リンクの輻輳状態を考慮した動的なミラーサーバ選択方式

和泉勇治[†] 宇津江康太[††]
加藤寧[†] 根元義章[†]

負荷分散システムとしてネットワーク上にミラーサーバを設置するサイトが増加し，ユーザは複数のミラーサーバからアクセスするサーバを選択可能となってきている．サーバ選択の際，スループットを計測することで，高精度な選択が可能となるが，測定トラヒックが膨大になる問題点がある．本論文ではサーバ選択における測定トラヒックを削減するために，ネットワークの混雑状況とサーバなどの負荷状況によるスループットへの影響を分類し，ネットワークの混雑状況からボトルネックを検出する指標を定義し，ボトルネックの位置情報を抽出する．さらに，この情報を基にサーバのグループ化を施した低コストなサーバ選択方式を提案し，実ネットワークにおいて有効性を検証する．

Dynamic Mirror Server Selection Method with Consideration about Congestion of Link

Yuji Waizumi,[†] Kouta Utsue,[††] Nei Kato[†] and Yoshiaki Nemoto[†]

Web sites which have mirror servers for load sharing are increasing. Users can select a server to access from mirror servers. Although user can select the most appropriate server by measuring its throughput when selecting a server, the measuring traffic becomes enormous amounts. In this paper, we analyze factors affecting the performance of network and define a index of bottleneck. Using the index of bottleneck and positional information of servers, we propose a server selecting system which can select the most appropriate server with lower measuring cost.

1. はじめに

近年，Linux[1] に代表されるフリーソフトが多くのコンピュータ上で利用されるようになってきている．そのようなフリーソフトの多くは，インターネット上に公開され，ネットワークを経由しての取得が可能となっている．Linux のような OS やコンパイラなどの開発環境は，配布されているサイズが数百 k バイトから数十 M バイトにも及ぶ大規模なものが多い．そのようなソフトウエアの配布を行う際，サーバやネットワークの負荷の集中を防ぎ，ユーザがそれらのソフトウエア（以下，コンテンツと呼ぶ）を短時間で効率的に取得できるよう，同一のコンテンツを配布しているミラーサーバが数多く運用されるようになってきている．

効率的なコンテンツ取得を目的として，CDS（Content Distribution Service/System）[2)~4)] のようなシステムが注目されている．それらのシステムでは，DNS（Domain Name System）のラウンドロビン[5)]，最小接続数，送出パケット数，CPU 負荷といった基準でアクセスするミラーサーバを選択する負荷分散技術が用いられている[6)]．これらの負荷分散技術は，サーバ側の情報のみを利用しているため，ユーザ/サーバ間のネットワークの混雑状況を反映した負荷分散が行われない．そのため，コンテンツ取得時間の短縮という観点から最適なミラーサーバの選択が困難であるという欠点がある．

ユーザ/サーバ間のネットワークの途中経路を考慮した手法として，BGP4[7)] などの経路制御プロトコルで得られる情報から，ユーザ/サーバ間を経由する AS（Autonomous System）[8)] 数を算出し，サーバを選択する方式も提案されている[10),11)]．これらの手法は，通過する AS 数という時間的に変化の少ない情報を基準とするため，時々刻々変化するネットワーク状態に追随することは不可能である．文献 9) では，ルー

† 東北大学大学院情報科学研究科システム情報科学専攻
Department of System Information Sciences, Graduate School of Information Sciences, Tohoku University
†† 株式会社日立東日本ソリューションズ
Hitachi East Japan Solutions

タ数とコンテンツ取得時間が必ずしも比例関係ではないことが示され，ルータ数を基準としてクライアントに最近隣のサーバを選択することがコンテンツ取得時間の短縮につながらないことを明らかにしている．

また，traceroute[13] などの測定ツールを利用し，サーバまでの距離として通過経路上のルータの数や各ルータまでの RTT（Round Trip Time）を比較することができる．しかし，実際のネットワークに存在する複数のサーバに対する RTT からコンテンツ取得時間を推定し，最適なサーバを確実に選択することは困難であることが示されている[14]．

ネットワーク特性を利用する方式として，サーバの数とコンテンツの転送量，転送遅延の関係を求めるため，ノード間の距離に応じた重み付けを利用することが提案されているが[15]，その動的な変動とコンテンツ取得時間の関係については特に検討されていない．

