

転送速度の向上と負荷分散を両立する 広域サーバ負荷分散方式の提案と評価

佐々木 健[†] 敷田 幹文[‡]

[†] 北陸先端科学技術大学院大学情報科学研究科
[‡] 北陸先端科学技術大学院大学情報科学センター
^{†‡} 〒 923-1292 石川県能美市旭台 1-1
E-mail: ^{†‡}{stakeshi, shikida}@jaist.ac.jp

あらまし 現在、インターネットにおいてサーバの負荷分散を目的とするミラーサーバ方式の使用が一般的となっている。また、リスク回避やネットワークの負荷分散を目的とする広域サーバ負荷分散方式が利用されつつある。しかし、広域分散配置下で、転送効率の向上と負荷分散を同時に実現するのは困難である。本論文では、転送速度の向上と負荷分散を両立する広域サーバ負荷分散方式を提案する。線形計画法を用いてクライアントの接続要求を複数同時判断することにより、ボトルネックリンクを回避しつつ、各配信サーバの負荷を分散する。そして、シミュレーションによって提案方式の評価を行い、有効性を示す。

キーワード ミラーサーバ、広域サーバ負荷分散、線形計画法

Proposal and Evaluation of Global Server Load Balancing Method Compatible with both Improvement of Throughput and Load Balancing

Takeshi SASAKI[†] and Mikifumi SHIKIDA[‡]

[†] School of Information Science, Japan Advanced Institute of Science and Technology
[‡] Center for Information Science, Japan Advanced Institute of Science and Technology
^{†‡} 1-1 Asahidai, Nomi-shi, Ishikawa, 923-1292 Japan
E-mail: ^{†‡}{stakeshi, shikida}@jaist.ac.jp

Abstract

Mirror server method is used for the purpose of server load balancing in the Internet now. Global server load balancing is used for the purpose of risk avoidance and network load balancing too. However, it is difficult to realize improvement of throughput and load balancing in wide area distributed placement. In this paper, we propose global server load balancing method compatible with both improvement of throughput and load balancing. This method manages plural connections from clients by linear programming, so it avoids a bottleneck link and server load balancing. And we evaluate proposed method by simulation and show the effectiveness.

Keywords Mirror server, Global server load balancing, Linear programming

1 はじめに

現在、インターネットにおいて、サーバの負荷を分散させるミラーサーバ方式の使用が一般的となっている。これは、多数の接続要求を受ける Web サーバでは標準的に用いられている。これらは、データセンターなどの一箇所に集中配置するミラーサーバである。また、各機関がサーバを提供しあい、大き

なサービスを提供する Ring-Server や SourceForge などが存在する。これらは広域分散配置型ミラーサーバに分類される。そのため、広域分散配置型はボランティア型サービスに用いられることが多かった。しかし、近年、サーバ設置地域の災害、インフラの故障などサーバ集中配置によって多大な影響を受けるリスクを回避する需要が生まれてき

た。また、大規模コンテンツを大量に配信することにより、配信サーバ近隣ネットワークの負荷圧迫が問題になっている。よって、ネットワークの負荷分散にも今後広域分散配置型ミラーサーバが必要になり、そのため、商業利用の場でも広域分散配置型ミラーサーバの使用が一般的となっていくと考えられる。

しかし、この方式を商業利用すると課題が存在する。それは、ミラーサーバの本来の目的である負荷分散と同時にサービス利用者への高いサービス品質の提供することの困難さである。広域分散配置型では、地域ごとの環境格差により各サーバごとにサービス品質の格差が生じやすい。商業利用では、負荷分散と同時に、クライアントに高いサービス品質を提供することが重要となる。

そこで本論文では、大規模コンテンツ配信サービスにおいて、クライアントの利益となる転送速度の向上とサーバ管理者への利益となるシステムの分散利利用を両立することを目的とし、転送速度の向上とサーバの負荷分散を両立する広域サーバ負荷分散方式の提案を行う。

本論文では、2章において広域分散配置型ミラーサーバにおいて重要な問題である接続サーバの決定における従来方式の特徴と問題点に触れ、3章では提案方式について述べ、4章でシミュレーションによる評価と議論を行う。

