

WWWにおけるラウンドロビン-トラフィック適応型サーバ選択方式

富沢 考弘 山本 潮 小野里 好邦

群馬大学 工学部 情報工学科

〒 376-8515 群馬県桐生市天神町 1-5-1

{tomizawa, kansuke, onozato}@nzt1.cs.gunma-u.ac.jp

Abstract

インターネットにおける WWW トラフィックの増加および集中がサーバに対するアクセス時間やサーバ、回線の負荷に影響を及ぼす。この問題を解決する方法としてミラーサーバの設置による負荷の分散が行なわれている。我々は、インターネットの混雑状況をラウンドトリップタイムの観測により推測し、適切なミラーサーバの選択を行なう TASS (Traffic Adaptive Server Selection) 方式を提案した。本稿では TASS 方式の問題点を解決するために、DNS のラウンドロビン機能を利用する RR-TASS 方式 (Round Robin-TASS 方式) を提案する。これは、DNS のラウンドロビン機能によりプライマリサーバをクライアントからのアクセスの度に順次変更させることにより、プライマリサーバのみに発生する負荷も分散させることが可能になる。また、シミュレーションにより本方式と TASS 方式とのアクセス時間の比較を行ない本方式の効果についても考察する。

1. はじめに

近年、WWW (World Wide Web) はブラウザの GUI による操作性と全世界からの情報取得の容易性などから急激に発展し、現在もお成長を続けている。初期的にはテキストのみであった Web ページも最近では音声、画像、動画等のマルチメディアコンテンツが組み込まれるようになり、その結果として Web ページサイズが増加傾向にある。これらがインターネットにおける WWW トラフィックの増加の原因となり、クライアントのサーバに対するアクセス時間、サーバや回線の負荷に影響を及ぼす。アクセス時間の短縮やサーバなどの負荷を軽減させる方法の一つとして、現在ではミラーサーバの設置が行なわれている。これは、同じコンテンツを保持する Web サーバを異なる場所に複数設置することにより、Web サーバの負荷やネットワークトラフィックを分散させることを可能にする。ここで、同一コンテンツをもつ Web サーバの中でオリジナルのコンテンツをもつサー

バをプライマリサーバ、負荷分散のためにコピーを持って設置されたサーバをミラーサーバと呼ぶ。

あるユーザがこれらのサーバ群からアクセスするサーバを決定する方法は (1) ユーザ自身による手動選択、(2) システムによる自動選択に大別されるが、前者は選択基準の曖昧さやユーザに対して余計な負荷を与えるという問題点がある。後者に関しては、DNS (Domain Name Service) のラウンドロビン機能を利用した方法が提案されている [2]。これは、一つのドメイン名に対して複数の IP アドレスを割り当てておき、そのドメインに対する IP アドレスの照会要求の度に順次登録された IP アドレスを返す方式である。この機能の使用により、ユーザのアクセスを各ミラーサーバに分散できるが、ネットワークの静的状況 (物理的な接続関係など) かつ動的状況 (混雑状況など) を考慮しないため、選択されたミラーサーバがそのユーザにとって最適なサーバである保証はない。

また、後者の別の方法として Distributed Director [1] のようなハードウェアレベル、Global Dispatch [3] のようなソフトウェアレベルでの自動選択方式が開発されているが、これらの方式の共通点として、情報を伝送するのに適切なサーバを決定する権限がサーバ自身ではなく仲介システムにあるため、サーバ自身は動いていても仲介システムのトラブルによりサーバが機能しなくなる、またクライアントとサーバの間に仲介システムが存在することによるトラフィックの増加などの問題点がある。

上記の問題点を克服するために我々はインターネットの混雑状況をラウンドトリップタイムの観測により推測し、適切なミラーサーバの選択を行なう TASS (Traffic Adaptive Server Selection) 方式 [4] を提案している。しかしながら、この方法では、クライアントからの要求がプライマリサーバに集中するためにミラーサーバよりも負荷が大きくなるという問題点があった。

