生する要求の平均棄却率を示す。また，方式 4 改で α に関係なく a_1,b_1 を固定した場合も参考に評価した。

この図から，方式 4 改では地点Aからの発生要求数が想定の 2 倍以上になっても地点Bの要求棄却率を一定値以下に保つことができることがわかる。なお，α に対応して a_1,b_1 を変更することが必要である。

3. 3　方式５の評価

＜シミュレーション条件＞

・今回は，計算時間を考慮するという基本的なことだけでどれだけの効果が得られるかを評価するため，方式 4 との比較を行う。

・センタ数：2（センタ 1，センタ 2）

・max_1=100，max_2=100

・アクセス遅延，計算時間は図 3 と同じ値を想定。

・要求種別 1（センタ 2 のみ利用可能）と要求種別 2（両方のセンタを利用可能）の発生比率は 3:7 を想定。

＜シミュレーション結果と考察＞

シミュレーションの結果，方式 5 は方式 4 に比べ，平均要求棄却率を最大で 3 割～4 割程度改善できることがわかった。

4．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，従来提案された複数資源の同時割当て方式に対して，i)短い遅延時間を必要とする要求が想定以上に発生した場合に他要求種別の品質劣化を防止する機能，ii) センタでの計算時間も含むトータル処理時間を考慮したセンタ選択機能，を追加した方式（方式 4 改，方式 5）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した。

今回は，基本評価を行うためアクセス地点数，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詳細な評価を行い，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: Impremention, Management, and Security”, CRC Press LLC, Aug. 2009.

[3] 粟野，栗林：“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”，成蹊大学理工学報告 Vol.49, No.2, pp.101-104 (Dec. 2012)

[4] 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.