2 広域分散配置型ミラーサーバにおける接続サーバの決定方式

従来の広域分散配置型ミラーサーバにおける接続サーバの決定方式について、方式の特徴と問題点を述べる。

2.1 転送効率の向上を目指した方式

転送効率を指標にしてクライアントが接続すべきサーバを決定し、クライアントにとって最もよいサーバを選択を行う。

文献 [1, 2] では、RTT(Round Trip Time) によるサーバ評価を行い、応答性によって接続するサーバを選択している。これは、Web コンテンツなどの小規模かつ多数のコンテンツの転送を必要なものには有効であるが、大規模コンテンツにおいては、応答性より転送速度の方が重要となる。文献 [3, 4] では、サンプルコンテンツ転送による各ミラーサーバの通信帯域の測定を行い、最も転送速度の速いサーバを決定している。この方式は大規模コンテンツ転送においてユーザ視点で考えた場合、重要な指標である転送速度の向上を目指している。

しかし、サーバの性能や、サーバ近隣ネットワークの性能差により、一部の効率の高いサーバだけ

にクライアントの接続が集中してしまう可能性がある。そのため、商業利用においての広域分散配置型ミラーサーバでは、多数の接続が発生するため、多大な接続の一転集中によりサーバがダウンする可能性がある。また、広域分散配置型の利点である、サーバ近隣ネットワークの負荷分散性などが失われてしまう。よって、負荷分散の観点では問題がある。

2.2 サーバの負荷分散を重視した方式

各ミラーサーバの負荷の分散を行い、ミラーサーバの本来の目的である負荷分散に注力し、サービスの高い可用性を実現する。

文献 [5] では、接続要求の到着率を基にした接続配分比率の算出をし、以後の接続要求に対して確率的な接続制御を行い、負荷分散を行う。それに対して、文献 [6] では、予測値を使用せず、各サーバの負荷格差を補完する確定的な判断を行い、各サーバの負荷の推移の完全なる制御を行う。これにより、突発的な負荷の上昇を防ぎ、サーバの処理性能の瞬間的な超過を回避してセッションの棄却率の低減を実現している。

しかし、負荷分散の観点から選択されたサーバがクライアントにとって有益とは限らない。各サーバにおいて、転送効率の格差がある場合、転送速度の低下によってユーザに多大な不満を与えてしまう可能性がある。

2.3 転送効率、負荷分散を相互判断する方式

各サーバを転送効率、負荷分散のそれぞれの観点によって評価を行い、評価結果を相互判断し、ユーザとサーバ管理者の利点の両立を行う。

文献 [7] では、BGP(Border Gateway Protocol) における ASpath(Autonomous System) によって、サーバとクライアントの距離を求め、クライアントをネットワーク的に近く、同時に各サーバの負荷が分散するように接続判断を行う方式である。クライアントを近くのサーバへ誘導することにより、縦横無尽な長距離転送を避け、ネットワークの負荷分散も実現する。また、ネットワークの近さだけでなく、転送速度や応答速度などを考慮して、転送効率の向上を目指している。

転送効率と負荷分散の相互判断処理部では、まず転送効率における評価を行い、次に負荷分散による評価を行う 2 ステップの判断アルゴリズムにより接続判断を行っている。

しかし、各サーバの転送効率からの評価と、負荷分散からの評価が逆になってしまった場合などでは妥協した結果となる。あるサーバは転送効率

が高いが、負荷が集中している場合、転送速度が負荷分散がどちらかを妥協させる判断を行わなければならない。その場合は両者の利点を同時に実現するのは難しい。この矛盾は多々起こる可能性がある。

3 転送速度の向上と負荷分散を両立する広域サーバ負荷分散方式

本論文において提案する広域サーバ負荷分散方式は、転送速度の向上と負荷分散を両立する接続制御を行う。ミラーサーバ配信環境下に、制御サーバを設置し、制御サーバが判断指標値を収集し、線形計画法による確定的な接続判断をする。クライアントをAS単位でグループ化し、AS単位で判断指標値の共有を行うことで、調査コストを削減する。クライアントからの接続要求を一定時間プールし、複数の接続要求を同時に判断することによって、ボトルネックリンクを回避して転送速度の低下を防ぎ、各サーバの負荷分散を行う。システム構成は、制御サーバと、各配信サーバへの通信ログ排出ソフトウェア、各クライアントへの専用ダウンロードソフトウェアの導入を想定している。

