第 III 部

ネットワークトラフィック統計情報の収集と解析

第 3 部
ネットワークトラフィック統計情報の収集と解析

第 1 章　MAWI ワーキンググループについて

MAWI（Measurement and Analysis on the WIDE Internet）ワーキンググループは、トラフィックデータの収集と解析を研究対象とした活動を行なっている。

MAWI ワーキンググループでは WIDE プロジェクトの特徴を活かした研究をするため、「広域」「多地点」「長期的」の三つの項目に重点を置いたトラフィックの計測・解析を行っている。広域バックボーンでのデータ収集はバックボーンを持っている WIDE だからできる事である。分散管理されるインターネットの状態を把握するためには、多地点で観測したデータを照らし合わせることが欠かせない。また、長期的にデータを収集し蓄積するために、ワーキンググループとしての継続的な活動が役に立つ。

計測技術はほとんどの研究分野で必要となるため、MAWI ワーキンググループは WIDE 内の他のワーキンググループと連係をとりながら活動をしている。具体的には、

- グローバルな視点からの DNS の挙動解析（dns ワーキンググループと共同）
- IPv6 普及度の計測（v6fix ワーキンググループと共同）
- ネットワークトポロジの観測（netviz ワーキンググループと共同）
- 長期的な経路変動の観測（routeview ワーキンググループと共同）
- sFlow/NetFlow を使ったトラフィック計測（roft ワーキンググループと共同）
- AI3 の衛星トラフィックの計測（ai3 ワーキンググループと共同）

などが上げられる。

また、国際協調として

- CAIDA[26]
- University of Waikato[285]
- ICANN RSSAC[115]
- ISC OARC[135]
- USC/ISI[288]
- INRIA[124]

などと共同して研究活動をしている。

第 2 章　MAWI ワーキンググループ 2005 年度の活動概要

今年度の報告書では、まず第 3 章において、集約型トラフィックプロファイラを使った国際線トラフィックの傾向を報告する。このツールは、WIDE バックボーンのトラフィックをニアリアルタイムかつ長期的にモニタリングする目的で 2001 年に開発され、それ以来利用されてきている。また、急増する分散型 DoS アタックの早期検出にも役立っている。

第 4 章では、サーバ選択の安定性と新アルゴリズムの提案を行う。インターネット上ではさまざまなサービスが複数のサーバーによって提供されている。クライントは通常最寄りのサーバを選択するが、単純な選択アルゴリズムでは系全体のゆらぎを増幅する場合がある。ここでは、大規模シミュレーションを使って問題点を解析し、効率的でかつ安定なサーバ選択アルゴリズムの要件を検証し、具体的な方式を提案する。

第 5 章では、急速に普及してきたブロードバンドのトラフィック解析を報告する。国内 ISP7 社の協力によりデータを収集し、バックボーンにおいてもブロードバンドユーザのトラフィックが支配的になってきた実態や、ヘビーユーザの分布について報告する。

第 6 章では、2005 年 3 月に行った第 5 回 CAIDA/WIDE 計測ワークショップについて報告する。CAIDA と WIDE は、2003 年度から正式に計測に関する包括的な共同研究を行っている。2005 年度も CAIDA との共同研究を継続し、DNS 計測、IPv6 トポロジ計測などの共同研究を行い、人材交流も実施した。

第 3 章　WIDE 国際線のトラフィック傾向

3.1 はじめに

WIDE インターネットのような広域なネットワークを運用し続けていくためには、トラフィックモニタリングを多地点、かつ長期間行い、ネットワークの現状に適した通信機器の設置、設定を行う必要がある。

しかし、現存するネットワークモニタリングツールは長期に渡ってトラフィックの傾向を収集し続けることが難しい。

そこで、WIDE プロジェクトの mawi ワーキンググループでは収集したトラフィックを効果的に集約することによって、ネットワークの特徴を抽出することのできるトラフィックモニタリングツール AGURI[35] の設計、実装を行った。

AGURI（Aggregation-based Traffic Profiler）は、1) トラフィック中の特徴的なフロー傾向を残しつつ、2) 短期間から長期間に渡って利用可能なトラフィックモニタリングツールである。

AGURI は以下に示す 4 種類のネットワークサマリ情報を作成する。

- 送信元 IP アドレス
- 受信先 IP アドレス
- IP バージョン＋プロトコル＋送信ポート番号
- IP バージョン＋プロトコル＋受信ポート番号

この 4 種類のネットワークサマリを定期的に出力することによって、ある短時間のネットワーク状態の特徴を知ることができる。

更に、AGURI は一度 AGURI で作成したネットワークサマリからもデータを入力することができ、複数のサマリを同時に入力することもできるので、ある短時間のサマリを組み合わせて AGURI に入力することによって、可変長の時間のネットワーク状態の特徴を知ることができる。

3.2 収集データ

WIDE プロジェクトでは以下に示す 2 地点において国際線のデータを収集している。

1. samplepoint1　trans-Pacific line
   (18 Mbps CAR on 100 Mbps link)
2. samplepoint2　US-Japan line
   (Japan side 60 Mbps POS)

WIDE プロジェクトで利用している 2 本の国際線のうち、1 本は他 AS と BGPpeer を張っているポイントを WIDE インターネットの入り口でデータ収集を行っている（samplepoint1）。

他の 1 本は WIDE の利用している国際線日本側（samplepoint2）でそれぞれデータ収集を行っている。

2005 年度の WIDE 報告書では、samplepoint1 と samplepoint2 で収集した WIDE 国際線の年間トラフィック傾向を図 3.2 から図 3.33 に示す。