以上から，従来のサーバ選択方式のいずれの方式でも問題となるのは，コンテンツ取得時間を正確に推定する指標が確立されていないことである．

このような問題に対して，筆者らは時間帯によって変動するネットワークの特性を考慮した，統計的な方式によるサーバ選択方式[12] を提案している．この選択方式は，短時間で経路が変化しない環境でユーザが最適ミラーサーバを選択し，数百 k バイトから数十 M バイトのコンテンツを短時間で取得することを目的とした手法で，コンテンツ取得時間の推定パラメータとして，ユーザ/サーバ間のスループットがサーバ選択における妥当な評価指標であるとしている．ここで，ユーザから測定するスループット S は，アプリケーションが転送したコンテンツサイズを B[bit]，転送開始時刻を T_s，転送終了時刻を T_e とするとき，

$$S = \frac{B}{T_e - T_s} \tag{1}$$

と定義している．現在の計算機の性能から，データを受信してからユーザにコンテンツを提示するまでの処理時間は，データの転送時間より十分小さいと考えられ，コンテンツ取得時間はデータの転送時間にほぼ一致するといえる．そのため，データの転送時間に相当する転送終了時刻と転送開始時刻の差である $T_e - T_s$ を含む式 (1) は，コンテンツ取得時間を推定する妥当な指標であると考えられる．文献 12) のサーバ選択方式では，式 (1) で定義されるスループットを用い，選択対象のサーバ間でスループットの逆転が生じる場合を履歴情報を基に自動判別し，高い選択精度で最高のスループットを示すサーバの選択を実現している．しかし，この方式では比較する対象となるすべてのミラーサーバ群に対し，定期的にスループットの測定を行うことから，測定のためにネットワークへ流すトラヒック量（以下，測定コスト）が大きく，改善すべき課題が残されている．

本論文では，文献 12) に示される手法のような測定トラヒックを発生させてスループットを測定する手法の前処理として，少ない測定コストでネットワークに点在するボトルネックを検出し，検出したボトルネックの位置情報からサーバのグループ化を行うことで，測定対象のサーバ数と測定トラヒックの発生を削減する手法を提案する．性能評価実験では，実際のネットワークに対して提案方式を適用し，文献 12) のサーバ選択方式とサーバの選択精度，測定コストの面で比較を行い，提案する手法の有効性を示す．

以下，2 章ではボトルネックを検出し，その位置情報からサーバのグループ化を行い，測定コストを削減したサーバ選択方式案を述べる．3 章では RTT を利用したボトルネック検出方式を提案し，その検出能力を評価する．4 章では提案手法の評価を行う．5 章は結論である．

2. リンクの輻輳状態を考慮したサーバのグループ化

複数のサーバとユーザ間のスループット変動の調査を行う．実際に稼働するサーバまでのスループットを定期的に測定し，その時間変化の特徴を分析する．そのうえで，スループットの測定コストの削減を可能とする，サーバのグループ化を用いたミラーサーバ選択方式を提案する．

2.1 途中経路が等しいサーバ群の特性

ユーザ/サーバ間の途中経路が等しいサーバ群 1（serverA，B，C，D），サーバ群 2（serverE，F，G，H）のスループット変動を**図 1** に示す．サーバ群 1 は RING サーバ（ring.htcn.ne.jp，ring.omp.ad.jp，ring.shibaura-it.ac.jp，so-net.ne.jp），サーバ群 2 は netscape の FTP サーバ（ftp-assg.netscape.com，ftp-au.netscape.com，ftp-eude.netscape.com，ftp-mide.netscape.com）である．ユーザからサーバまでの途中経路は traceroute コマンドで測定し，既知である．測定期間を 24 時間，測定間隔を 15 分間とし，serverA〜H まで定期的に 1 Mbyte のコンテンツを転送し，式 (1) からスループットを算出する．図 1 のサーバ群 1 までのスループットを左縦軸とし，サーバ群 2 までのスループット変動を右縦軸とする．

図 1 からサーバ群 1，2 に，それぞれにスループットが同程度まで低下する時間帯が存在するのが確認で

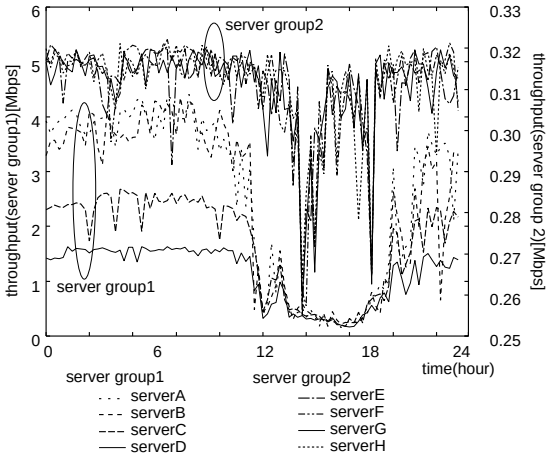


図 1 途中経路が等しいサーバ群のスループット変動
Fig. 1 The change of throughput to servers.

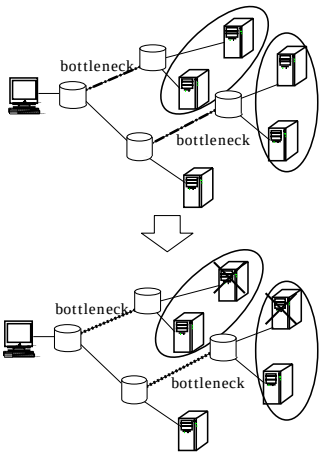


図 2 ボトルネック検出によるサーバのグループ化
Fig. 2 Grouping servers by detection of bottleneck.