本稿では TASS 方式の問題点を解決するために、DNS のラウンドロビン機能を利用する RR-TASS 方式 (Round Robin-TASS 方式) を提案する。これは、DNS のラウンドロビン機能によりプライマリサーバをクライアントからのアクセスの度に順次変更させることにより、プライ

マリサーバのみに発生する負荷も分散させることが可能になる。また、シミュレーションにより本方式と TASS 方式とのアクセス時間の比較を行なうことにより本方式の効果についても考察する。

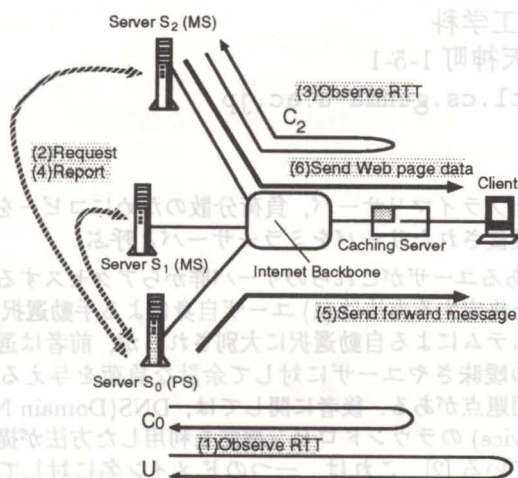


图 1. TASS 方式

2. TASS 方式

TASS 方式は、ICMP を使用してパケットによるラウンドトリップタイムの観測でネットワーク状況を推測し、適切なサーバの選択を行なう方式である。図 1 に TASS 方式の動作を示す。まず、アクセスを受けたプライマリサーバは、Web サーバ/クライアント間、Web サーバ/キャッシングサーバ間のラウンドトリップタイムを観測する。ここで、観測結果の評価基準として閾値 K_u 、 K_c を導入する。 K_u はサーバ/クライアント間が混雑していると判断されるラウンドトリップタイムの閾値、 K_c はサーバ/キャッシングサーバ間が混雑していると判断されるラウンドトリップタイムの閾値を表す。ラウンドトリップタイムを評価した結果、プライマリサーバが伝送するのに適切でないと推測した場合、プライマリサーバは各ミラーサーバに対してミラーサーバ/キャッシングサーバ間のラウンドトリップタイムの観測要求を行なう。各ミラーサーバはラウンドトリップタイムを観測し、その結果をプライマリサーバに返す。プライマリサーバは最小値のラウンドトリップタイムを返したミラーサーバに転送するようクライアントにメッセージを送り、そのサーバがクライアントに Web ページデータを伝送する。この方式では、評価値としてラウンドトリップタイムを使用することでネットワークの状況を考慮したアクセス分散を行なうことができる。また、閾値を導入することで制御パケットによる追加トラフィックを極力減

少させている。しかし、TASS 方式では以下のような問題が生じる。まず、TASS 方式はプライマリサーバが中心となって制御を行なうため、制御用パケットがプライマリサーバに集中してしまう。また、TASS 方式では閾値 K_u 、 K_c を導入しているが、プライマリサーバはこれらの値を越えるまで、つまりクライアント/プライマリサーバ間の回線の状態が悪くなるまでミラーサーバには転送しない。これらが原因となってプライマリサーバはミラーサーバに比べ、負荷が集中し、回線状況が悪化してしまう。TASS 方式では必ず最初にプライマリサーバにアクセスすることから、プライマリサーバの負荷はそのままアクセス時間に影響を及ぼすことになる。

3. 提案方式

3.1. 概要

2節で述べた問題を解決するために、TASS方式でのプライマリサーバの役割を果たすサーバの選択、つまり最初にアクセスするサーバの選択にDNSのラウンドロビン機能を利用したRound Robin-TASS(RR-TASS)方式を提案する。TASS方式では全てのサーバがプライマリサーバまたはミラーサーバとして機能が分散されていたが、本方式では全てのサーバがプライマリサーバとミラーサーバの両方の機能を保持する必要がある。そして、DNSのラウンドロビン機能により全てのサーバに均等にアクセスを振り分け、最初にアクセスしたサーバをプライマリサーバ、残りのサーバをミラーサーバとして動作させる。プライマリサーバとなったサーバはTASS方式のプライマリサーバと同じ動作をする。つまりクライアントからのアクセスを他のサーバに転送するかアクセスされたサーバ自身が伝送するかを決定する。