図の出力は時期を四半期ごとに、対象を 1) 送信元

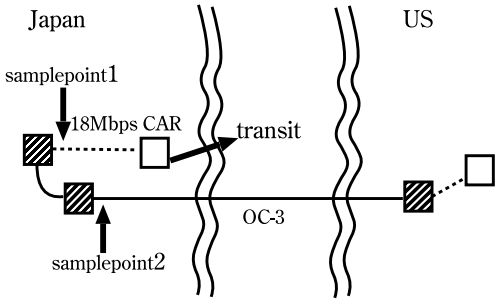

図 3.1.　データ収集地点

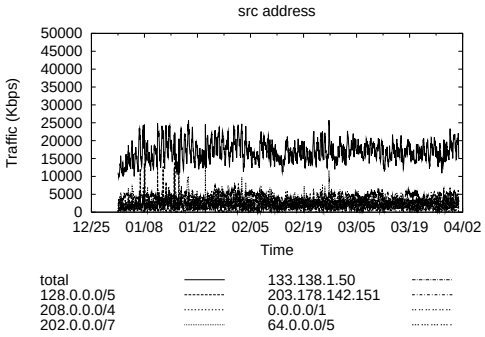

図 3.2.　送信元 IP アドレス（1 月–3 月）

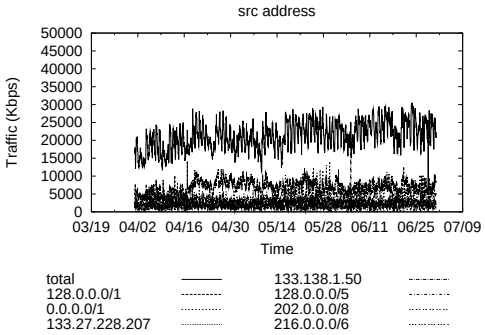

図 3.3.　送信元 IP アドレス（4 月–6 月）

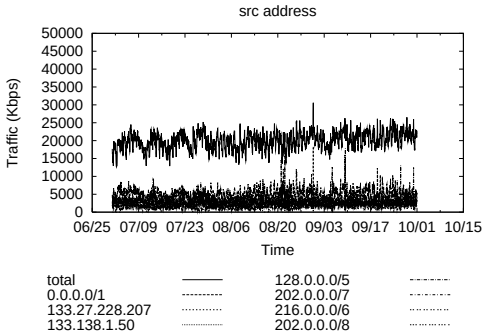

図 3.4. 送信元 IP アドレス（7 月–9 月）

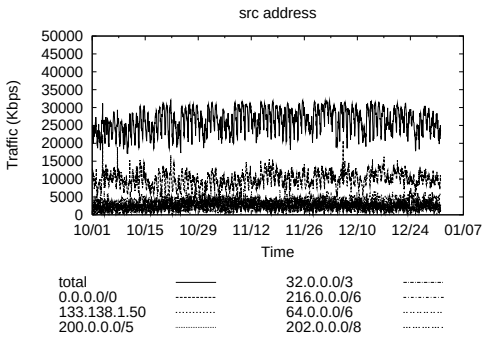

図 3.5. 送信元 IP アドレス（10 月–12 月）

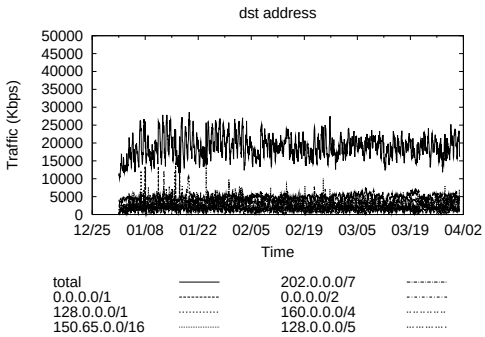

図 3.6. 宛先 IP アドレス（1 月–3 月）

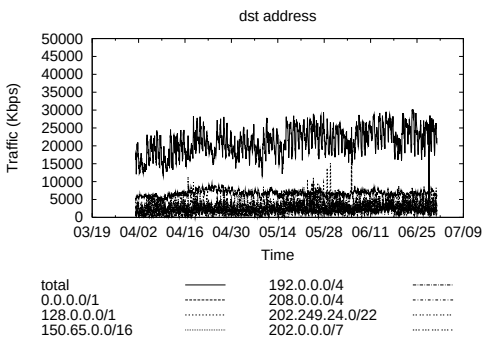

図 3.7. 宛先 IP アドレス（4 月–6 月）

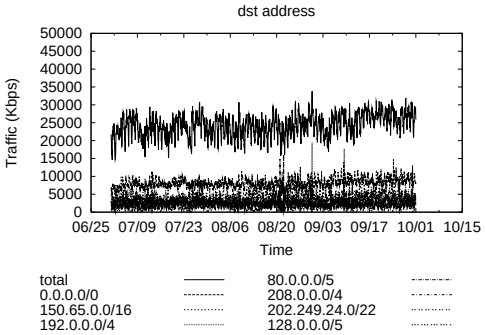

図 3.8. 宛先 IP アドレス（7 月–9 月）

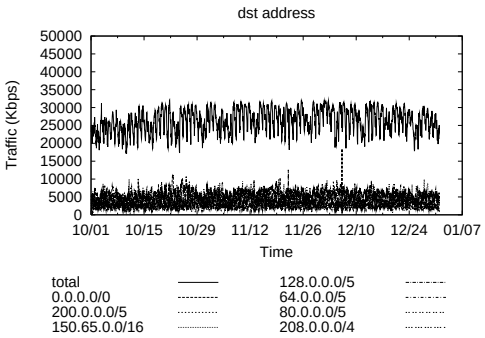

図 3.9. 宛先 IP アドレス（10 月–12 月）

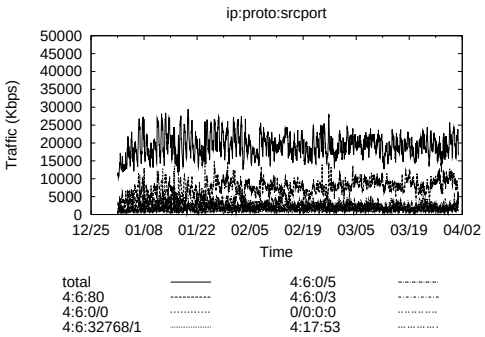

図 3.10. 送信元ポート番号（1 月–3 月）

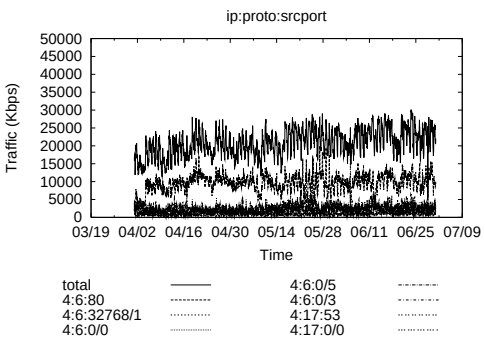

図 3.11. 送信元ポート番号（4 月–6 月）

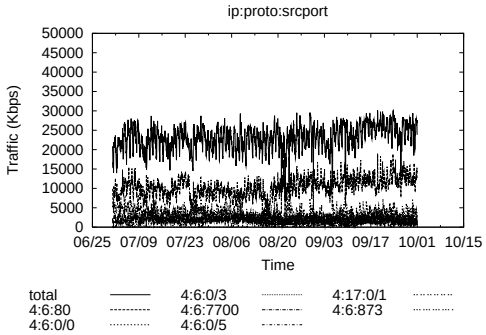

図 3.12.　送信元ポート番号（7 月–9 月）

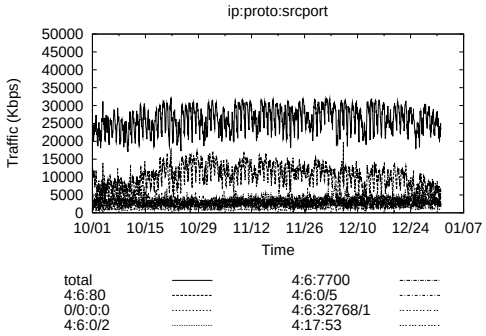

図 3.13.　送信元ポート番号（10 月–12 月）

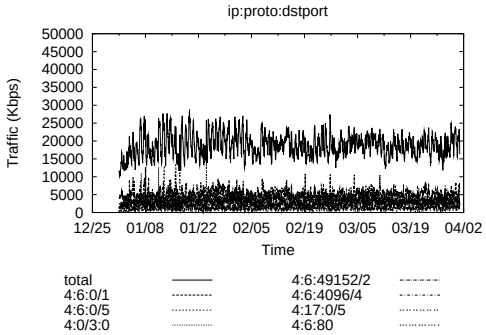

図 3.14.　宛先ポート番号（1 月–3 月）

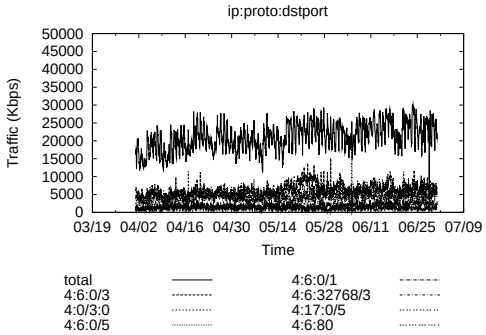

図 3.15.　宛先ポート番号（4 月–6 月）

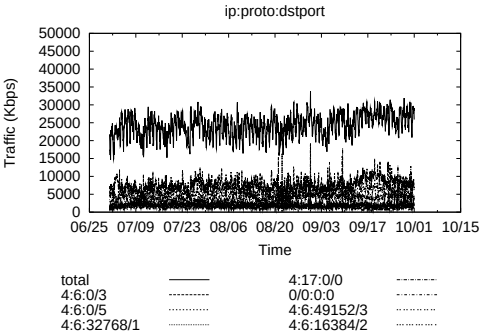

図 3.16.　宛先ポート番号（7 月–9 月）

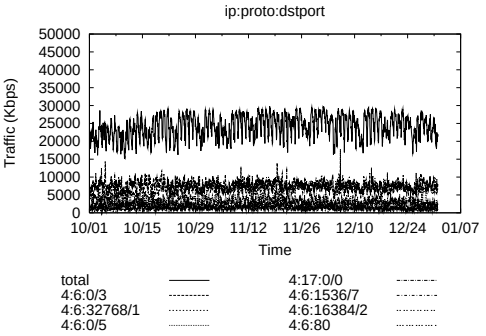

図 3.17.　宛先ポート番号（10 月–12 月）

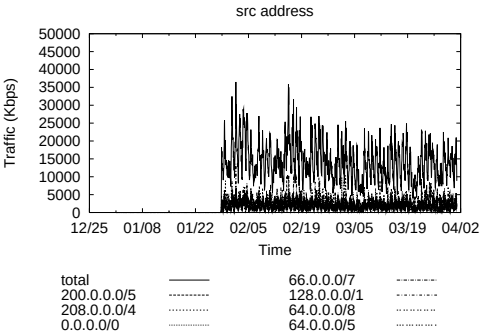

図 3.18.　送信元 IP アドレス（1 月–3 月）

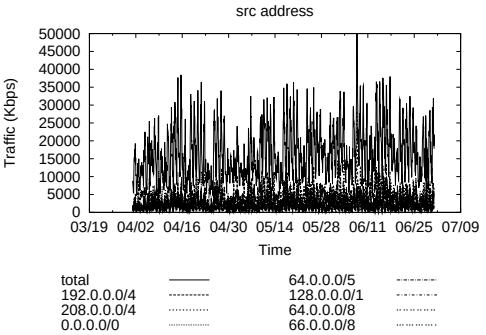

図 3.19.　送信元 IP アドレス（4 月–6 月）

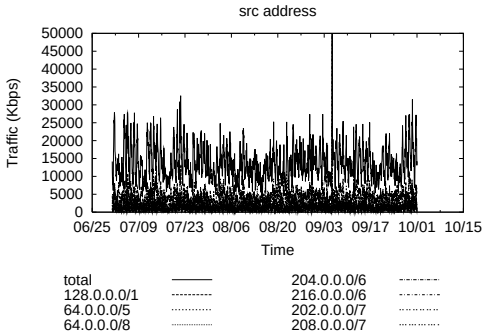

図 3.20. 送信元 IP アドレス（7 月–9 月）

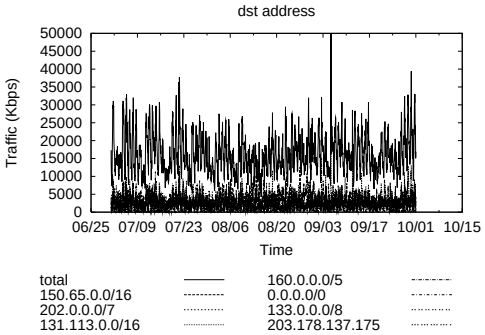

図 3.24. 宛先 IP アドレス（7 月–9 月）

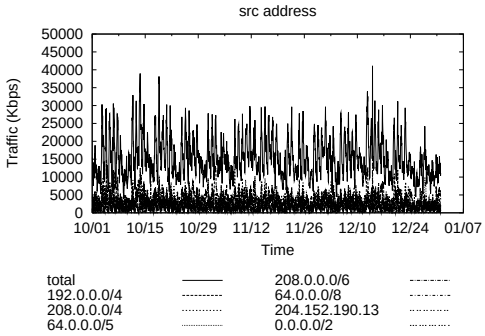

図 3.21. 送信元 IP アドレス（10 月–12 月）

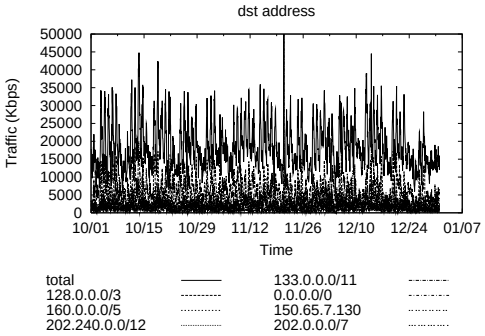

図 3.25. 宛先 IP アドレス（10 月–12 月）