きる．

この原因は，ネットワークにおける輻輳によるものであると考えられる．複数のルータにおいて同時刻に輻輳が生じ，同程度までスループットが低下するとは考えにくい．また，サーバ側に関しても同様に，複数のサーバにおいて同時刻に，同程度までスループットが低下するような負荷の増加が生じることは非現実的であるといえる．つまり，ユーザ/サーバ間の経路中の 1 つのルータにおける輻輳が原因で，スループットの低下が生じていると推測できる．本論文では，サーバ群とユーザ間のスループットを同時刻において同程度まで低下させる原因のルータが 1 つであると仮定し，そのルータを経由するリンクをボトルネックと呼ぶこととする．サーバ選択において，各時刻でボトルネックの検出を行えば，その位置情報からスループットが同じ程度となるサーバ群とその時間帯の抽出が可能となると考えられる．

2.2 サーバのグループ化による測定コストを削減したサーバ選択手法

2.1 節の解析から，ユーザ/サーバ間のスループットは，途中経路でボトルネックとなる同一のルータにより制限されることが明らかとなった．このボトルネックを検出することにより，スループットの測定対象となるサーバをグループ化し，測定コストを削減したサーバ選択方式を提案する．

グループ化されたサーバ群は，同程度のスループットとなるため，そのグループを代表するサーバを選択し，そのサーバに対してのみスループットを測定することで，グループ化されたサーバ群のスループットの推定が可能となる．ボトルネックの位置情報からグループ化されたサーバは，代表サーバを除きスループット測定対象から削除する（**図 2**）．つまり，提案手法はスループット測定のための測定トラヒック量自体を削減するのではなく，サーバのグループ化により，スループットを測定するサーバ数を削減し，従来手法と比べ測定コストを削減したサーバ選択を実現するものといえる．

3. RTT を利用したボトルネック検出

2.2 節で述べたグループ化を実現するためには，高精度にボトルネックを検出する手法が必要である．通常，途中経路のルータで輻輳が生じている場合，ルータ内のキューにパケットが複数存在し，パケットの伝送遅延が通常より増加すると考えられる．その伝送遅延の通常からの増加量を定量的に評価できる指標が必要である．また，複数のサーバまでの複数の経路のどれにボトルネックが生じているかを判断するために，ボトルネック検出のための指標は，異なる経路でも共通の判断基準が設定できるように正規化される必要性もある．

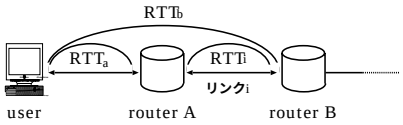
3.1 ボトルネック指数

通常からの RTT の増加量を表すことと，正規化されていることの条件を満たす指標として，ボトルネック指数を提案する．

ボトルネック指数 BI を以下に示す．

$$BI = \frac{RTT_i - \overline{RTT}_i}{RTT_{all}} \tag{2}$$

ここで，RTT_i はユーザ/サーバ間で経由する任意のリンク i の RTT，$\overline{RTT}_i$ は過去 N 時間の RTT_i の平均でリンク i の通常の RTT を表す．RTT_{all} は BI



$RTT_i = RTT_b - RTT_a$

図 3 RTT_i の測定法
Fig. 3 The measurement method of RTT_i.

を算出する時点でのユーザ/サーバ間の経路上の各リンクの RTT の合計で，各リンクの RTT の通常からの増加具合を正規化する項である．この正規化項は，各リンクの RTT_i が，サーバまでの経路の RTT のうち何割を占めるかを表すために用いている．これにより，BI の最大値が 1 になるように正規化され，異なる経路でも同一の基準で比較が可能になる．RTT の測定に関しては，ICMP の echo パケットを利用するため，RTT の観測に用いられるパケットサイズは 64 bytes となる．また，経路の途中でパケットがロスした場合は，測定結果として利用者が設定したタイムアウト値を測定結果として用いる．4 章の実験では，クライアント/サーバ間の RTT が 1 秒を超えるものがなかったため，タイムアウト値を 1 秒とした．ping コマンドにより RTT を測定する間隔は，観測対象となるネットワークの特性により変化すると考えられるが，本論文では，文献 12) に従い 15 分とした．

RTT_i はユーザからの ping コマンドにより測定可能な RTT_a，RTT_b の差分から容易に求めることができる（**図 3**）．文献 16)，17) ではバックボーンを流れるトラヒックの変化には 24 時間の周期性を持つ場合があることが示されている．したがって，リンク i の通常の RTT を表す平均 RTT（$\overline{RTT}_i$）は，過去 1 日分（$N = 24$）の RTT から求めることとする．