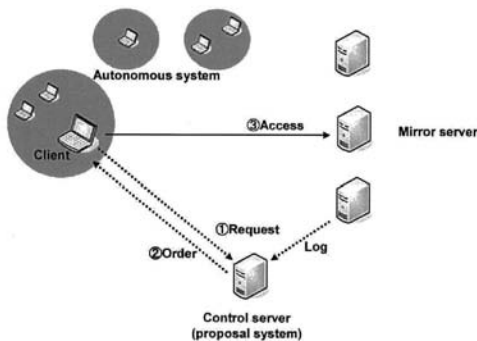


図 1: システム概要図

3.1 判断指標値の調査

本方式が行う線形計画法では、クライアントが所属するASから各ミラーサーバまでの転送速度と、各ミラーサーバの接続負荷数を用いる。そのため、制御サーバは、この指標値の調査を行う。転送速度は、以下の二種類の方法により把握する。

- 各サーバに存在するログ情報からの算出
- クライアントからのサンプルコンテンツ転送による測定

転送速度は、各ミラーサーバが持つ通信履歴ログによる算出によって把握する。しかし、過去に接続を行っていない、もしくは長期間接続を行っていないASからの転送速度は把握できない。その場

合、クライアントにサンプルコンテンツ転送による測定をさせ、その結果を制御サーバが収集することで把握を行う。これらの制御は、専用ダウンロードソフトウェアによって自動的に行う。このコンテンツサイズは文献[3,4]より500KBとする。測定結果はAS単位で共有化するため、調査回数を抑えられ、クライアント数の増加に伴う測定頻度の増加は発生しないので、調査コストの増加は抑えられる。

次に、接続負荷数は、制御サーバが判断した接続配分値を記録して、各サーバの負荷数の把握を行う。しかし、この記録は実際の各サーバの負荷数とずれる可能性がある。クライアントが転送途中に切断した場合、または、制御サーバを並列に複数設置した場合である。そのため、制御サーバは、転送速度判断に用いるサーバのログ情報を用いて、実際の負荷数の把握を行い、記録値の調整を行う。よって、ミラーサーバからの通信ログの取得は、頻繁に行う必要がある。

3.2 線形計画法による接続配分値の算出

線形計画法により複数の接続要求を同時に判断し、ボトルネックリンクを回避して負荷の分散を行う。そのため、ボトルネックを検知する必要がある。これは各ASからすべてのミラーサーバに対する転送速度の比較を行い、最大転送速度を持つミラーサーバに対して、他のミラーサーバの速度減少率を比較をすることによって、ボトルネックリンクを経由している配信サーバの経路を発見する。

まず、すべての AS_j からミラーサーバ i までの通信経路評価値 s_{ij} を算出する。これは対象サーバ i の経路速度 T_{ij} を AS_j が取りうる最大の速度 T_{jmax} で割り、速度の減少を求める。

$$s_{ij} = \frac{T_{ij}}{T_{jmax}}$$

線形計画法により、 AS_j からの接続数 A_j を分配し、ミラーサーバ i に割り当てる接続配分値 x_{ij} と、評価値 s_{ij} の積が最大となるよう配分値を算出する。この目的関数は、評価値が高い経路が選択されるにつれて値が上昇するので、目的関数の増加と共に全体の転送速度が上昇する。

線形計画法の算出式は以下の通りである。ただし、 L_{total} はサービス全体の接続負荷数、 L_i はミラーサーバ i の接続負荷数、 P_i はサーバの性能格差を考慮した負荷配分比率、 α は負荷の許容誤差率とする。この負荷配分比率と許容誤差率は管理者が静的に決める必要がある。

$$\text{目標関数: } \sum_{i=1}^n \sum_{j=1}^m s_{ij} x_{ij} \rightarrow \text{最大}$$

制約条件:

$$\begin{aligned}
 L_{1min} &\ll x_{11} + \dots + x_{1n} + L_1 \ll L_{1max} \\
 L_{2min} &\ll x_{21} + \dots + x_{2n} + L_2 \ll L_{2max} \\
 &\vdots \\
 &\vdots \\
 L_{min} &\ll x_{m1} + \dots + x_{mn} + L_1 \ll L_{max} \\
 x_{11} + x_{21} + \dots + x_{m1} &= A_1 \\
 x_{12} + x_{22} + \dots + x_{m2} &= A_2 \\
 &\vdots \\
 &\vdots \\
 x_{1n} + x_{2n} + \dots + x_{mn} &= A_n \\
 L_{imin} &= L_{total}P_i(1 - \alpha) \\
 L_{imax} &= L_{total}P_i(1 + \alpha) \\
 P_1 + P_2 + \dots + P_m &= 1
 \end{aligned}$$