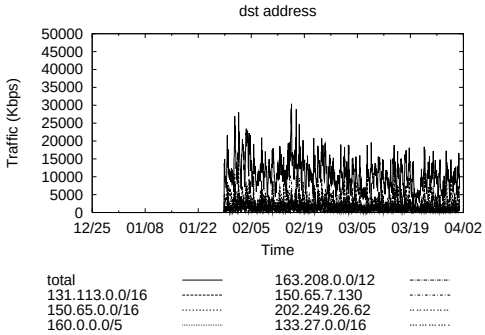

図 3.22. 宛先 IP アドレス（1 月–3 月）

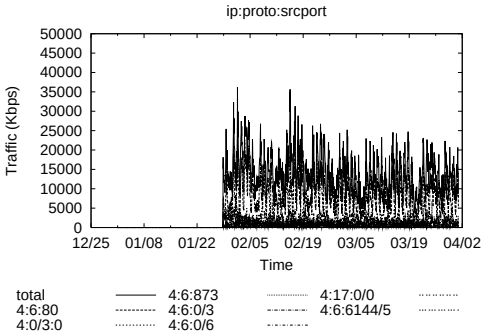

図 3.26. 送信元ポート番号（1 月–3 月）

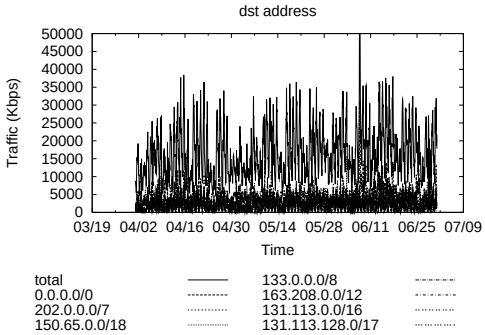

図 3.23. 宛先 IP アドレス（4 月–6 月）

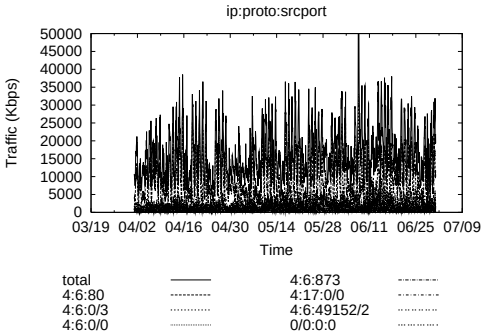

図 3.27. 送信元ポート番号（4 月–6 月）

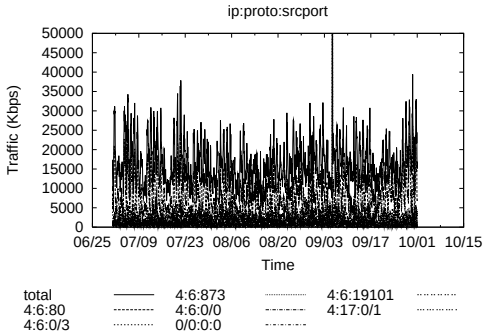

図 3.28.　送信元ポート番号（7 月–9 月）

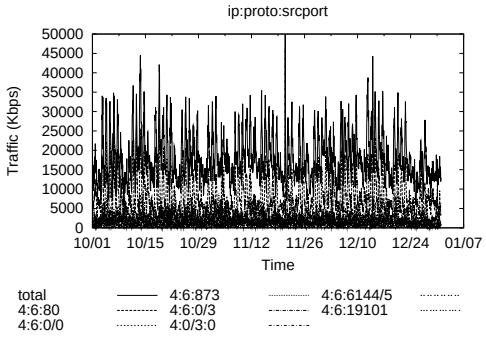

図 3.29.　送信元ポート番号（10 月–12 月）

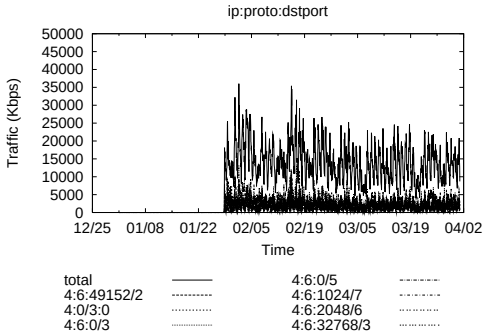

図 3.30.　宛先ポート番号（1 月–3 月）

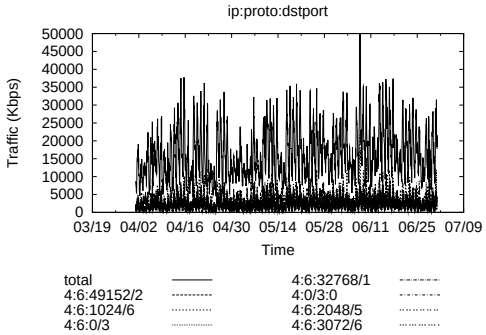

図 3.31.　宛先ポート番号（4 月–6 月）

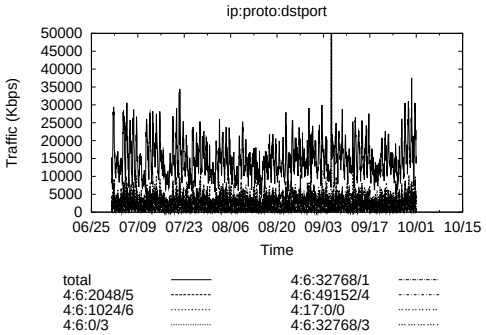

図 3.32.　宛先ポート番号（7 月–9 月）

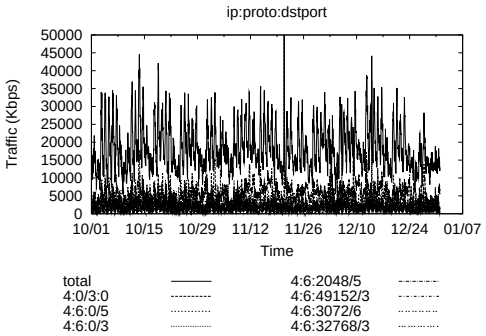

図 3.33.　宛先ポート番号（10 月–12 月）

IP アドレス、2) 宛先 IP アドレス、3) 送信元ポート番号、4) 宛先ポート番号とする。（表 3.1、表 3.2）

図 3.2 から図 3.33 に示された長期的トラフィック傾向から抽出できた情報を表 3.3、表 3.4 に示す。

図中に出て来る “4:6:80” とは IP バージョンが 4、プロトコル番号が 6（つまり TCP）、送信元ポート番号が 80（つまり HTTP）ということを示している。

ここに示した図は 2 つの情報を持っている。

- 折れ線グラフ

  回線を占めているトラフィックの属性を視覚的に見ることができる。

  今回取り上げた WIDE インターネット国際線の例では、全トラフィック量の推移と HTTP データの割合を把握できる。

- 項目

  折れ線グラフの下にリストアップされる項目数は、AGURI によって設定することができる。この項目は全トラフィック中の占有率順にリストアップされるため、回線を使用している組織や使われているアプリケーションを検知することができる。

  送信元、宛先 IP アドレスからは、特定の組織の

表 3.1. トラフィック傾向一覧表（samplepoint1）

	1 月–3 月	4 月–6 月	7 月–9 月	10 月–12 月
送信元 IP アドレス	図 3.2	図 3.3	図 3.4	図 3.5
宛先 IP アドレス	図 3.6	図 3.7	図 3.8	図 3.9
送信元ポート番号	図 3.10	図 3.11	図 3.12	図 3.13
宛先ポート番号	図 3.14	図 3.15	図 3.16	図 3.17

表 3.2. トラフィック傾向一覧表（samplepoint2）

	1 月–3 月	4 月–6 月	7 月–9 月	10 月–12 月
送信元 IP アドレス	図 3.18	図 3.19	図 3.20	図 3.21
宛先 IP アドレス	図 3.22	図 3.23	図 3.24	図 3.25
送信元ポート番号	図 3.26	図 3.27	図 3.28	図 3.29
宛先ポート番号	図 3.30	図 3.31	図 3.32	図 3.33