3.2 ボトルネックの検出能力の評価

3.2.1 ボトルネック指数とスループットの関係

ボトルネック指数を用いて，実際のネットワークに存在するボトルネックの検出能力の評価を行う．

測定対象を世界 56 カ国に存在する任意のサーバ群 295 個とし，測定期間を 48 時間，測定間隔を 15 分とする．測定対象サーバに対し，prtraceroute[18] コマンドで途中経路を調査し，サーバまでの途中経路が調査可能なサーバ群に対して，途中に経由するルータとサーバまでの RTT とユーザ/サーバ間のスループットを測定する．

prtraceroute コマンドによって途中経路が調査可能であったサーバ数を**表 1** に示す．

表 1 から，約 63%のサーバが途中経路が調査可能であることが分かる．

表 1 途中経路の調査
Table 1 The investigation of the path.

total	success	fail
295	186	109

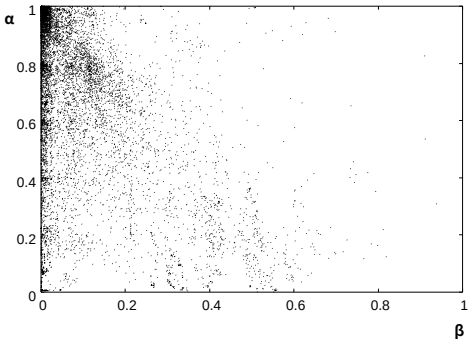


図 4 α と β の関係
Fig. 4 The relation of α and β.

途中経路が調査可能であった 186 個のサーバ中には anonymous でログインが拒否されたり，サーバにコネクションが張れないことにより，スループットの測定が困難なサーバが存在し，そのようなサーバを除いた 151 個のサーバに対してスループットの測定が可能であり，151 個のサーバを対象として実験を行った．このサーバ数は，295 個のサーバの約半数しか選択対象にならないことになるが，151 個のサーバが同様のコンテンツを保持するミラーサーバであるため，短時間でのコンテンツ取得を目的とした場合，実運用上の大きな問題にはならないと考えられる．

また，ボトルネック指数を利用したボトルネックの検出能力を評価する際に以下の評価指標 α，β を用いる．

$$\alpha = \frac{S_t}{S_{max}} \tag{3}$$

$$\beta = (BI_{tmax}) \tag{4}$$

ここで，S_t は時刻 t におけるスループットで，式 (1) で定義されるものであり，実験では 1 M バイトのファイルを転送した際の値である．S_{max} は測定期間中で最高のスループット，BI_{tmax} は各計測時刻 t におけるユーザ/サーバ間の経由するリンクの中で最大のボトルネック指数である．

ボトルネック指数がスループットの低下を推定できている場合，β が増加するにともない α が低下することになる．α に対する β の値をプロットしたグラフを**図 4** に示す．

図 4 から，β が大きくなるほど，α の値が小さい範

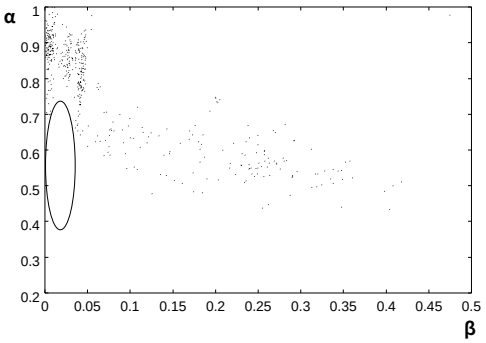


図 5 サーバに負荷がない場合
Fig. 5 No server overhead.

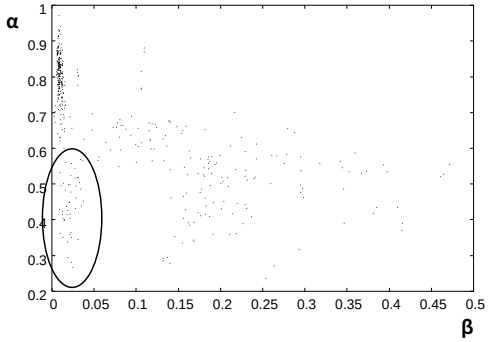


図 6 サーバに負荷がある場合
Fig. 6 Any server overhead.

囲に分布しているのが分かる．α と β の相関係数は -0.601 となり，比較的強い相関を持っていることから，ボトルネック指数が大きいリンクがサーバまでの途中経路に存在する場合に，そのサーバまでのスループットは通常より低下している傾向があるといえる．したがって，ボトルネック指数を利用することで，複数のサーバに対し共通の判断基準を設定し，ボトルネック検出が可能であるといえる．

3.2.2 サーバの負荷のスループットへの影響

図 4 において，α と β がともに小さい値を持つ点が複数存在する．これらの点は，ボトルネック指数の観点からはボトルネックは存在しないが，コンテンツを転送して観測されるスループットが低いことを表している．ボトルネック指数は RTT のみを用いて算出されるため，サーバが高負荷の状態でコンテンツの送出性能が低下していることは，ボトルネック指数の値に反映されることはない．そのため，スループットが低下してもボトルネック指数が増加せず，α と β がともに小さい値を持つ点が存在してしまうと考えられる．