本方式によりプライマリサーバに偏っていた制御パケットやクライアントの要求によるトラフィックが分散できるため、プライマリサーバ及びプライマリサーバ側の回線の負荷の集中が軽減され、アクセス時間の短縮が見込める。

3.2. 動作

以下に本方式における、Webサーバおよびクライアントの動作を説明する。

Step1 : クライアントはサーバ S_n に対して Web ページの伝送要求を行なう。このときクライアントが伝送要求するサーバ S_n は DNS のラウンドロビン機能により決定される (図 2 中の (1)).

Step2 : 伝送要求を受けたサーバ S_n はラウンドトリップタイム U_n および C_n の観測を行なう (図 2 中の (2)). これらの値を以下のように TASS 方式と同様に閾値 K_u, K_c と比較してネットワーク状況を推測する.

1. $U_n \leq K_u$ かつ $C_n \leq K_c$
(ネットワーク状況が良好)
2. $U_n \leq K_u$ かつ $C_n > K_c$
(サーバ/キャッシングサーバ間がやや混雑…キャッシングサーバ/クライアント間の回線帯域の広さでカバー可能)
3. $U_n > K_u$ かつ $C_n \leq K_c$
(キャッシングサーバ/クライアント間が混雑…転送による改善不可能)
4. $U_n > K_u$ かつ $C_n > K_c$

比較した結果、1,2,3のときは、サーバ S_n がクライアントに対してデータの伝送を行ない、RR-TASS 方式の動作はここで終了する。4のときは、転送によるアクセス時間の改善が期待できるため、他のサーバへアクセスを転送することを試みる。

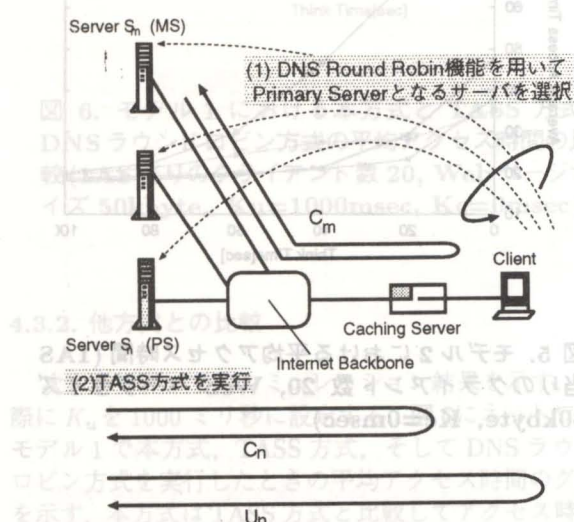


図 2. RR-TASS 方式

Step 3: サーバ S_n は全ミラーサーバに対して各ミラーサーバ/キャッシングサーバ間のラウンドトリップタイムの観測要求を行なう。これを受けて各サーバはミラーサーバ/キャッシングサーバ間のラウンドトリップタイム $C_m (m = 1, 2, \dots, N \text{ ただし } n \text{ を除く})$ の観測を行なう。

Step 4: 各サーバで観測したラウンドトリップタイムをサーバ S_n に報告する。

Step 5: プライマリーサーバは最小の C_m あるいは C_n を報告したサーバ S_{min} を選択し、そのサーバへの転送要求をクライアントに対して伝送する。

Step 6: クライアントは選択されたミラーサーバ S_{min} に対してリクエストを行なう。

Step 7: サーバ S_{min} はクライアントに対してデータの伝送を行なう。

4. シミュレーション

4.1. シミュレーションモデルの仮定

本方式におけるシミュレーションモデルの仮定を以下に示す。

- プライマリサーバからミラーサーバへの転送要求が送信されて来た場合、クライアントは指定されたミラーサーバに対して伝送要求を行なう。
- サーバから Web ページデータが完全に伝送されるまで、新たな伝送要求は行なわない。
- Web サーバから Web ページデータが完全に伝送されたのち、平均 t 秒の指数分布に従う時間 (思考時間) をおいて新たな伝送要求を DNS により指定されたサーバに対して行なう。DNS はクライアントからの問い合わせを受けるごとに登録されている Web サーバを順次返す。