表 3.3. 識別された IP アドレス

graph	IP アドレス	hostname
図 3.3、3.4	133.27.228.207	saba.w3.mag.keio.ac.jp
図 3.2–図 3.5	133.138.1.50	csl.sony.co.jp 配下のホスト
図 3.2	203.278.142.151	download.sfc.wide.ad.jp
図 3.21	204.152.190.13	ftp.netbsd.org
図 3.22–図 3.25	150.65.7.130	ftp.jaist.ac.jp
図 3.18	202.249.26.62	cache.unibraw.ai3.net
図 3.19	203.178.137.175	ftp.nara.wide.ad.jp

表 3.4. 識別されたポート番号

graph	ポート番号	プロトコル／アプリケーション
図 3.10–3.13、図 3.14、3.15、3.17、図 3.26、3.31、3.33	4:6:80	HTTP
図 3.10、3.11、3.13	4:17:53	DNS
図 3.12、3.13	4:6:7700	BitTorrent
図 3.12、図 3.26–3.29	4:6:873	rsync

IP アドレス空間と特定のホストを検出できた。特に 2004 年度の WIDE 報告書と比較した場合、2004 年度に観測された ai3.net、jaist.ac.jp、keio.ac.jp を宛先としたトラフィックを引き続き抽出できた。それに加え、csl.sony.co.jp など AS2500 番に接続されている組織のアドレスブロックを抽出できた。

また、特定のホストにトラフィックが集中している様子も観察できた。抽出された‘133.27.228.207’という IP アドレスは、慶應義塾大学院に設置されているプロジェクトのサーバである。download.sfc.wide.ad.jp、ftp.jaist.ac.jp、ftp.nara.wide.ad.jp などの WIDE プロジェクト内に設置されている公開 FTP サーバのトラフィックを観測できた。‘202.249.26.62’という IP アドレスはインドネシア BRAWIJAYA 大学に設置されている web キャッシュサーバである。

送信元ポート番号からは、特定のポートを使用したアプリケーションを検出できた。

今年度も引き続き HTTP トラフィックの観測に加えて、rsync トラフィックの観測もできた。また、今年度から BitTorrrent という Free Speech Tool のトラフィックを検出できた。

以上のように、折れ線グラフで表された情報とリストアップされた項目から、全トラフィックを構成

している特徴的な要素を抽出することができた。

3.3 結論

本節では、AGURI を用いた WIDE インターネット国際線のトラフィック傾向を述べた。

WIDE インターネットのような広域なネットワークを運用し続けていくためには、トラフィックモニタリングを多地点、かつ長期間に行い、ネットワークの現状に適した通信機器の設置、設定を行う必要がある。

しかし、現存するネットワークモニタリングツールは長期に渡ってトラフィックの傾向を収集し続けることが難しい。

WIDE プロジェクトの mawi ワーキンググループでは収集したトラフィックを効果的に集約することによって、ネットワークの特徴を抽出することのできるトラフィックモニタリングツール AGURI を用い長期に渡る国際線のトラフィック傾向を明らかにした。

実際に AGURI を用いて WIDE インターネット国際線でデータを収集し、対象とした国際線のトラフィックの傾向を明らかにした。

WIDE プロジェクトでは、AGURI の開発をすすめると共に、WIDE インターネットのバックボーンにおいて AGURI を運用し続けている。これらのデータは `http://mawi.wide.ad.jp/mawi/` から参照可能である。

第 4 章　ネットワーク変動に対するサーバ選択アルゴリズムの安定性について

現在サーバ-クライアントモデルの通信ではベストサーバセレクションが広く利用されている。しかし、ベストサーバセレクションは効率はよいがネットワーク変動に非常に弱い。また既存のレシプロカルアルゴリズムはネットワーク変動に強いが効率が悪い。

我々は、既存のアルゴリズムの性質を検証するためにシミュレーションを行った。また、問題を視覚的に認識するためネットワーク変動がおこった際のサーバ負荷の変化を可視化した。

この結果、レシプロカルアルゴリズムは、ある程度の確率でコストの大きなサーバを選択するため、性能が低下することがわかった。そこで、我々は、性能のよい少数のサーバから構成されるワーキングセットを選択し、そのなかから、サーバを一定の確率で選択する 2 ステップの方式を提案する。この方式がサーバのコスト変動への適応性、負荷分散、スケーラビリティ、効率の面において非常に優秀であることをシミュレーションにより示した。

4.1 背景

現在サーバクライアント型サービスが広く利用されており、この型のサービスでは、クライアントはサーバを何らかの方法で決定し、選択されたサーバへリクエストを送る。

サーバ選択アルゴリズムとして、ベストサーバセレクションが広く利用されているが、サーバの負荷が偏りがちである。サーバの負荷の偏り自体は効率を考えると仕方がなく、サーバ配置の観点からは負荷の高いサーバがはっきりわかることでサーバの増強や追加を行うことで負荷を分散できる可能性がある。

しかし、負荷の偏りがあった場合にネットワーク変動が起きるとクライアントが一斉に移動し、クライアントが再選択するサーバも集中した場合、さらなるネットワーク変動の原因となる可能性がある。

ベストサーバでは上記の問題が顕著にあらわれ、他の手法はサーバの負荷をある程度分散できるが効率面に問題がある。

本報告書では、ネットワーク変動に強く、効率がよいサーバ選択アルゴリズムを提案する。

4.2 従来のサーバ選択アルゴリズム

本節では、従来から利用されている一般的なアルゴリズムについて説明、考察する。

4.2.1 ベストサーバセレクション

ベストサーバセレクションはホップ数や RTT などのコストから、利用できるサーバ中で最適なものを選択する方式である。

クライアントは最も効率のよいサーバを一意に選択するため効率は常に最良である。

しかし、問題点としてネットワーク変動の増幅や、振動の原因となることが上げられる。

4.2.2 ユニフォームセレクション

ユニフォームセレクションはコストに関わらず無作為に利用可能なサーバからサーバを選択する方式である。

この方式では、負荷の集中が起こらず、それに起因する問題も起こらない。

しかし、あるクライアントからのサーバ性能を考慮しないため性能は悪い。

4.2.3 レシプロカルセレクション

サーバ性能のコストの逆数に比例する確率でサーバを選択する方式である。サーバを選択する確率をきめる関数によりその挙動は変化する。

性能の低いサーバもある程度の確率で選択されるため、サーバの集中をある程度に押さえられ変動にも強い。

一方、コストが悪いサーバも選択されることから性能もベストサーバに比べてよくない。

4.2.4 実例：DNS の場合

DNS[188, 189] は比較的少ないサーバ群から実際に問い合わせをするサーバを選択する。

DNS の実装として広く利用されている、BIND[134] バージョン 8 および 9 で利用されているアルゴリズムはレシプロカルセレクションの一種と考えられる。また、DJBDNS[50]、Microsoft Windows Internet Server はユニフォームセレクションを採用している。

BIND で利用されている実践的手法は、多くの場合において効率を向上させるが、検討の余地がある。クライアント群からみてコストの差が小さな 2 台のサーバが存在する場合に、サーバ負荷が大きく片寄る可能性があることや、多くのサーバを利用するサービスには適当でないからである。これについては [248] で述べられている。

4.3 既存の手法のシミュレーションによる評価

我々は、既存の手法の評価を行うため、ある程度の規模のネットワークでコスト変動が起こるシミュレーションを行なった。

4.3.1 シミュレーショントポロジおよびシミュレーション内容

サーバ負荷に偏りがあるトポロジを機械的に生成するために、スケールフリーなネットワーク構造をもとにしたトポロジ生成を行った。トポロジ構築のためのルールを以下に示す。

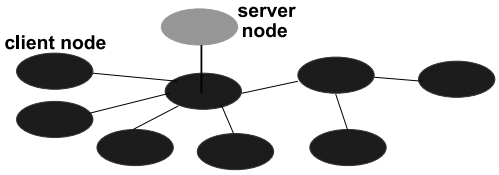

図 4.1. Concept of the simulation topology.