ボトルネック指数とスループットの関係に対するサーバの負荷の影響を検証するため，日本国内に負荷が掛かっていない状態の web サーバを構築し，アメリカのカリフォルニア州からサーバまでの途中経路で通過する各ノードまでの RTT，サーバまでのスループットを定期的に測定する実験を行う．測定期間は 48 時間であり，測定間隔は 3 分間である．

サーバが無負荷の場合の α と β 関係の実験結果を**図 5** に示し，一時的にサーバの CPU 負荷を増大させた場合の α と β の関係を**図 6** に示す．

図 5，6 からサーバに負荷がない環境では α と β が小さい範囲には点がないが，サーバを高負荷状態にした場合には，α と β がともに小さい範囲に点が存在するのが確認できる．したがって，図 4 の α と β のともに小さい範囲に存在する点は，サーバの負荷によりコンテンツの転送性能が低下したことを表しているといえる．

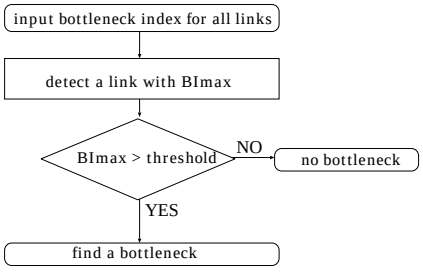


図 7 ボトルネック検出のフローチャート
Fig. 7 The flowchart of the detection of bottleneck.

3.3 ボトルネック検出アルゴリズム

ボトルネック指数を利用したボトルネック検出のフローチャートを**図 7** に示す．

まず，RTT の測定からユーザとサーバとの間に経由する各リンクのボトルネック指数を算出する．次にユーザと比較の対象となる各サーバ間において最大のボトルネック指数をとるリンクを検出する．その最大のボトルネック指数（BI_{max}）が閾値以上の場合は，最大のボトルネック指数をとるリンクがユーザとサーバ間におけるボトルネックと判断する．閾値以下の場合はユーザとサーバ間にはボトルネックはないと判断する．

4. サーバのグループ化を用いたサーバ選択方式の実装と評価

4.1 サーバのグループ化アルゴリズム

図 8 にサーバのグループ化のフローチャートを示す．

まず，ユーザとサーバの間でボトルネックと判断されたリンクを入力する．ユーザと複数のサーバの間で同じリンクがボトルネックと判断された場合にそのリンクを途中経路におけるボトルネックとして持つサーバ群をグループ化する．次に，グループ化されたサー

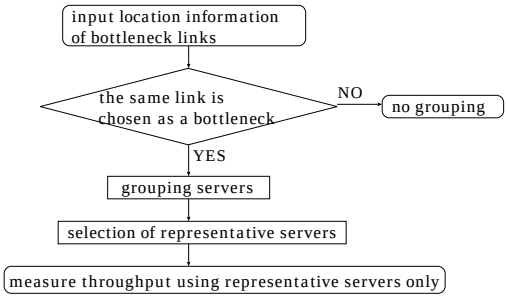


図 8 サーバのグループ化のフローチャート
Fig. 8 The flowchart of the grouping servers.

バ群の中でスループットを測定する代表サーバを選択し，そのサーバ以外のグループ化されたサーバはスループットを測定する対象のサーバから除外する．これにより，ボトルネック指数から同程度のスループットとなるサーバが測定対象から外され，測定コストの削減が行われる．同じリンクが他のサーバまでの途中経路におけるボトルネックと判断されなかったサーバはグループ化は行わず，スループットを測定する対象のサーバとなる．

4.2 測定対象サーバの選択法

ボトルネックの位置情報に基づきグループ化されたサーバ群の中でも途中経路の最後のリンクが物理的に狭い帯域のサーバの場合，その影響で常時スループットが低くなってしまう．したがって，そのようなサーバは，代表してスループットを測定するサーバとしては不適切である．そのようなサーバを考慮した，グループ化されたサーバ群中の測定対象サーバの選択方法は，文献 12) でのサーバ選択基準に従って以下のように行う．

サーバ i，j がグループ化されたとする．グループ化された時刻でサーバ i，j に関して取得された過去 1 時間以内のスループットの平均 m_i，m_j と標準偏差 d_i，d_j を求める．過去のスループットがサーバのグループ化により削除されていた場合は，その値を日付が異なるが同じ時間帯に観測されたスループットの中で最近の値を用いて補間する．平均の関係が $m_i > m_j$ のとき，

$$(m_i - d_i) > (m_j + d_j) \quad (5)$$

が成立すれば，平均が高いサーバ i を選択する．式 (5) が成立しない場合は，サーバ i と j のうち，ユーザとサーバの間の Hop 数が最も少ないサーバを選択する．最小 Hop 数のサーバを選択することにより，よりネットワークへ流す測定のためのトラヒックの影響が軽減される．

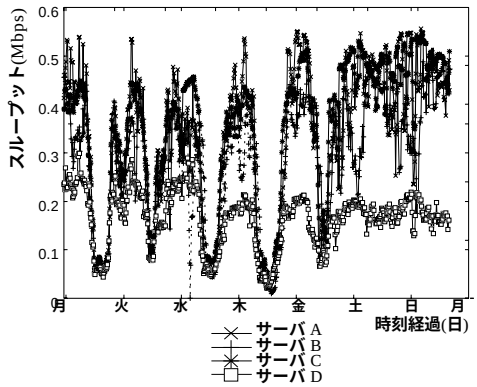


図 9 スループットの時間変動
Fig. 9 Time jitter of throughput to servers.

