

広域分散 Web サーバにおける経路情報を用いたサーバ選択

下川 俊彦* 中川 郁夫† 山本 文治‡ 吉田 紀彦+

*九州大学, †インテック・ウェブ・アンド・ゲノム・インフォマティクス
‡アイアイジェイメディアコミュニケーションズ, +長崎大学

* 〒 812-8581 福岡市東区箱崎 6-10-1

toshi@csce.kyushu-u.ac.jp, ikuo@intec.co.jp, bunji@iij-mc.co.jp, yoshida@cis.nagasaki-u.ac.jp

あらまし

従来の単純なサーバクライアントモデルでは、処理要求がサーバに集中し、十分なサービスを提供することができない。この問題を解決するための手法に CDN (Content Delivery Network) がある。CDN を構築するためには、利用者からの処理要求を、分散配置したコンテンツ配布用サーバに適切に誘導する必要がある。これを実現するものがサーバ選択機構である。

本研究ではインターネット上の経路情報と DNS を利用することで、利用者から透過かつサーバやクライアントから非依存に実現可能なシステムを開発した。本システムを日食中継システムに適用し、評価を行った。その結果、処理要求の 80.9% を近傍のサーバへ誘導することができた。

キーワード

CDN, DNS, BGP, サーバ選択, リクエストルーティング

Routing-information-based server selection in widely distributed web servers

Toshihiko Shimokawa* Ikuo Nakagawa†
Bunji Yamamoto‡ Norihiko Yoshida +

*Kyushu University, †INTEC Web and Genome Informatics Corporation
‡IIJ Media Communications, +Nagasaki University

*Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan

toshi@csce.kyushu-u.ac.jp, ikuo@intec.co.jp
bunji@iij-mc.co.jp, yoshida@cis.nagasaki-u.ac.jp

abstract

In a traditional simple server client model, many service requests overconcentrate to the server. Now, CDN (Content Delivery Network) is used to solve this problem. To construct a CDN, request-routing system is required. It navigates a user request to one of appropriate servers.

In this paper, we present a new server selection mechanism. This system refers routing information, and is based on DNS servers. It is independent of servers and clients. We evaluate the system in a solar eclipse live web servers at 21 June, 2001. In this experiment, we navigated 80.9% of requests to one of their neighbor server.

Keywords

CDN, DNS, BGP, server selection, request routing

1 はじめに

FTTH, ADSL, CATV を代表とするいわゆるブロードバンドネットワークの、一般利用者への普及が急速に進んでいる。

従来の単純なサーバクライアントモデルでは、サーバへの処理要求が集中し、サーバシステムの処理性能やネットワーク性能に不足が出てきている。この結果、従来の多数の利用者へ十分なサービス品質を提供することが困難になってきている。

一般に、高性能な計算機や、広帯域なネットワークは高価である。そこで、経済的にこの状況に対処する手法が必要となってきた。この問題に対する解決策の一つが Content Delivery Network (CDN) である。これは、web やストリーミングのようなコンテンツ配布系のサービスを対象にしている。利用者の近くに配布用サーバを配置することで負荷の分散を計り、高品質なサービスを提供するシステムである。

C. Day らは、CDN を構成する要素は次の4つであると述べている [1]。

1. コンテンツ配送基盤¹
2. 要求ルーティング基盤²
3. 配布基盤³
4. 課金基盤⁴

この中で、要求ルーティング基盤とは、利用者からのアクセス要求を、利用者の近くにある配布用サーバに誘導するための機構である。これは、従来より我々が研究開発を進めているサーバ選択機構の機能と同様のものである。我々の機構は、サーバ選択機構 TENBIN

¹content-delivery infrastructure

²request-routing infrastructure

³distribution infrastructure

⁴accounting infrastructure

および、経路情報サーバ RADIX からなる [2, 3]。これまでに、本機構をストリームサーバのサーバ選択に適用し実用実験を行ってきた [4]。本論文では、本機構を広域分散配置した web サーバのサーバ選択に適用した事例について評価する。

2章では、我々が構築している TENBIN と RADIX によるサーバ選択機構について概略を説明する。3章では、今回の実験を行った 2001年6月21日の皆既日食中継システム、ならびにその際のサーバ選択ポリシーについて説明する。4章では実験結果を評価する。5章で本論文のまとめを行う。