- 1 台目のノードを設置
- 2 台目のノード以降は既存のノードを 1 台選択し、そのノードに接続するように設置する。2 台目以降のノードは生成時に既存のノードに接続されるが、このノードは既存のノードからのエッジの数に比例した確率で選ばれる。エッジを多く持つノードはより高い確率でさらに多くのエッジを得る。
- ノードを 10 台設置するごとにサーバを設置する。
- サーバはその時点でエッジを最も多く持つノードにのみ接続される。選択したノードにすでにサーバが接続されていた場合は、次にエッジが多いノードに接続する。
- サーバのクライアント数が 20 を超えるとノードを 2 つに分割し、サーバを新たなノードにサーバを新たに接続する。ノードに接続されていたエッジの半数は新たなノードに接続しなおされる。
- ノードを 100 台設置するごとにその時点でのエッジの数に比例して、2 台のノードを選択し、その 2 台を接続するエッジを生成する。すでにエッジで接続されていた場合は、改めて選択しなおす。
- ただしトポロジ生成時はベストサーバセレクションによりクライアント数を算出することとする。

図 4.1 にサーバを設置する際の概念を示す。

ある程度大規模なネットワークを構築するため 500 ノードを設置した。サーバの分割などにより最終的にノード数は 510、サーバは 60 台となった。

サーバコストはエッジに重みをつけることにより表現した。エッジコストの初期値は 10 とする。

このトポロジ上で各アルゴリズムのシミュレーションを行った。シミュレーションは、サーバと接続されたノード間のコストを変更し、すべてのクライアントから各アルゴリズムを用いてサーバを選択し、各種データを取得する。1 ステップごとに一台のサーバをランダムに選択し、このサーバと接続するノー

ド間のコストを $1 \leq c \leq 40$ の値に変化させる。この時 c の値はランダムで決定する。各クライアントは各ステップ毎 100 回サーバセレクションを実行し、サーバをクライアントが選んだ回数をサーバの負荷とする。以上の規則でコストを変化させ次のステップでもとに戻す。このステップの組みを 50 回、合計 100 ステップ行った。

4.3.2 視覚化

サーバ負荷を視覚でとらえることにより各アルゴリズムでの負荷集中の度合いをわかりやすくし、コストが変動した場合のサーバ負荷の移動や新たなサーバの設置によるサーバ負荷の変化を認識しやすくする。

トポロジの描画には Tulip[284] を利用した。Tulip はグラフを視覚化するためのツールであり、ノードの表示位置決定および表示を行う。

Tulip を用いて視覚化したトポロジを図 4.2 に示す。図中の青いノードはクライアントを示し、それ以外の色のノードはサーバが設置されたクライアントであり、その色でサーバ負荷を表している。緑色のサーバの負荷は 600 未満、黄色のサーバ負荷は 600–1099、オレンジのサーバ負荷は 1100–1599、そして赤色は 1600 以上の負荷をもつサーバである。

サーバ負荷の違いを色の変化で示したことで、サーバ負荷の移動が目で捕らえられるようになった。

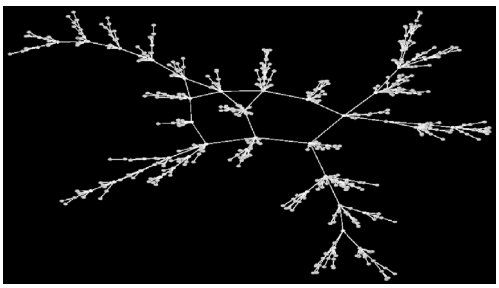

図 4.2. The simulation topology.

4.3.3 ベストサーバセレクション

図 4.3 にベストサーバセレクション利用時の各クライアントから選択されたサーバまでのコストの平均と最大値を示す。また、図 4.4 にステップ毎の各サーバに接続しているクライアント数を示す。

図 4.3 から、平均コストは 22 程度となり、ノード間のコストが 10 で固定であることを考えると非常に効率がいいことがわかる。しかし、図 4.4 から、負荷が集中しているサーバ周辺のコストが変動した際

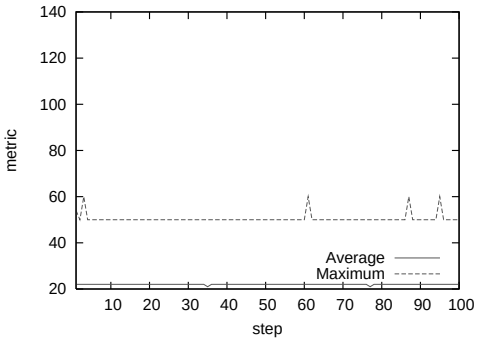

図 4.3. Average and maximum costs of best-server algorithm.

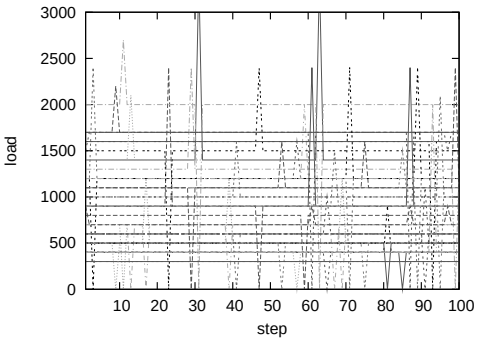

図 4.4. Server load of best-server algorithm.

に多くのクライアントが別の一台のサーバに移動していることがわかる。

新たにサーバを追加しても最も性能がよいサーバへ負荷が集中するため、サーバの新規追加では負荷分散が困難である。

これらの結果は我々の予想および、実インターネット上で観測されている状況と同一である。

4.3.4 ユニフォームセレクション

図 4.5 にユニフォームセレクション利用時の各クライアントから選択されたサーバまでのコストの平均と最大値を、図 4.6 にステップ毎の各サーバに接続しているクライアント数を示す。

ユニフォームセレクションではコスト平均が 77 程度とベストサーバセレクションの約 3.5 倍の値を示しているが、サーバの負荷は非常に平均的であり、コストの変動が起きても、クライアントの移動はほとんどみられない。性能は期待できないが、サーバ負荷の変動が小さいという事前の考察を裏付ける結果が得られたといえる。

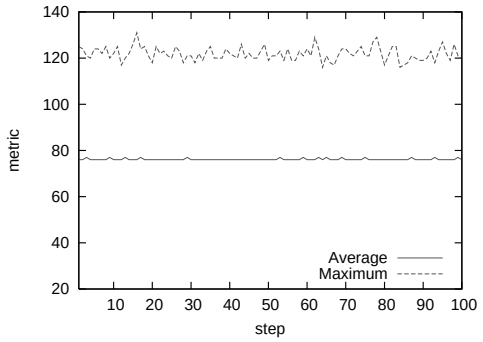

図 4.5. Average and maximum costs of uniform algorithm.

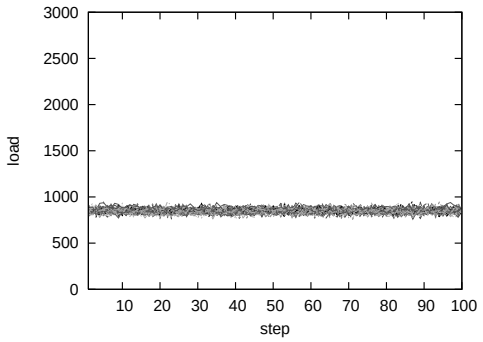

図 4.6. Server load of uniform algorithm.

4.3.5 レシプロカルセレクション

レシプロカルセレクションでは、利用するアルゴリズムを決定する必要があり、我々はコストを c として $1/c$ と $1/c^2$ の関数を利用した。

4.3.5.1 関数 $1/c$ を利用したレシプロカルセレクション

関数 $1/c$ を利用したときの、最大値と平均値を図 4.7 に、各ステップ毎のサーバに接続するクライアント数を図 4.8 に示す。平均値は 67.5 程度であり、ベストサーバセレクションの約 3 倍、ユニフォームセレクションの 14%程度の性能向上しか望めない。一方サーバ変動に関してはユニフォームサーバセレクションよりは、大きいがベストサーバセレクションと比べれば十分小さいといえる。

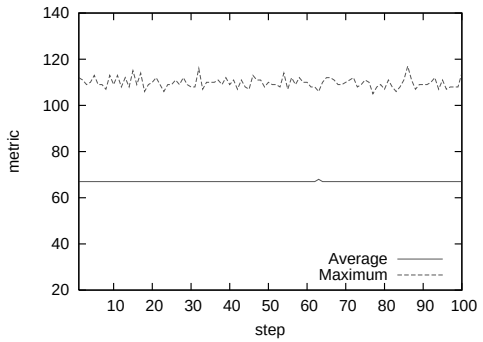

図 4.7. Average and maximum costs of reciprocal algorithm ($1/cost$).

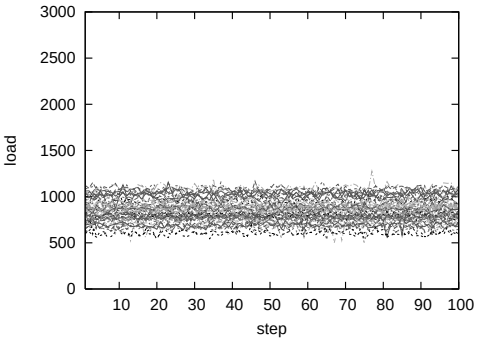

図 4.8. Server load of reciprocal algorithm ($1/cost$).

4.3.5.2 関数 $1/c^2$ を利用したレシプロカルセレクション

関数 $1/c^2$ を利用したときの、最大値と平均値を図 4.9 に、各ステップ毎のサーバに接続するクライアント数を図 4.10 に示す。

コストがより小さいサーバを選択する確率が 2 次関数的に向上するため、$1/c$ を利用した場合よりも、よりベストサーバセレクションに近い挙動を示す。しかし平均値は 55.5 程度とベストサーバセレクションの値の 2 倍のコストがかかっている。一方サーバの負荷の集中については、あるサーバのコストが大きく変化しても、負荷は複数のサーバに分散し吸収されるため、別のサーバに大きな影響を与えることはない。

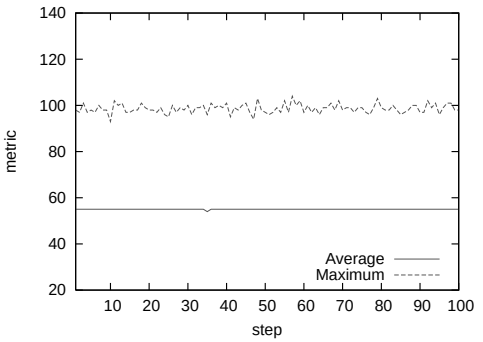

図 4.9. Average and maximum costs of reciprocal algorithm ($1/cost^2$).