4.3 実環境におけるサーバ選択実験

提案方式のサーバ選択精度，測定トラヒックの削減能力の検証のための運用実験を行う．対象とするサーバは，同じコンテンツを保持し異なる 13 のネットワーク上に存在する 13 個のサーバとする．実験期間を 2002 年 6 月 16 日から 23 日までの連続した 7 日間（月曜日〜日曜日）とし，測定間隔を 15 分で計 504 回のサーバ選択を行った．1 回の測定ごとに以下の処理を行う．測定対象のサーバまでに経由する各ノードまでの RTT を測定し，ボトルネック指数を導出する．導出したボトルネック指数からボトルネックを検出し，サーバのグループ化を行う．グループ化によって削除されなかったサーバ群に対して，1 Mbyte のコンテンツを転送してスループットを測定し，サーバ選択を行う．サーバ選択方式は文献 12)（以下，従来方式とする）と同様である．提案方式では，サーバのグループ化により過去のスループットデータが削除されている場合がある．その場合には，過去に同じ時間帯に観測された中で最近のスループットの値で補間し，従来方式のサーバ選択アルゴリズムを駆動する．この処理により，選択精度に差が生じると考えられるが，図 1 で用いたサーバ A〜D の 1 週間の変動を示す**図 9** から，同時刻の場合，同等のスループットが得られることが分かり，極端な精度低下をともなわずに測定コストは削減されることが期待できる．選択精度と測定コストの両面において，従来方式と比較し，その効果を検証する．

4.3.1 選択精度の比較

図 10 にボトルネックと判断する閾値に対する平均スループットを示す．

図 10 から提案方式はボトルネックと判断する閾値を低く設定するほど，つまり，ボトルネックと判

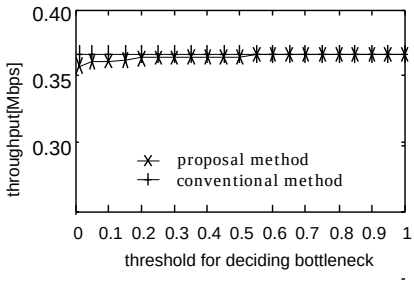


図 10 選択指標
Fig. 10 An index of selection.

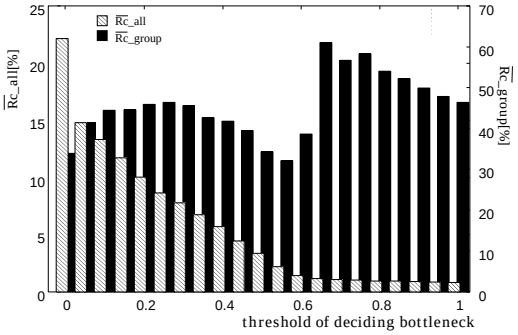


図 11 測定コスト削減率
Fig. 11 Measurement cost reduction rate.

断する感度を上げグループ化を行う頻度を増やすほど，選択精度が低下しており，その差は，最大で 9 Kbps となっているのが確認できる．実際のコンテンツ取得時間は，文献 12) の従来方式において，1 [Mbytes]/366423.35 [bps] = 22.89 [sec] であり，提案方式おいては，1 [Mbytes]/357773.43 = 23.44[sec] である．コンテンツ取得時間の差は 0.55 sec であり，2.4%の増加となり，選択精度の低下は非常に小さいことが分かる．また，まったくサーバ選択を行わなず特定のサーバのみを用いてスループットを測定した場合，コンテンツ取得時間は最小で 28.3%，最大で 325.6%の増加となり，サーバ選択の有効性も確認できた．

4.3.2 測定コストの比較

ある 1 つのサーバ i のスループットを 1 回測定するための測定コスト $cost_i$ を

$cost_i$ =(ユーザ/サーバ i 間の Hop 数)
×(測定用コンテンツサイズ)

として定義する．測定のためにネットワークへ流したコンテンツサイズとユーザ/サーバ間の Hop 数の積をとるのは，測定によるネットワークへの負荷もコストとして考慮するためである．比較には，次式で定義されるコスト削減率を用いる．

$$Rc = \frac{\sum_{i \in prop} cost_i}{\sum_{all} cost_i} \qquad (6)$$

式 (6) の分母はすべての測定対象サーバの測定コストの総和で，従来手法によって 1 回のサーバ選択が行われる際に必要な測定コストである．また，分子は提案手法によって測定対象と判断されたサーバに対する測定コストの和である．サーバのグループ化が行われた場合，測定対象となるサーバ数が減少するため Rc は 1 未満となる．しかし，ボトルネックが検出されず，サーバのグループ化が行われなかった場合，Rc の値は 1 となる．