4.2. ネットワークモデル

ネットワークモデルを図 3、図 4 に示す。各 Web サーバはインターネットバックボーンにそれぞれ固有の帯域で接続されている。

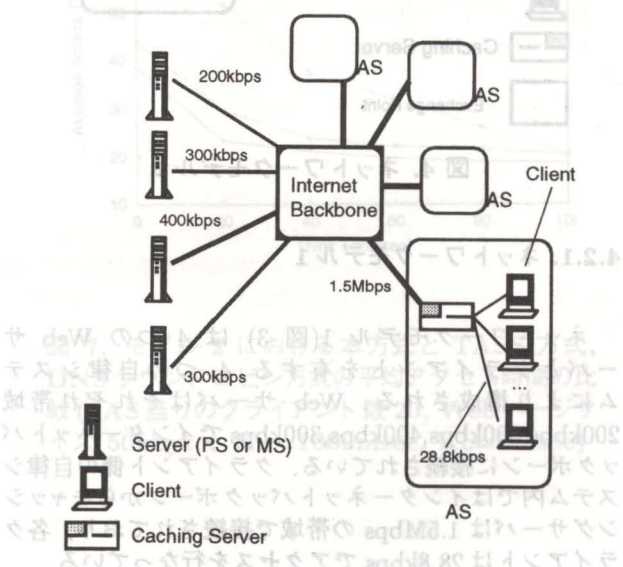


図 3. ネットワークモデル 1

この帯域が小さい場合、サーバ回線能力の不足や、自律システム内のホットスポットの存在を表している。クライアント側の回線はインターネットバックボーン/キャッシングサーバとキャッシングサーバ/クライアント群、クライアント回線の3つに分割されている。キャッシングサーバはクライアント側の各自律システムに1つ存在し、自律システム内のクライアントは全てそのキャッシングサーバに接続されているものとする。Web ページデータの伝送は 2048byte のパケットに分割され伝送されるものとする。1つのクライアントへの Web ページデータの伝送において、Web ページデータはパケットとして1つずつクライアントに送られ、1つのパケットがクライアントに到着するまで新たなパケットの伝送は行なわれないものとする。このパケット伝送の動作はウィンドウサイズが常に1である場合を表している。以下各モデルについて詳しく述べる。

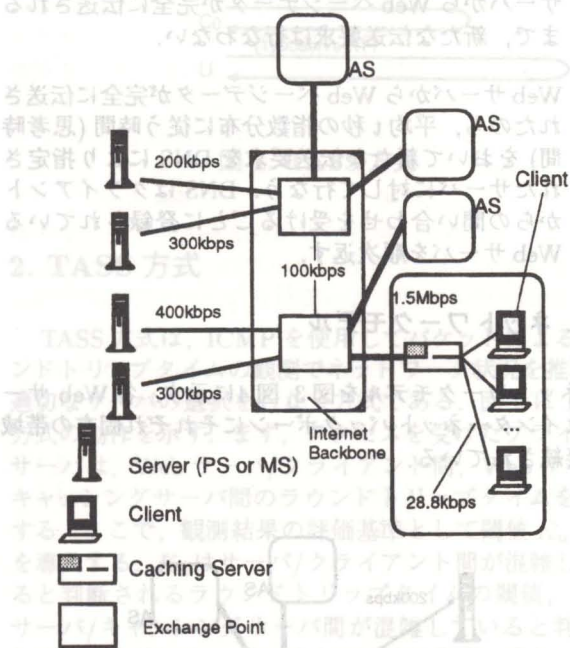


図 4. ネットワークモデル 2

4.2.1. ネットワークモデル 1

ネットワークモデル 1(図 3) は 4 つの Web サーバとクライアントを有する 4 つの自律システムにより構成される。Web サーバはそれぞれ帯域 200kbps, 300kbps, 400kbps, 300kbps でインターネットバックボーンに接続されている。クライアント側の自律システム内ではインターネットバックボーンからキャッシングサーバは 1.5Mbps の帯域で接続されており、各クライアントは 28.8kbps でアクセスを行なっている。