許容誤差率内で、各ミラーサーバの負荷が収まるように、接続配分値を算出する。

3.3 複数同時判断を行う接続制御

本方式はクライアントからの接続要求を複数同時判断することにより、ボトルネックリンクを回避し、転送速度の向上と負荷分散を両立する接続制御を行う。

図2の単数接続判断において、4番目に発生した接続要求は、各ミラーサーバの負荷分散の観点からボトルネックリンクを通過させるしかない。しかし、図3のように複数同時判断では4つの接続要求を同時に判断し、ボトルネックリンクを迂回させた上での負荷分散が可能である。複数の接続

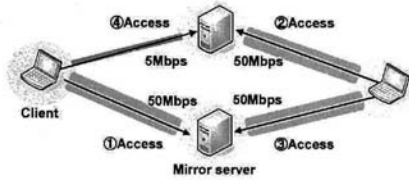


図2: 単数接続判断

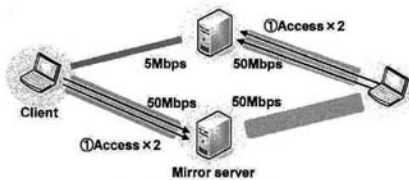


図3: 複数同時接続判断

要求を同時に判断するため、クライアントからの接続要求を一時プールする。この時間は、1秒から5秒を想定しており、クライアントに不満がない範囲内で長いことが望ましい。本方式は大規模コンテンツの転送を目的としており、この時間間隔は許容しうる範囲だと考える。プールの後に接続配分計算を行う。そのため、配分計算のタイミングを一定間隔にでき、接続要求数の増加に伴う計算コストの増加は発生しない。配分値の算出

後、制御サーバは結果を元に各クライアントを誘導し、ミラーサーバへの接続を指示する。

4 シミュレーションと評価

提案方式のミラーサーバへの接続配分をシミュレーションし、評価を行う。比較評価を行うために、以下の4方式においてそれぞれ評価する。

- 提案方式 (Proposed)
- 負荷分散重視方式 (LB): 最も接続負荷数が低いサーバを選択。
- 転送速度重視方式 (Speed): 最も転送速度が速いサーバを選択。
- 相互判断方式 (LB + Speed): 各サーバの負荷不均衡が許容誤差範囲内で収まっているサーバの中で最も速いサーバを選択。単数接続判断。

4.1 評価値の定義

シミュレーション結果を転送速度と負荷分散それぞれの観点から評価を行うため、以下の評価値を定義する。

最高速度達成率 (S_r) 各ASが最高速度を達成している割合の平均である。そのため、理想値が1となり速度低下と共に下降する。以下の式で算出し、 n は参加ASの数とする。

$$S_r = \frac{1}{n} \sum_{j=1}^n (AS_j \text{の平均速度} \div AS_j \text{の最高速度})$$

負荷不均衡偏差 (L_d) 各サーバの負荷を理想負荷配分値と比較した誤差率を出し、すべてのサーバの誤差率を標準偏差にしたものである。理想値が0となり、各サーバの負荷が不均衡になると上昇する。以下の式で算出し、 m は配信サーバの数とする。

$$L_d = \sqrt{\frac{\sum_{j=1}^m (\text{サーバ}_j \text{の誤差率} - \text{平均誤差率})^2}{m - 1}}$$

4.2 シミュレーション設計

今回作成したシミュレーションについて説明する。提案方式のシミュレーションでは、まず、BRITE[8]により、ネットワークトポロジーを作成する。これは、ASと各リンク回線からなるネットワークとなっており、各回線の転送速度を定義されている。このネットワーク上に、ミラーサーバを設置するASの位置を決め、各ASから接続要求を発生させる。接続要求を線形計画法の算出プログラムにより、配分値を決定し、配分後の各ミラーサーバの負荷数、各通信セッションごとの転送速度を算出

して、これらを秒ごとに出力する。また、環境の変化を模擬して、各回線の速度変化を与えている。

この処理をシミュレーション実行上限時間分繰り返し、 S_r 、 L_d を算出して評価する。他の方式では、接続判断部を置き換えて、これと同様の処理を行っている。また、これらはC言語にて作成している。

4.3 配信ネットワーク、接続要求数の影響

図4のネットワークにおいてシミュレーションを行う。これは、ASが10個存在する5つのネットワークを低速のボトルネックリンクで接続した物である。ネットワーク内の回線は30Mbpsから100Mbpsの間の速度で生成し、ボトルネックリンクは10Mbpsとしている。そのため、自ネットワーク以外のミラーサーバへ接続を行う場合、必ずボトルネックリンクを経由するしかなく、大幅な転送速度低下が発生する。この環境下で以下の2通りのシミュレーションを行う。