ボトルネックと判断する閾値に対する測定コスト削減率を**図 11** に示す．図 11 には，504 回の測定すべての Rc の平均 $\overline{Rc_{all}}$ と，グループ化が行われた場合のみ，つまり，$Rc < 1$ を満たす Rc の平均 $\overline{Rc_{group}}$ が示されている．$\overline{Rc_{all}}$ は，ネットワークにボトルネックが存在せず，提案手法による測定コスト削減が行われなかった場合も含めた評価で，$\overline{Rc_{group}}$ は，ネットワークにボトルネックが存在し，測定コスト削減が行われた場合の評価ということになる．

図 11 からボトルネックと判断する閾値を低く設定するほど，ボトルネックを高感度に検出し，大幅な測定コストの削減が可能であることが分かる．また，$\overline{Rc_{group}}$ の変動から，ネットワークにボトルネックが存在し，サーバがグループ化された場合には大幅なサーバの削除が行われ，提案手法の測定コスト削減に対する効果が大きいことを確認できる．

4.4 ボトルネック検出のための測定トラヒックの影響

ボトルネックを検出するために RTT を測定する際の測定トラヒック量を検討する．今回の実験では 64 byte の ICMP echo [19] パケットを 43 個のノードに 15 分ごとに送信し，RTT を測定した．したがって，ボトルネックを検出するため，24 時間に発生する測定トラヒックの総量は 0.246 Mbyte である．一方，スループットを測定するために各サーバから 1 Mbyte のコンテンツを 15 分ごとに行うと，24 時間に発生する測定トラヒックの総量は 1,248 Mbyte である．したがって，1 日にスループットを測定するためにに必要な総トラヒック量に対する，ボトルネックを検出するために必要な総トラヒック量の比は

$$\frac{0.246\,\mathrm{Mbyte}}{1248\,\mathrm{Mbyte}} = 1.97 \times 10^{-4}$$

となり，およそ 0.0002%となる．したがって，ボトルネックを検出するための測定コストはほぼ無視できるといえる．

4.5 考　　察

本論文では，測定トラヒックを発生させスループットを計る手法の前処理として，ICMP の echo パケットを用い算出した RTT により，低コストな測定でボトルネックを検出し，選択対象となるサーバのグループ化を行うことで測定トラヒック量の削減を実現した．文献 12) の手法と比較して，選択精度，測定コスト削減に関し良好な結果が得られた．本手法は，文献 12) の手法のみでなく，netperf[20] や ttcp[21] などの測定トラヒックを発生させスループットを計る手法の測定対象を削減する前処理として利用可能であると考えられる．しかし，RTT の測定間隔やグループ化を行った後のスループット測定に用いるデータサイズの決定方法などの課題が残っている．これらは，観測対象とするネットワークや利用者の要求精度によって変化すると考えられ，今後の課題といえる．

また，TCP を使いコンテンツの取得を行う利用方法に対し，ICMP によってボトルネックを評価している．TCP においてスループットが低下する要因として遅延とパケットロスがある．本手法において，遅延は RTT で評価し，パケットロスはタイムアウト値を RTT として用いることで実現している．タイムアウト値を RTT として用いることで提案しているボトルネック指数が増大し，パケットロスによりスループット低下も評価可能となっているが，タイムアウト値の設定方法についても今後の課題となると考えられる．

本手法では，1 つのルータが原因でボトルネックが発生していることを仮定し，手法の構築を行っている．ボトルネックの存在を仮定せず，すべてのサーバに対しスループットの測定を行う手法と比較して，選択精度の低下が小さいという実験結果から，1 つのルータが原因でボトルネックが発生している仮定は妥当であるものと考えられるが，インターネットにおいて一般的に仮定が成り立つかの検証も必要であると考えられる．

5. む　す　び

コンテンツ取得時間短縮のために，負荷分散技術として，コンテンツを複数のサーバに分散配置するサーバのミラー方式が広く利用されている．それらのサーバ群の中から最適なサーバを選択する方式として，選択の指標にスループットを用いることで高精度にサーバの選択が実現可能となっている．しかし，時々刻々変化するネットワークの状態を選択の判断基準に反映するため，定期的にスループットを測定しなければならず，ネットワークに対し大きな測定負荷を与える問題があった．

本論文では，RTT の通常からの増加具合から高精度にボトルネック検出可能なボトルネック指数を提案し，検出したボトルネックの位置情報を基に，スループットが同程度まで低下しているサーバをグループ化する手法を提案した．ボトルネックを基準としたサーバグループ化により測定対象のサーバ数削減が可能となり，サーバ選択システムとして測定コストの削減を実現した．実運用ネットワークにおいて，グループ化を行わない場合との，選択精度，測定コストの変化を調査し，同程度の選択精度と測定コストの大幅な削減が可能であることを示し，提案手法の有効性を確認した．