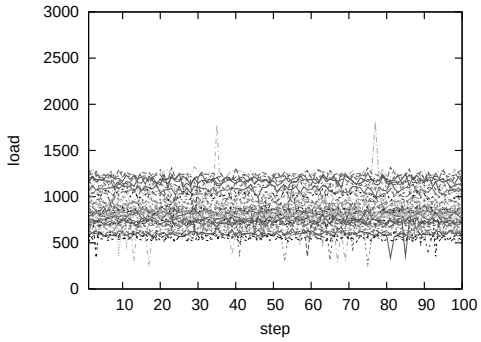

図 4.10. Server load of reciprocal algorithm $(1/cost^2)$.

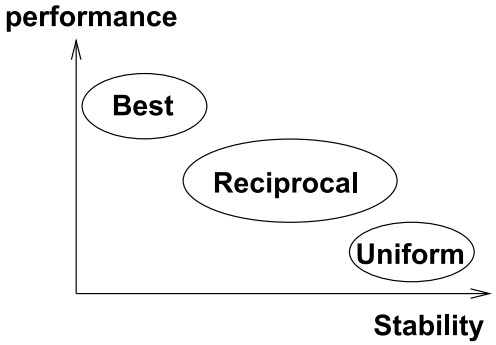

図 4.11. 各サーバセレクション手法の特性

4.3.6 各手法の比較と考察

これまでの考察およびシミュレーションで確認できた各サーバセレクション手法の特性を図 4.11 に示す。ベストサーバセレクションを利用した場合は、各クライアントが最もコストが小さいサーバを一意に決定するため、サービスの提供を受けられるまでの時間が非常に小さい。しかし、負荷が集中しているサーバ周辺でネットワークの変動があると、その負荷を別の一台または少数のサーバで吸収することになるため、別のサーバへの影響が非常に大きいといえる。また、クライアント群からみた最良のサーバが一意に決まるため、サーバ追加で思うような負荷分散をすることが難しい。

ユニフォームセレクションを利用した場合は、サービスの提供を受けられるまでの時間は非常に長いが、ネットワーク変動が起きても、影響があるサーバの数は非常に少なくてすむ。ユニフォームセレクションを利用している場合は、サーバをクライアントが集中している場所に置いても性能向上ができない。

レシプロカルセレクションは上記 2 手法の中間に位置する手法であるが、存在するすべてのサーバを利用するため、ある程度の確率でコストの大きなサーバを選択することになり、ベストサーバセレクションと比べ効率は低い。

これらの手法の考察から、一般的にネットワーク変動を伝搬させないためには、複数のサーバから確率的に利用するサーバを選択すればよいが、従来の手法では十分な性能が期待できないということが言える。

次節ではこの問題について考察し、あらたな選択手法を提案する

4.4 2 ステップサーバセレクション

ネットワーク変動が起きてもほかのサーバに大きな影響をおよぼさないためには各クライアントは利用するサーバを一意に決定するのではなく複数のサーバの中から一台のサーバを選択をすれば十分であると考えられる。しかし、これまでの手法では性能面でベストサーバセレクションを大きく下回る。

これまでのサーバセレクションの問題点は、1 つのアルゴリズムで、性能の向上と負荷分散の性能向上を目指していたことにある。そこで我々は、性能の向上と負荷分散を行うアルゴリズムを分離した、2 段階でのサーバセレクションを提案する。

第一段階では、コストがあまりにも大きいサーバを選択しないために、まずある程度のコストで利用できる W 台で構成されるサーバセットを選択する。第二段階では、第一段階で構成した W 台のサーバセットで負荷分散を行い、実際に利用するサーバを選択する。これによりコストがある程度小さいサーバのみを分散して利用することができる。

4.4.1 ワーキングセットの選択

ワーキングセットの選択は、性能のよい少数のサーバに利用するサーバを絞ることで性能の向上のために行う。

柔軟にサーバメトリックの変動に対応するためには短い間隔でメトリックのチェックを行う必要がある。しかし、利用できるすべてのサーバのコストをある程度の間隔で検査する必要があるため、非常に大きな負荷となる。実際に利用されるサーバはサーバセットに含まれるものと、そのあとに続くコストが小さなサーバ群であるため、コストが小さなサーバほどチェック間隔を短くすることでそのコストを低減する。

存在するサーバの数を N とすると、最も性能のよいサーバのランクは 1 となり、最も性能の悪いサーバのランクは N となる。ここでランクを i とすると、サーバ i の確認間隔 $q(i)$ は以下の式で表される。

$$q(i) = \frac{C}{i}$$

ここで C を変更することで耐規模性が変化する。サーバ N 台分のコストの合計確認回数 $Q(N)$ は以下の式で表される。

$$Q(N) = \sum_{i=1}^{N} q(i) = C \sum_{i=1}^{N} \frac{1}{i}$$

N が大きくなるほど曲線は水平に近くなるため、N が大きくなっても十分な耐規模性をもつといえる。また、ワーキングセットに含まれるサーバのコストチェックがもっとも回数が多いが、実際にリクエストを出す際にコストを確認すればさらに負荷を軽減できる。

4.4.2 ワーキングセットからのサーバ選択

ワーキングセットからサーバを確率的に選択することで、サーバ負荷を分散する。これによりネットワーク変動があったときの影響を小さくできる。我々のアルゴリズムでは、サーバセットからのサーバ選択にレシプロカルサーバセレクションを採用した。

4.4.3 シミュレーション結果

前述のトポロジ上で我々の提案する手法の性能評価を行った。本シミュレーションでは、W の値は 4 とした。本方式では、サーバセットに含まれていたサーバに問題がおこれば次点のサーバと入れ替えが起こるため、サーバセットは小さくてよいためである。シミュレーションでのサーバコストには、round-trip time（rtt）を利用した。サーバの処理時間やネットワークのディレイを含み、実際にクライアントがサービスをうける時間であるためである。また、一時的な変動の影響を排除するため、smoothed rtt（srtt）を利用した。srtt は以下の式で計算される。

$$srtt = \alpha \times srtt + (1 - \alpha) \times rtt$$

α には一般的に 0.7–0.8 が利用されるが迅速にコストの変化に適応するため今回は 0.6 という値を用いた。

ワーキングセット中からのサーバ選択のためのレシプロカルアルゴリズムには、サーバの負荷が偏り過ぎないよう $1/c$ のアルゴリズムを採用した。

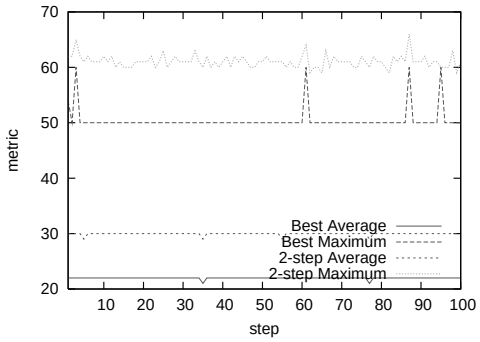

図 4.12. The maximum and average values of the 2-step algorithm compared with the best-server algorithm.

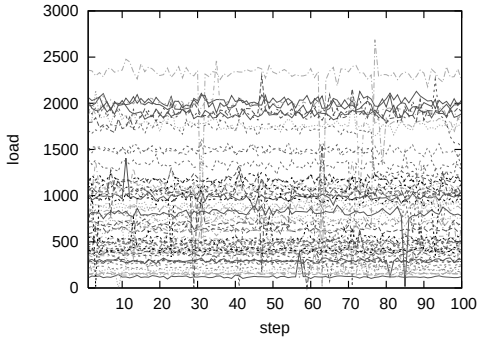

図 4.13. The server load of the 2-step algorithm.

ベストサーバセレクションと提案手法の最大値と平均値を図 4.12 に、各ステップごとのサーバに接続するクライアント数を図 4.13 に示す。図 4.12 より、平均値は 30 程度であり、ベストサーバセレクションと比較して、性能低下を 36%ほどに押さえられていることがわかる。また、図 4.13 から、サーバへの負荷集中はある程度見られるものの、ある程度の性能を求めると負荷の集中は避けられない。負荷が集中しているサーバ周辺のコストが変動した際に、その負荷がごく少数のサーバに移動することによる影響を避けることの方が重要である。ただし、本手法の負荷の偏りは、ベストサーバセレクション利用時と比べて小さい。

図 4.14 は 575 ステップ前後で、あるクライアントから見た性能が高い、7 台の負荷をプロットしたものである。532 ステップ目でサーバ 5 の負荷が激減しているが、この時サーバ 5 と接続されているノード間の $srtt$ が 10 から 38 へ大幅に増加している。こ

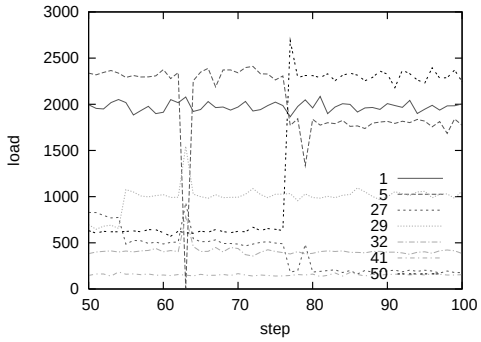

図 4.14. The loads of 7 servers.