2 サーバ選択への経路情報の利用

CDN はその網の展開範囲によって次の2種類に分類することができる。

- ISP⁵内に閉じた CDN
- 複数の ISP に跨った CDN

前者は、対象を特定の ISP 内の利用者だけに絞ったものである。最近では多くの ISP が付加サービスとして CDN の構築を進めつつある。後者は、インターネット全域の利用者を対象とする。AKAMAI[5]のように、既にサービスを行っている事業者も存在する。また、前者のタイプの CDN を複数束ねて一つの CDN を構築する場合もある。このような CDN は、後者のタイプと捉えることができる。

現在のインターネットでは、各 ISP が独立した自律システムを構成していると考えて良い。ここで、自律システムとは、経路制御 (BGP) における自律システムである。従って

⁵Internet Service Provider

CDN を次の 2 種類に分類しなおすことができる。

- 自律システム内に閉じた CDN
- 複数の自律システムに跨った CDN

本研究が対象とする環境は後者である。つまり、選択の対象となる配布用サーバは、自律システムを跨って広域に分散している。このような環境では BGP の経路情報を用いることで、各配布用サーバと利用者の位置関係を把握することが可能である。そこで、サーバ選択の判断材料として経路情報を用いる。次節で、我々が構築している経路情報を利用するサーバ選択機構について説明する。

2.1 TENBIN と RADIX によるサーバ選択機構

TENBIN と RADIX によるサーバ選択機構の基本的なアイデアは単純である。

- 経路情報を元にサーバを選択する
- DNS の枠組みを通じて選択結果を返す

この各々を、TENBIN と RADIX が担当する。

TENBIN[2] は DNS を基盤として用いたサーバ選択機構である。TENBIN は、柔軟なサーバ選択ポリシーを利用可能という特徴を持つ。サーバ選択ポリシーとは、サーバ選択を行う際のルールのことである。

RADIX[3] は経路情報サーバであり、インターネットの広域での経路制御で標準的に利用されている BGP の情報を取り扱うことが可能である。経路情報交換の通信路を遮断せずに、プログラムモジュールの更新が可能という特徴を持つ。

TENBIN と RADIX によるサーバ選択機構を図 1 に示す。

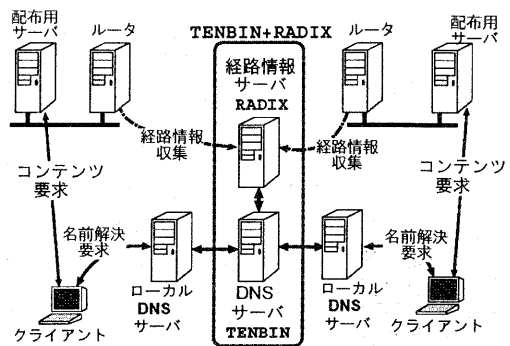


図 1: TENBIN+RADIX

配布用サーバが分散配置されていることが前提である。前述のように、これら分散配置した配布用サーバを選択するための情報としては BGP による経路情報を利用する。そこで、各配布用サーバから経路情報を収集する。実際には、本サーバ選択機構が判断するのに十分な経路情報を、配布用サーバが持つことは少ない。また、配布用サーバから直接経路情報を収集することは運用上も困難である。そこで、配布用サーバが所属するネットワークのルータが持つ経路情報を、経路情報サーバ RADIX に集める。

利用者から見た場合、本サーバ選択機構は単なる DNS サーバに見える。より詳細に見ると、クライアントは従来通り近傍の DNS サーバ(以下、ローカルDNS)へ DNS 問い合わせを発行する。問い合わせを受けたローカルDNSが、その問い合わせを解決するためにDNSの通信プロトコルに従ってDNSのツリーを辿ると、それがTENBINに到達する。ここで、TENBINはRADIXが収集した情報を用い、最適な配布用サーバを選択する。「最適」の判断方法としては、経路の有無、経路長等があり、サービス(コンテンツ)毎に変更することが可能である。この判断方法が、前述の「選択ポリシー」となる。選択した配布用サーバのIPアドレスをDNSの通信プ