参　考　文　献

1) http://www.linux.org/
2) Akamai. Akamai content delivery network. http://www.akamai.com
3) Digital Island. http://www.digitalisland.com
4) Mirror Image. http://www.mirror-image.com
5) Brisco, T.: DNS Support for Load Balancing, RFC 1794 (Apr. 1995).
6) Foundry Networks, Inc.: ServerIron XL. available at http:www.foundrynet.com/datasheets/serverironspec.html
7) Rekhter, Y. and Li, T.: A Border Gateway Protocol 4 (BGP-4), RFC1771 (Mar. 1995).
8) Hawkinson, J. and Bates, T.: Guidelines for creation, selection and registration of an Autonomous System (AS), RFC1930 (Mar. 1996).
9) Kamiya, H., Ohta, K., Kato, N., Mansfield, G. and Nemoto, Y.: An Improved Content Search Engine — Usage of Network Configuratin Informatin, *Proc. IEEE TENCON '98*, Vol.1, pp.21–24, New Delhi, India (Dec. 1998).
10) 廣津登志夫，高田敏弘，栗原　聡，菅原俊治：転送履歴を利用した URL Resolver における最適経路選択，第 1 回インターネットテクノロジーワークショップ（WIT'98）(Aug. 1998).
11) Cisco Systems, Inc.: DistributedDirector. available at http://www.cisco.com/warp/public/cc/cisco/mkt/scale/distr/index.html
12) 真壁　知，太田耕平，加藤　寧，Mansfield, G.，根元義章：ネットワークの負荷変動を考慮した動的なミラーサーバ選択方式，電子情報通信学会論文誌 B，Vol.J84-B, No.3, pp.435–442 (Mar. 2001).
13) Jacobson, V.: traceroute. available at ftp://ftp.ee.lbl.gov/traceroute.tar.Z
14) Johnson, K.L., Carr, J.F., Day, M.S. and Kaashoek, M.F.: The Measured Performance of

Content Distribution Networks, *Proc. 5th International Web Caching and Content Delivery Workshop*, Lisbon, Portugal (May 2000). available at http://www.terena.nl/conf/wcw/Proceedings/
15) 中庭明子，高橋 潤，榎原博之，岡田博美：信頼性を考慮した分散ミラーサーバ配置最適化モデル，信学技報，IN2000-9, pp.7–12 (Apr. 2000).
16) Thompson, K., Miller, G.J. and Wilder, R.: Wide-Area Internet Traffic Patterns and Characteristics, *IEEE Network*, Vol.11, No.6, pp.10–23 (Nov./Dec. 1997).
17) 川原亮一，石橋圭介，平野聡之，斎藤 洋，大原久樹，佐藤大輔：インターネットバックボーントラヒック測定分析，信学技報，IN99-44, pp.13–20 (Sep. 1999).
18) Prtraceroute. available at http://www.ripe.net/ripencc/pub-services/db/irrtoolset/prtraceroute/
19) Postel, J.: Internet Control Message Protocol, RFC 792 (Sep. 1981).
20) http://www.netperf.org/netperf/NetperfPage.html
21) http://www.dtic.mil/ttcp/

(平成 15 年 4 月 28 日受付)
(平成 15 年 10 月 16 日採録)

和泉 勇治

平成 8 年東北大学工学部情報工学科卒業．平成 10 年同大学大学院情報科学研究科修士課程修了．平成 13 年同研究科博士課程修了．同年同大学助手，平成 15 年同講師，現在に至る．ニューラルネットワーク，文字認識，コンピュータネットワークの構築，管理等の研究に従事．電子情報通信学会会員．

宇津江康太

平成 13 年東北大学工学部情報工学科卒業．平成 15 年同大学大学院情報科学研究科修士課程修了．在学中ネットワーク管理の研究に従事．現在，株式会社日立東日本ソリューションズ．

加藤 寧（正会員）

昭和 63 年東北大学大学院修士課程修了．平成 3 年同大学大学院後期課程修了．同年同大学大型計算機センター助手，平成 7 年同大学院情報科学研究科助手，平成 8 年同助教授，平成 15 年同教授，現在に至る．工学博士．コンピュータネットワークの構築，管理，文字認識，ニューラルネットワーク等の研究に従事．電子情報通信学会会員，IEEE 各会員．

根元 義章（正会員）

昭和 43 年東北大学工学部通信工学科卒業．昭和 48 年同大学大学院博士課程修了．同年同大学助手，昭和 59 年同大学電気通信研究所助教授，平成 3 年同大学大型計算機センター教授，平成 7 同大学大学院情報科学研究科教授，平成 10 年より同大学大型計算機センター長併任．平成 13 年より同大学情報シナジーセンター長併任．工学博士．マイクロ波伝送路回路，衛星利用ネットワーク，情報伝送システム，画像処理，文字認識等の研究に従事．昭和 56 年 IEEE・MTT・Micro Wave Prize 受賞．電子情報通信学会会員，IEEE 各会員．