れにより、サーバ 5 に接続していたクライアント群は別のサーバへ移動しているが、一台のサーバへ移動するのではなく複数のサーバへ移動しているため、サーバ 5 以外のサーバの負荷はサーバ 5 の負荷ほど大きく変化しない。

また 578 ステップ目では、サーバ 50 のコストが 10 から 3 へ変化し、サーバセット外であったサーバ 50 がサーバセット中に移動し、負荷が急激に上昇しているが、この負荷減少分も複数のサーバに分散しているのがわかる。

4.5 考察

今回のシミュレーションの妥当性をシミュレーションで利用したトポロジとサーバまでのコスト変動について考察する。

4.5.1 トポロジ

トポロジは 4.3.1 節に記述した手順に従い自動的に生成した。この生成手順で構築したトポロジには最終的に、510 台のノードと 60 台のサーバから構成されている。

構築されたトポロジは、多くのノードに接続されたノードがリングをつくり、その他のノード群がそこに接続されている。サーバはリング周辺に密に分布しており、実際のトポロジに近いといえる。

4.5.2 サーバコスト変動

シミュレーション部分は、初期状態からさまざまなサーバのコストを変化させ、その影響を確認するものである。

本シミュレーションでは、サーバは常にクライアントから要求された処理を行えるだけの性能を持っており、あるサーバを利用するクライアントの数が増えても、負荷に影響をおよぼさない点で実環境とは異なるといえる。実環境ではサーバの負荷が増大するにしたがって性能が低下する場合、性能低下は *rtt* が大きくなるように見える。あるサーバの *rtt* が増大するモデルは今回のシミュレーションそのものであり、シミュレーションの結果からこのような状態が起きた際の各アルゴリズムの傾向は予想できる。

4.6 まとめ

現在広く利用されているベストサーバセレクションでは、負荷が集中しているサーバ周辺のネットワークに変動がおこった際に、クライアントが別の少数のサーバに再接続することにより、更なるネットワーク変動の原因となる可能性となる問題点を指摘した。また、既存の手法のサーバ負荷の変化を視覚的に把握するためにサーバ負荷を色の変化で示すようなトポロジを表示するシミュレーションを行った。その上で、従来のサーバセレクションアルゴリズムについて考察を行い、ベストサーバセレクションに近い性能を持ちつつ、負荷が集中するサーバ周辺のネットワーク変動が起きた場合でも、別の複数のサーバによりその影響を吸収できる 2 ステップアルゴリズムを提案した。

サーバの負荷の偏りを小さくすれば、サーバリソースを適切に利用でき、ネットワーク変動にも強くなるが、性能が低下する。ある程度の負荷の偏りは許容し、ネットワークが変動した際に、負荷が集中していたサーバの負荷をいかに分散して吸収するかが重要である。また、負荷の偏りは新たなサーバの追加や、サーバの再設置などで解消できる。

2 ステップアルゴリズムでは、サーバの負荷の偏りはある程度あるが、ネットワーク変動が起きたときにも、影響があったサーバの負荷をある程度分散して吸収できる。また、ベストサーバセレクションと異なり、各クライアントが利用するサーバを一意に決定しないため、ある程度の負荷分散効果も期待できる。

2 ステップアルゴリズムでは、サーバ負荷の偏りはある程度認められるが、各クライアントがサーバを一意に決定しないため、ある程度の負荷分散でき、ネットワーク変動が起きたときでも、影響があったサーバの負荷を分散吸収することができる。つまりサーバのコスト変動への適応性、負荷分散、スケーラビリティ、効率を実現した。

今後はネットワークトポロジ、サービスとサーバセレクションの一般的な関係について議論する。また、クライアントによるサーバセレクションとサーバプレイスメント問題には関連性があるはずなのでそれについて検証する。

第5章 The impact of residential broadband traffic on Japanese ISP backbones

(This report appeared in ACM SIGCOMM CCR SPECIAL ISSUE: Measuring the internet's vital statistics. vol.35(1), pp.15–22, January 2005. The title of this paper is "The Impact of Residential Broadband Traffic on Japanese ISP Backbones".)

5.1 Introduction

The availability of residential broadband access has made tremendous advances over the past few years, especially in Korea and Japan where both the penetration rate and the average line speed are much higher than other countries. A government survey shows that there are 14.5 million broadband subscribers in Japan as of February 2004; 11 million DSL subscribers, 2.5 million CATV Internet subscribers, and 1 million FTTH subscribers[1]. The number of broadband access subscribers is still increasing as shown in Figure 5.1[1]. At the same time, broadband access technologies are shifting to higher speed such as 50 Mbps DSL and 100 Mbps FTTH.

As residential broadband access becomes widespread, we are observing an unprecedented traffic increase on commercial backbone networks. Figure 5.2 shows the aggregated peak traffic at major IXes (JPNAP[195], JPIX[141], and NSPIXP[205]) in Japan, and illustrates the growth in backbone traffic[1]. The impact of residential broadband traffic is not only in volume but also in usage patterns. The peak hours have shifted from office hours to evening hours, and emerging file sharing or other

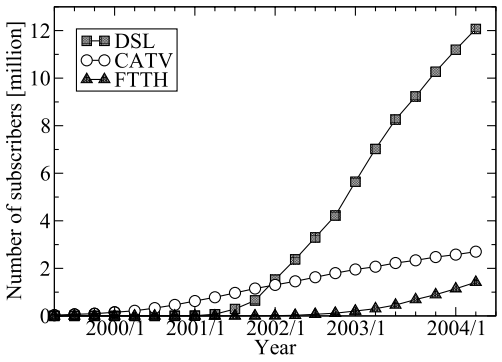

Fig. 5.1. Increase of residential broadband subscribers in Japan.

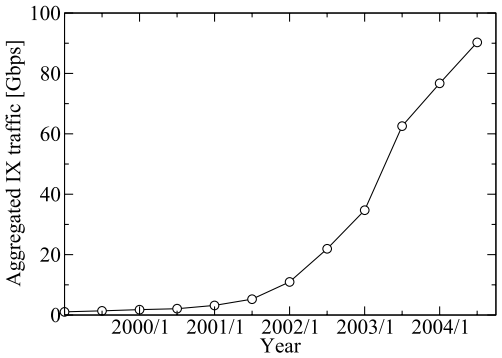

Fig. 5.2. Traffic growth at the major Japanese IXes.

peer-to-peer communications with audio/video contents exhibit behavior considerably different from traditional world wide web[99, 250]. There are striking differences in traffic patterns from earlier observations[81, 85, 87, 176, 245].

Although a drastic change in backbone traffic has already been observed, it is difficult to plan for the future because residential broadband traffic is undergoing a transformation; new innovations in access networking technologies continue to be developed, and new applications as well as their usage are emerging to take advantage of low-cost high-speed connectivity.

There is a strong concern that, if this trend continues, Internet backbone technologies will not be able to keep up with the rapidly growing residential traffic. Moreover, commercial ISPs will not be able to invest in backbone networks simply for low-profit residential traffic.

It is critical to ISPs and policy makers to

understand the effects of growing residential broadband traffic but it is difficult both technically and politically to obtain traffic data from commercial ISPs. Most ISPs are collecting traffic information for their internal use but such data contain sensitive information and are seldom made available to others. In addition, measurement methods and policies differ from ISP to ISP so that it is in general not possible to compare a data set with another set obtained from a different ISP.

In order to seek out a practical way to investigate the impact of residential broadband traffic on commercial backbone networks, we have formed an unofficial study group with specialists including members from seven major commercial ISPs in Japan.

Our goal is to identify the macro-level impact of residential broadband traffic on ISP backbones. More specifically, we are trying to obtain a clearer grasp of the ratio of residential broadband traffic to other traffic, changes in traffic patterns, and regional differences across different ISPs. As the first step, we have collected aggregated bandwidth usage logs for different traffic groups. Such statistics will provide reference points for further detailed analysis, most likely by sampling methods. In this paper, we report findings in our data sets that residential broadband traffic presents a significant impact on ISP backbones.

5.2 Methodology

There are several requirements in order to solicit ISPs to provide traffic information. We need to find a common data set which all the participating ISPs are able to provide. The required workload and investment for ISPs to provide the data set should not be high. The data set should be coarse enough not to reveal sensitive information about the ISP but be meaningful enough so that the behavior of residential broadband traffic can be analyzed. It is also desirable to be able to cross-check the consistency of the results with other data sets. The data sets should be summable in order to aggregate them with those provided by other ISPs.