4.2.2. ネットワークモデル 2

ネットワークモデル 2(図 4) はネットワークの相互接続ポイント間のボトルネックを示したモデルである。Web サーバはそれぞれネットワークモデル 1 と同じ回線帯域で接続されている。クライアント側の自律システム内ではインターネットバックボーンからキャッシングサーバは 1.5Mbps の帯域で接続されており、各クライアントは 28.8kbps でアクセスを行なっている。

また相互接続ポイントは 2 つあり自律システムとサーバはそれぞれ 2 つずつ相互接続ポイントに接続されている。相互接続ポイント間の回線は混雑状態を想定し 100kbps の狭帯域に設定する。

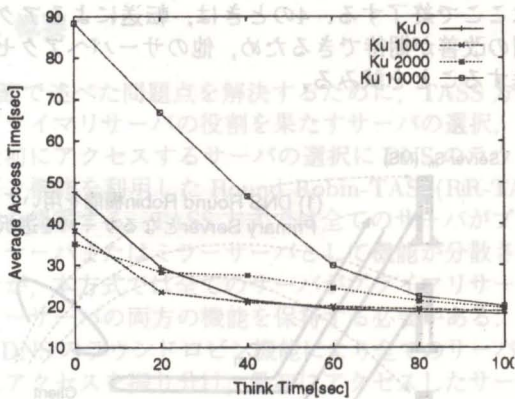


図 5. モデル 2 における平均アクセス時間 (1AS 当りのクライアント数 20, Web ページサイズ 50kbyte, $K_c=0\text{msec}$)

4.3. 数値例

4.3.1. K_u によるアクセス時間への影響

まず K_u の値を変化させることによる RR-TASS 方式の動作の影響について考察する。 K_u が 0 ミリ秒のとき、プライマリサーバでの伝送が適さないと判断されるため、必ずミラーサーバに転送要求を行なう。そのため、 K_u が 0 ミリ秒のときはネットワーク内における制御パケットが最も増加するので、ネットワーク内、特にネットワークモデル 2 では相互接続ポイント間の回線が混雑しアクセス時間が増大する。 K_u を増加させていくとプライマリサーバ自身も伝送を行なうようになり、ネットワーク内における制御パケットが減少していく。そのためネットワークの混雑度が減少し、アクセス時間が短縮される。さらに K_u を増加させるとプライマリサーバ自身が

必要以上に伝送を行ってしまうため、プライマリサーバ回線が混雑してしまいアクセス時間が増大する。この結果より、本方式においてはアクセス時間を最も短くする K_u が存在すると考えられる。今回シミュレーションをネットワークモデル2で行なった結果、図5より K_u が1000ミリ秒前後のときが適切な値として考えられる。

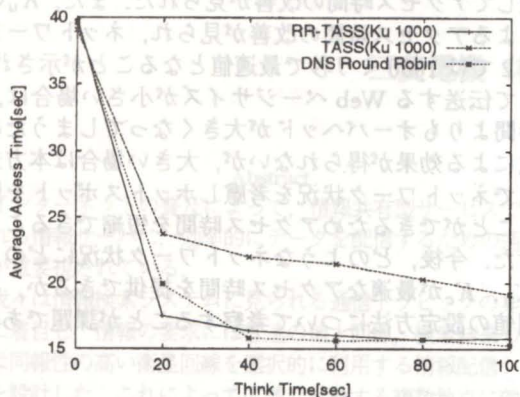


図 6. モデル 1 における本方式と TASS 方式、DNS ラウンドロビン方式の平均アクセス時間の比較 (1AS 当りのクライアント数 20, Web ページサイズ 50kbyte, $K_u=1000$ msec, $K_c=0$ msec)