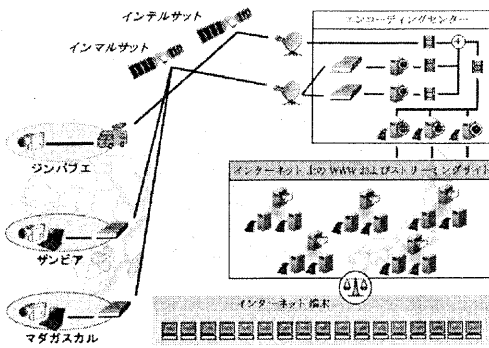


図 2: 日食中継システム

ロトコルに従って、前述のローカル DNS に返答する。最後に、ローカル DNS から得た IP アドレスに対して、クライアントが処理要求を発行する。この結果、利用者は最適な配布用サーバからコンテンツの提供を受けることができる。

3 日食中継

ライブ!エクリプス実行委員会 [6] では 1997 年 3 月の皆既日食の世界初のインターネット多地点中継を皮切りに、過去 7 回におよび皆既日食と金環日食、皆既月食のインターネット中継を手がけてきている。1999 年 8 月の皆既日食中継では、ストリームサーバに対して 16 万アクセス、ピーク時 5648 ストリームという数字を記録している。

2001 年 6 月 21 日、今世紀初の皆既日食が起きた。ライブ!エクリプス実行委員会では今回もインターネット中継を実施した。

3.1 日食中継システム

今回の日食中継で用いた中継システムの概略図を図 2 に示す。アフリカの 3 個所で撮影した映像を日本国内に設置したエンコーディ

ングセンターに衛星回線経由で集信した。エンコーディングセンターでは、3 箇所からの映像を編集し番組を構成した。この番組をインターネット中継用のストリーミングメディアにエンコードした。エンコードされたストリームを、広域分散配置したストリーミングサイトに配送した。利用者がアクセスするための web サーバは 6 つのサイトに広域分散配置した。この web サーバにはストリームアクセス用の web ページや日食に関する情報を掲載した、

利用者が本 web サーバにアクセスを要求する際、web サーバの実体が分散配置されていることを意識する必要はない。我々が構築したサーバ選択機構により、近傍のサーバに誘導される。この際、どのようにして近傍のサーバを選択したかについては、次節で説明する。

今回の日食中継においては、ストリームサーバも広域分散配置した。この際のサーバ選択にも我々のサーバ選択機構も利用したが、本論文ではこの結果については割愛する。

3.2 日食中継における Web サーバ選択

前節で述べたように、今回の日食中継においては Web サーバおよびストリームサーバのサーバ選択に TENBIN と RADIX を利用した。本論文では、このうち Web サーバのサーバ選択について述べる。

基本的な選択ポリシーは単純なものである。各利用者と全 web サーバ間の距離を比較し、最も近いものへ誘導する。この際、「距離」としては、自律システムパス長を用いた。これは、各 web サーバが存在する自律システムから、利用者が存在する自律システムへパケットを送信する際に通過する自律システムの数

である。この自律システムパス長が短いほど、web サーバと利用者との距離が近いと判断することができる。

しかし、今回の日食中継においては、web サーバ設置サイトによっては、特定の条件に合致する場合にのみコンテンツの提供を行いたいという要望があった。そこで各々のサイトの要望に答えるためのルールを作成した。最終的に、これら複数のルールの組合わせで選択ポリシーを構成した。次節以降ではこれらのルールについて述べ、最後に実際に利用した選択ポリシーについて説明する。

3.2.1 自律システムパス長選択

今回の選択ポリシーの基本は、BGP における自律システムパス長の長短で選択するというルールである。しかし、自律システムパス長の比較においては、あまり長い自律システムパス長どうしを比較しても意味がないことが知られている [7]。そこで、今回は最短の自律パス長が 3 以下の場合にのみ、このルールによる選択を行った。

3.2.2 提供する経路情報によるサービス範囲の制御

多くの web サーバ設置サイトからは、全経路情報を収集した。しかし、一部の web サーバ設置サイトからは、その web サーバ設置サイトが望む経路情報のみの提供を受けた。