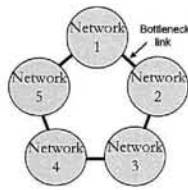


図4: シミュレーション・ネットワーク1

- Sim1.1:ミラーサーバを各1台ずつ、計5台配置。毎秒10アクセスを各ASからランダムに発生。
- Sim1.2:ミラーサーバを計8台不規則に配置。毎秒100アクセスをランダムに発生。

ミラーサーバの設置とアクセス数を変更し、比較を行う。また、両シミュレーションにおいて、提案方式、相互判断方式のそれぞれの負荷許容誤差率は5%とし、提案方式の接続プール時間は1秒間としている。転送コンテンツサイズを1GBとし、1時間のシミュレーションを行う。図5に結果を示す。これは、毎秒ごとの S_r 、 L_d を1時間分の平均として算出している。

結果、Sim1.1では、提案方式は相互判断方式と比較して同じ負荷許容誤差率で S_r が0.06改善した。また、Sim1.2では、相互判断方式はSim1.1の場合と比較して0.07減少している。しかし、提案方式は0.12改善し、そのため、Sim1.2において、 S_r の差は、0.26となった。Sim1.2において提案方式は S_r が0.98、 L_d が0.04となり、これは、転送速度の向上と負荷分散を極めて高い水準で両立している。

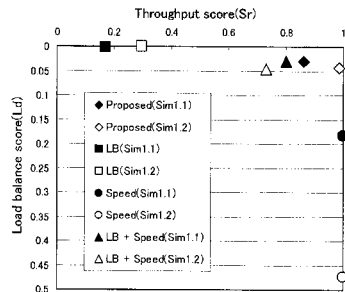


図5: Sim1.1とSim1.2の結果比較

提案方式は複数同時判断によるボトルネック回避を行っている。これは、ボトルネックと感じる高速の転送環境を持つクライアントの代わりに、ボトルネックと感じない低速の環境のクライアントを接続させることによって転送速度の低下を防いでいる。よって、転送環境の格差が大きいクライアントが接続プール時間内に多数の接続がある場合に効率が向上する。そのため、対称配置の場合、各ネットワークの環境が似通い、クライアントごとの格差が少なくなり、相互判断方式との差が少ない。また、接続要求数の増加は、複数同時判断を行う上で、ボトルネック回避する接続配分の組み合わせが増えるので、効率が向上する。よって、提案方式は、格差の大きいクライアントが同時に多数接続する場合に最も効果を発揮する。これは、現実のネットワークにおいては、クライアントごとの環境の格差が激しいため、適用性が高い。また、負荷分散が重要となる多数の接続が発生するサービス環境下において効果を発揮するので実用性が高い。

4.4 国際分散時の接続要求数格差の影響

図6のネットワークにおいてシミュレーションを行う。図4は、ASが30個存在する4つのネットワークをボトルネックリンクで繋いだものであり、内部は30Mbpsから100Mbps、ボトルネックリンクは10Mbpsとしている。

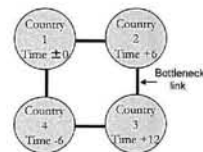


図6: シミュレーション・ネットワーク2

- Sim2:ミラーサーバを2台、1台、1台、なし、計4台不規則配置。接続要求は実際の通信ログを使用。

ミラーサーバは不規則に設置し、また、各 AS から発生する接続要求は、北陸先端大の FTP サー “ftp.jaist.ac.jp” の平成 18 年 12 月 1 日における実際の AS から接続要求推移を使用する。実際に接続要求のあった 471 個の AS から日本の AS、更に連続的な接続要求のある上位 30 位の AS を抽出し、AS ごとの接続要求を BRITE により生成させた AS30 個のネットワークの各 AS に適用する。また、ネットワーク地域ごとに時差があるとし、6 時間ずつずらして接続要求推移を再生、国際的分散配信をシミュレーションする。

図 7 に 1 時間ごとの上位 30 位の AS が発生する接続要求数を示す。時間ごとに接続数の格差があるため、これを時間をずらして各地域に適用すると、時間ごとの各地域の接続要求数に激しい格差が発生する。そのため、負荷を均衡にするには、多くの地域外通信が必要となり、転送速度の効率を維持するのは大変困難な状況である。