4.3.2. 他方式との比較

次に各モデルでのシミュレーション結果を示す。その際に K_u を 1000 ミリ秒に設定する。図 6 にネットワークモデル 1 で本方式、TASS 方式、そして DNS ラウンドロビン方式を実行したときの平均アクセス時間のグラフを示す。本方式は TASS 方式と比較してアクセス時間の短縮が見られる。DNS ラウンドロビン方式と比較すると思時間 40 秒未満についてはアクセス時間の短縮が見られ、40 秒以上はあまり変わらない。思考時間が長くなるにつれて Web サーバがある瞬間に処理するクライアントからの要求数が減少するため、ネットワークの混雑も減少する傾向にある。しかし、ある程度思考時間が長くなるとネットワークはほとんど混雑しない状況になる。今回の結果では、思考時間が 40 秒前後まではネットワークが混雑する状況が発生するために、混雑していないミラーサーバが Web ページを伝送する機能が働く本方式の方が混雑していてもそのままそのサーバが伝送する DNS ラウンドロビン方式よりもアクセス時間が短くなっている。しかしながら、思考時間がそれより長くなると、ネットワークが混雑する状況がほとんどなくなるために本方式も DNS ラウンドロビン方式と同様の動きをすることになるが、プライマリサーバとクライアン

ト間のネットワークの混雑状況を測定するために、その分のオーバーヘッドがアクセス時間にかかる結果となっている。

図 7 はネットワークモデル 2 で本方式、TASS 方式、そして DNS ラウンドロビン方式を実行したときの平均アクセス時間のグラフである。本方式は他の方式と比較してアクセス時間の短縮が見られる。DNS ラウンドロビン方式と比較して、本方式の使用により Web ページデータを伝送するとき相互接続ポイント間の回線を通る回数が減少するのでアクセス時間が短縮される。また TASS 方式より効果が得られた理由として次のように考えられる。TASS 方式では、プライマリサーバ回線が混雑するまで直接プライマリサーバが Web ページ伝送を行ってしまう、プライマリサーバ/クライアント間の回線が混雑した時はじめて他のミラーサーバに転送する。つまり、転送をはじめる時点では、プライマリサーバ/クライアント間の回線のみが混雑していて他のサーバにトラフィックが分散されていない。したがって、制御パケットや混雑するまでの Web ページデータはプライマリサーバから伝送するため、プライマリサーバ回線の混雑がそのままアクセス時間に影響を及ぼす。しかし、本方式ではプライマリサーバとして働くサーバが DNS のラウンドロビン機能により順次変更されるため、プライマリサーバとして動作することにより生じるトラフィックが各サーバに分散され、TASS 方式よりもアクセス時間が改善したと考えられる。

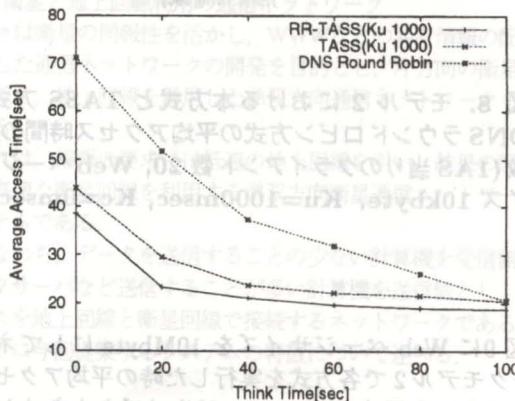


図 7. モデル 2 における本方式と TASS 方式、DNS ラウンドロビン方式の平均アクセス時間の比較 (1AS 当りのクライアント数 20, Web ページサイズ 50kbyte, $K_u=1000$ msec, $K_c=0$ msec)

4.3.3. Web ページサイズによるアクセス時間への影響

図8に Web ページサイズを 10kbyte にしてネットワークモデル 2 で各方式を実行した時の平均アクセス時間のグラフを示す。この場合、Web サーバから Web ページのみを伝送する時間が短くなるので、各サーバからクライアントへの Web ページの伝送時間の差は小さくなる。本方式は制御パケットとして数種類のパケットを用意し、パケットサイズはそれぞれ 2048byte としているため、制御パケットの伝送時間が全アクセス時間に対して大きな割合を占めることになる。この結果として、制御パケットを利用しない DNS ラウンドロビン方式が最も小さくなったと考えられる。