今回のルールでは、ある web サーバ S から、ある利用者が存在するネットワーク N への経路情報が無い場合、ネットワーク N に存在する利用者からの処理要求に対して web サーバ S を選択することはない。このことを利用して、RADIX へ提供する経路情報を制限することで、サービス提供範囲を制限することが可能である。

3.2.3 特定サイト優先選択

自サイトから、ある処理要求元が存在するネットワークへの経路情報が存在する場合、優先的に自サイトを選択して欲しいという要求を持つ web サーバ設置サイトがあった。これは、一見 3.2.2 と似ているが異なるものである。3.2.2 は「経路情報が無い」場合に選択されない、ことを述べたものであった。従って、経路情報があった場合には他のいずれかのルールで選択が行われることになる。一方本ルールは「経路情報がある」場合に積極的に選択する、というものである。

3.2.4 要求元自律システム毎選択

経路情報を調査することで、処理要求元つまり利用者が存在する自律システムを知ることができる。この要求元の自律システムに応じた選択を行うルールである。これは 3.2.3 と似ているが、静的に指定しておくという点で異なる。他のルールによる制御ではカバーできないような場合に利用した。

3.2.5 国内・海外選択

国内のみ、もしくは海外のみにサービス提供したいという web サーバ設置サイトがあった。この要求に応じるためのルールである。

判断する材料としては、3.2.4 で述べた要求元の自律システムを用いた。しかし、自律システムを元に国内か海外かを一般的に判断することはできない。今回は JPNIC が発行した自律システムかどうか⁶で判断した。

⁶<ftp://ftp.nic.ad.jp/jpnic/ipaddress/as-numbers.txt> に載っているかどうかで判断した。

3.2.6 ラウンドロビン選択

他の全てのルールでの選択に失敗した場合に用いるルールである。サーバ設置サイトによっては、想定範囲外へのサービス提供を希望しない場合があった。そこで、このルールでのサービス提供を許可したサイトのみをラウンドロビンで選択するようにした。

3.2.7 固定比率選択

このルールは、「最適」なサーバの選択を行うものではない。ある web サーバ設置サイトが、日食中継のようなイベントにおけるアクセス状況全般を調査したいという要求を持っていた。このサイトの要求を満たすためのルールが本ルールである。

このサイトの要求を完全に満たすためには、全ての処理要求を処理させる必要がある。しかし、それではこのサイトがボトルネックになるのは明らかである。そこで、固定の比率でこのサイトを選択するようにした。今回の中継においては、この比率は $\frac{1}{10}$ とした。

3.3 日食中継における選択ポリシー

上述のルールを組み合わせることで構成した選択ポリシーを疑似コードで示す。

```
if 固定比率 {
  サーバ A を選択
}

if 要求元自律システムに対するサーバが指定? {
  指定されているサーバを選択
}

if サイト B への経路が存在
  サーバ B を選択
```

```
}

if 要求元は国内? {
  国内向サイトの中から最短自律システムパス長を検索
  if 最短自律システムパス長 < 4 {
    最短自律パス長のサーバを選択
  } else {
    国内向ラウンドロビンサーバから選択
  }
} else {
  海外向けサイトの中から最短自律システムパス長を検索
  if 最短自律システムパス長 < 4 {
    最短自律パス長のサーバを選択
  } else {
    海外向けラウンドロビンサーバから選択
  }
}
```

4 実験結果

2001年6月19日 21:11 ~ 2001年6月25日 16:00 の間の TENBIN のログを集計した。この間、日食のライブストリームの中継を行ったのは6月20日 20:20 ~ 6月20日 22:35 である。この前後も、ライブストリーム以外の日食情報等のコンテンツは提供された。

3.2 で述べた各ルールが適用された数を表 1 に示す。

「経路情報無し」は RADIX が要求元への経路情報を持たなかった場合である。「通信エラー」は、TENBIN と RADIX 間の通信に失敗した場合である。

3.2.7 で述べたように、固定比率選択は最適なサーバの選択を行うものではない。従って、以下の評価においては固定比率選択を除いて検討を行う。

あり 6月19日 21:34 に修正した。

表 1: 適用された選択ルール

選択ルール	回数
自律システムパス長 (海外)	113149
自律システムパス長 (国内)	70259
ラウンドロビン (海外)	30779
固定比率	28943
特定サイト優先	22428
要求元自律システム毎	5035
ラウンドロビン (国内)	4527
経路情報無し	10393
通信エラー	4005
合計	289518