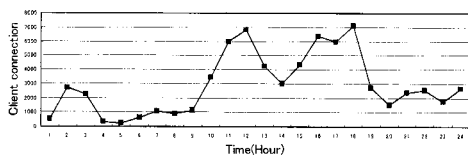


図 7: FTP ログにおける日本の AS 上位 30 位の 1 時間ごとの接続要求数

転送コンテンツサイズは 1GB とし、提案方式、相互判断方式の負荷許容誤差率 20%、提案方式の接続プール時間は 5 秒とする。このシミュレーションを 24 時間分を行い、結果の平均を図 8 に示す。

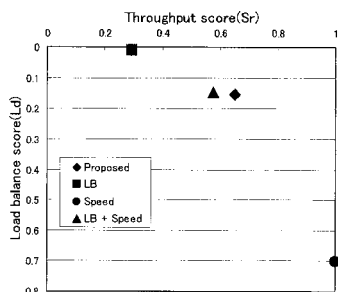


図 8: Sim2 結果

結果、提案方式は、相互判断方式と比較して、同じ負荷不均衡率内で、 S_r が 0.07 向上した。また、両立が困難な状況の中で、 $S_r = 0.65$ 、 $L_d = 0.15$ を達成した。

この結果により、地域ごとの発生接続数の時間ごとの地域格差を吸収して分散を行い、転送速度の低下を防いでいるのが分かる。そのため、一日の発生負荷のピーク時に合わせてサーバの設置を行わなければならなかったのが、提案方式を用い

ることにより一地域で活性化した接続負荷を、接続発生が少ない時間帯の地域の配信サーバに転送することができる。広域分散配置環境下で、システムを効率よく分散利用でき、また、転送効率の低下を防ぐことができる。

実際の接続要求を使うことにより、現実の秒単位で激しく変動する負荷推移への適用性も確認できた。提案方式は複数同時接続判断により転送速度向上を行うため、接続要求数が少ない時には効率が低下する。しかし、単数接続判断となった場合は従来の相互判断方式と同じ働きとなるだけなので、提案方式の優位性は変わらない。

5 おわりに

転送速度の向上と負荷分散を両立する広域サーバ負荷分散方式の提案を行った。

提案方式において不規則なミラーサーバ配置への高い適用性が確認できた。これは、現実のネットワークにおいて、ネットワークの形的にクライアントを均等に分けるようミラーサーバを設置するのは不可能なので、不規則性の対応は有効である。また、国際的なサーバ分散配置に対する適用性も確認できた。時差による地域ごとの時間あたりの接続要求発生格差を吸収し、高速のクライアントが低速の国際回線の使用を避けるため、高い有効性を発揮する。

今後、線形計画法の構成式を変更して、ストーリーミングなど用途の拡大を考える。

参考文献

- [1] 横田裕思, 木村成伴, 海老原義彦: DNS フィルタ方式によるミラーサーバ選択法の提案と実装, 情報処理学会論文誌, Vol. 44, No. 3, pp. 682-691 (2003).
- [2] 和泉勇治, 宇津江康大, 加藤肇, 根元義章: 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, 情報処理学会論文誌, Vol. 45, No. 1, pp. 65-73 (2004).
- [3] 金一怡, 敷田幹文: コンテンツのサンプリングスルーブットによるミラーサーバ選択法の提案, 情報処理学会研究報告, 2004-DPS-120, pp. 55-60 (2004).
- [4] 奥村巧, 敷田幹文: 複数パラメータを用いた複合型調査システムによるミラーサーバ選択法の提案, 情報処理学会研究報告, 2005-DPS-125, pp. 85-90 (2005).
- [5] M.Dahlin: Interpreting Stale Load Information, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 11, No. 10, pp. 1033-1047 (2000).
- [6] 佐竹伸介, 稲井寛: Web サーバクラスタにおける定期的な負荷情報に基づく非確率的なサーバ選択法, 電子情報通信学会論文誌, Vol. J88-B, No. 10, pp. 1968-1978 (2005).
- [7] 荻野司, 松田和宏, 須藤一顕, 針山欣之, 向阪正彦: 広域分散配置 Web サーバにおける最適サーバ検索システムの検討, 情報処理学会研究報告, 2000-DSM-018, pp. 7-12 (2000).
- [8] A.Medina, A.Lakhina, I.Matta and J.Byers: BRITE: An Approach to Universal Topology Generation, in *MASCOTS '01* (2001).