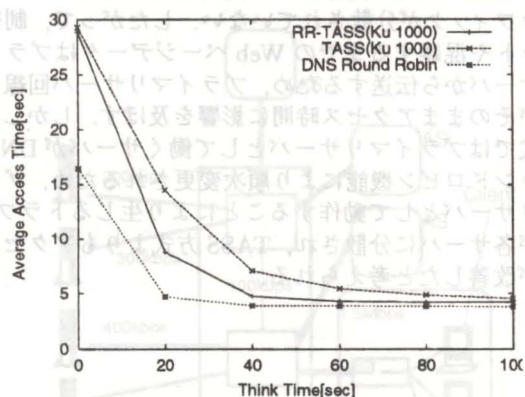


図 8. モデル 2 における本方式と TASS 方式、DNS ラウンドロビン方式の平均アクセス時間の比較 (1AS 当りのクライアント数 20, Web ページサイズ 10kbyte, $K_u=1000\text{msec}$, $K_c=0\text{msec}$)

図 9 に Web ページサイズを 10Mbyte にしてネットワークモデル 2 で各方式を実行した時の平均アクセス時間のグラフを示す。Web ページサイズを大きくした場合にも本方式の使用による効果が見られる。このシミュレーションでは、平均アクセス時間に対して思考時間が小さいため思考時間による影響が少なくなり各方式とも x 軸に平行になっている。また Web ページサイズを大きくしたことから全体的にアクセス時間が増加したが、DNS ラウンドロビン方式はネットワーク状況を考慮しないため相互接続ポイント間の回線の混雑による影響を特に受け、他の方式に比べアクセス時間が大幅に増加している。

5. まとめ

本稿では DNS のラウンドロビン機能を利用し TASS 方式の問題点を解決する方式を提案した。またシミュレーションを行ない本方式と TASS 方式および DNS ラウンドロビン方式のアクセス時間の比較を行なった。

その結果、TASS 方式、DNS ラウンドロビン方式と比較してアクセス時間の改善が見られた。また、 K_u の制御によるアクセス時間の改善が見られ、ネットワークモデル 2 では 1000 ミリ秒で最適値となることが示された。そして伝送する Web ページサイズが小さい場合は、伝送時間よりもオーバーヘッドが大きくなってしまったため本方式による効果が得られないが、大きい場合は本方式の使用でネットワーク状況を考慮しホットスポットを回避することができるためアクセス時間を短縮できることが示された。今後、どのようなネットワーク状況にどのような K_u , K_c が最適なアクセス時間を提供できるか、これら閾値の設定方法について考察することが課題である。

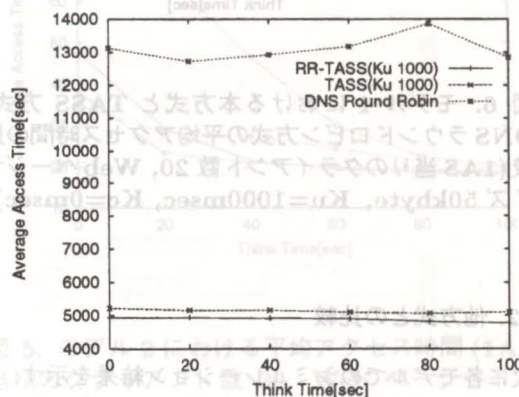


図 9. モデル 2 における本方式と TASS 方式、DNS ラウンドロビン方式の平均アクセス時間の比較 (1AS 当りのクライアント数 20, Web ページサイズ 10Mbyte, $K_u=1000\text{msec}$, $K_c=0\text{msec}$)

参考文献

- [1] Cisco System, Inc. *Cisco DistributedDirector*. <http://www.cisco.com>.
- [2] A. Paul and L. Cricket. *DNS and BIND*. O'Reilly & Associates, Inc, 1992.
- [3] Resonate, Inc. *Resonate Global Dispatch*. <http://www.resonateinc.com>.
- [4] 竹内, 小野里, 山本, 清水, 木村. インターネットにおけるトラフィック適応型サーバ選択方式. 情報処理学会第 55 回全国大会, 5U-1, 1997.