表2に、各ルールの割合を示す。表から分かるように、全処理要求の 80.9% を近隣のサーバへ誘導することができた。

表 2: 経路情報によるヒット率

選択ルール	回数	割合 (%)
自律システムパス長 (海外)	113149	43.4
自律システムパス長 (国内)	70259	27.0
特定サイト優先	22428	8.6
要求元自律システム	5035	1.9
その他	49704	19.1
合計	260575	100.0

自律システムパス長選択を行った場合について、そのパス長をまとめたものを表3、表4、表5に示す。

ここで、自律システムパス長1とは、処理要求元とwebサーバ提供サイトが同一自律システムに存在することを示す。自律システムパス長2とは、処理要求元とwebサーバ提供サイトが隣接した自律システムに存在することを示す。

表3、表5には、自律システムパス長4の場合が現れている。これはシステムのバグで

表 3: 自律システムパス長 (海外分)

パス長	回数	割合 (%)
1	189	0.2
2	55161	48.4
3	57774	50.7
4	25	0.0
合計	113949	100.0

表 4: 自律システムパス長 (国内分)

パス長	回数	割合 (%)
1	6940	9.9
2	32823	46.7
3	30496	43.4
合計	70259	100.0

表 5: 自律システムパス長 (合計)

パス長	回数	割合 (%)
1	7129	3.9
2	87984	47.8
3	88270	47.9
4	25	0.0
合計	184208	100.0

5 まとめ

我々は、TENBIN と RADIX という二つのシステムを組み合わせることで、経路情報を用いたサーバ選択機構を構築した。

本機構を、本年6月の皆既日食中継システムに適用した。本中継では、webサーバを

6つのサイトに広域分散配置した。各サイトから経路情報を収集し、この情報に基づいてサーバ選択を行った。この結果、サーバ選択の対象となった要求の80.9%という高い割合を近傍のサーバに誘導することができた。

このように、コンテンツ配布用サーバを広域に分散配置した場合、経路情報を用いたサーバ選択が効果的であることを明らかにできた。

今回の中継システムは、複数のCDNを束ねたメタCDNと考えることができる。このようなメタCDNの構築手法としてCDN peeringという手法が提案されている。今後は、メタCDNに適したサーバ選択機構に付いて検討を行う必要がある。

謝辞

本実験を行う機会を与えてくださった、ライブ!エクリプス実行委員会の皆さんに感謝致します。本研究を進めるにあたり有意義な議論をして頂いた九州ギガポップ(QGPOP)プロジェクト、インターネット技術研究委員会(ITRC)、WIDEプロジェクトの皆さんに感謝します。

なお、本研究の一部は平成13年度科学研究費補助金奨励研究(A)13780249によるものである。

参考文献

- [1] M. Day, B. Cain, G. Tomlinson, P. Rzewski, "A Model for Content Internetworking (CDI)," draft-day-cdn-model-08.txt, Oct. 2001
- [2] 下川俊彦, 吉田紀彦, 牛島和夫, "多様な選択ポリシーを利用可能なサーバ選択機

構," 電子情報通信学会論文誌, J84-D-I:9, 1396-1403, Sep. 20001.

- [3] 中川 郁夫, "RADIX - a new route server implementation," <http://www.toyama.net/%7Eikuo/radix/index.html>, August, 1999.
- [4] 下川俊彦, 木場雄一, 中川郁夫, 山本文治, 吉田紀彦, "広域分散ストリーム配信における経路情報の利用," 信学技法 Vol.101 No.176, pp.1-8, July, 2001.
- [5] Akamai Technologies, Inc., <http://www.akamai.com/>
- [6] ライブ!エクリプス実行委員会, <http://www.live-eclipse.org/>
- [7] 荻野司, 松田和宏, 須藤一顕, 針山欣之, 向阪正彦, 殖栗俊明, "広域分散配置 Web サーバにおける最適サーバ探索システムの検討", 分散システム/インターネット運用技術シンポジウム2001 論文集, pp. 41-50, Feb. 2001.