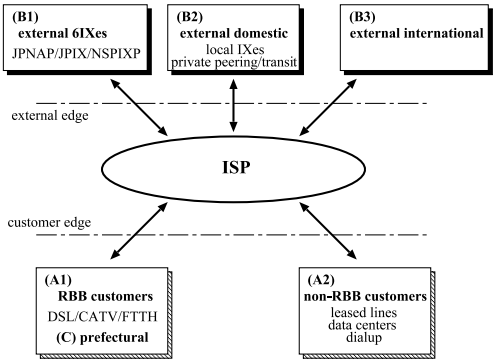

Fig. 5.3. 5 traffic groups at ISP boundary for data collection.

We found that most ISPs collect interface counter values of almost all routers in their service networks via SNMP, and archive per-interface traffic logs using MRTG[206] or RRDtool[207]. Thus, it is possible for the ISPs to provide aggregated traffic information if they can classify router interfaces into a common set.

Our focus is on traffic crossing ISP boundaries which can be roughly divided into customer traffic, and external traffic such as peering and transit. For practical purposes, we selected the 5 traffic groups shown in Figure 5.3 for data collection. The descriptions of the groups are in Table 5.1. It is impossible to draw a strict line for grouping (e.g. residential/business and domestic/international) on the global Internet so that these groups are chosen by the existing operational practice of the participating ISPs. We re-aggregate each ISP's aggregated logs, and only the resulting aggregated traffic is used in our study so as to not reveal a share of each ISP.

Our main focus is on **(A1) RBB (Residential Broadband) customers** but other items are used to understand the relative volume of (A1) with respect to other types of traffic as well as to cross-check the correctness of the results. **(A2) non-RBB customers** is used to obtain the ratio of residential broadband traffic to total customer traffic. The total customer traffic (A) is (A) = (A1) + (A2). **(B1) external 6IXes**

Table 5.1. Descriptions of traffic groups.

traffic group	description	notes
(A1) RBB customers	residential broadband customer lines	includes small business customers using RBB
(A2) non-RBB customers	includes leased lines, data centers, dialup lines	may include RBB customers behind leased lines
(B1) external 6IXes	links for 6 major IXes (JPNAP/JPIX/NSPIXP in Tokyo/Osaka)	
(B2) external domestic	external domestic links other than the 6IXes (regional IXes, private peering, transit)	domestic: both link-ends in Japan. includes domestic peering with global ASes
(B3) external international	external international links	
(C) prefectural	RBB links divided into 47 prefectures in Japan	prefectural links from 2 RBB carriers

and **(B2) external domestic** are used to estimate the coverage of the collected data sets. **(B3) external international** is used to compare domestic traffic with international traffic. The total external traffic (B) is (B) = (B1) + (B2) + (B3). **(C) prefectural** is to observe regional differences. This group covers only 2 major residential broadband carriers who provide aggregated links per prefecture to ISPs; other carriers' links are not based on prefectures. This group is a subset of (A1).

In general, it is meaningless to simply sum up traffic values from multiple ISPs since a packet could cross ISP boundaries multiple times. Customer traffic is, however, summable because a packet crosses customer edges only once in each direction, when entering the source ISP and exiting the destination ISP. The numbers for external traffic are overestimated since a packet could be counted multiple times if it travels across more than 2 ISPs. However, the error should be relatively small in this particular result since these ISPs are peering with each other.

We collected month-long traffic data that was sampled every two hours from the participating ISPs because a 2-hour resolution is the highest common factor for month-long data. This is because both MRTG and RRDtool aggregate old records into coarser records in order to bound the database size. In MRTG, 2-hour resolution records are maintained for 31 days in order to draw monthly graphs. RRDtool does not have fixed aggregation intervals but most operators configure RRDtool to maintain 1-hour or 2-hour resolution records for a period longer than needed for monthly graphs.

We developed a perl script to read a list of MRTG and RRDtool log files, and aggregate traffic measurements for a give period with a given resolution. It outputs "timestamp, in-rate, out-rate" for each time step. Another script produces a graph using RRDtool. We provided the tools to the ISPs so that each ISP can create aggregated logs by themselves. It allows ISPs not to disclose the internal structure of their network or unneeded details of its traffic.

The biggest workload for the ISPs is to classify the large number of per-interface traffic logs and create a log list for each group. For large ISPs, the total number of the existing per-interface traffic logs exceeds 100,000. To reduce the workload, ISPs are allowed to use the internal interface of a border router instead of a set of external (edge) interfaces if the traffic on the internal interface is an approximation of the sum of the external interfaces. In this case, we instruct the tool to swap "in" and "out" records since the notation in the per-interface logs depicts the perspective of the routers but inbound/outbound records in our data sets signify the ISPs' point of view.

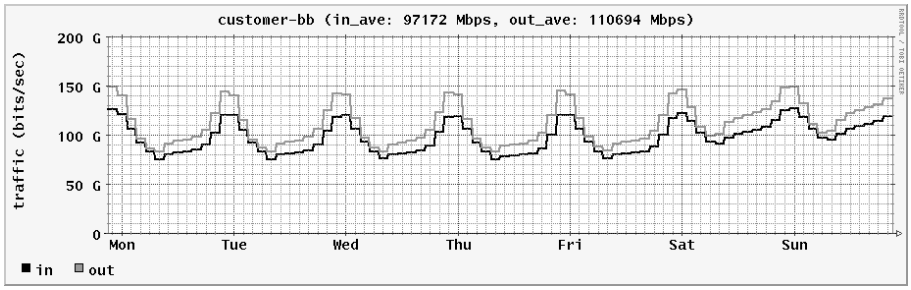

Fig. 5.4. Aggregated RBB customer weekly traffic in September 2004. Darker vertical dotted lines indicate the start of the day (0:00 am in local-time).

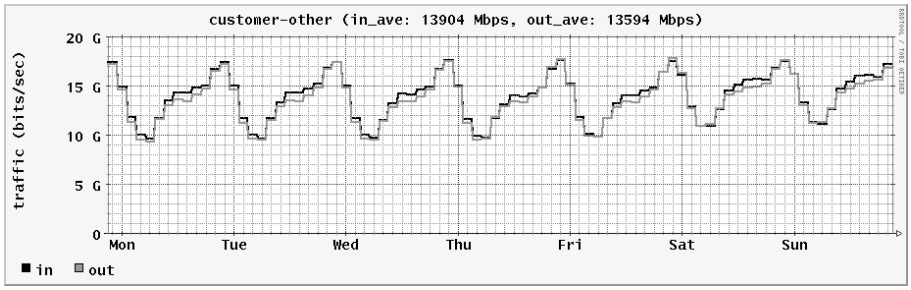

Fig. 5.5. Aggregated non-RBB customer weekly traffic in September 2004.

5.3 Results

We analyzed traffic logs for September and October in 2004 from seven major ISPs in Japan. Each ISP provided traffic logs with 2-hour resolution for those two months. The results were obtained by aggregating all the traffic logs provided by the seven ISPs. 2-hour boundaries were computed in UTC by MRTG and RRDtool so that they fell on odd hours in Japanese Standard Time that is nine hours ahead of UTC.

For weekly data analysis, we took the averages of the same weekdays in the month. We excluded two holidays in September and one holiday in October from the weekly analysis since their traffic pattern is closer to that of weekends. We also excluded another two days in October from the weekly analysis as one ISP failed to record traffic logs during this period.

5.3.1 Customer Traffic

Figure 5.4 shows the weekly traffic of RBB customers, consisting of DSL/FTTH/CATV residential users (A1). This group also includes small business customers using residential broadband access. Note that the plot is the mean rate and not the peak rate, even though the peak rate is often used for operational purposes. The residential broadband customer traffic has already exceeded 100 Gbps in total. The inbound and outbound traffic are almost equal, and about 70 Gbps is constant for both directions, probably due to peer-to-peer applications which generate traffic independent of daily user activities. The diurnal pattern indicates that home user traffic is dominant, i.e., the traffic increases in the evening, and the peak hours are from 21:00 to 23:00. Weekends can be identified by larger daytime traffic although the peak rates are close to weekdays. The outbound traffic to customers is slightly larger than the inbound, even though it is often assumed that home users' downstream traffic is much larger than upstream. We believe that peer-to-peer applications contribute significantly to the upstream traffic.

Figure 5.5 shows the weekly traffic of non-RBB customers (A2). This group contains leased lines, data centers, and other customers (e.g., dialup

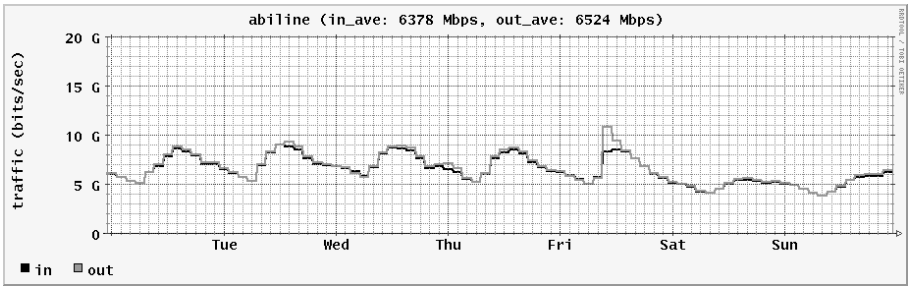

Fig. 5.6. Aggregated total traffic from ABILENE in October 2004. Time is in CDT.

customers). It also includes leased lines used to accommodate residential broadband access within the customer networks (e.g., second or third level ISPs) since ISPs do not distinguish them from other leased lines. As a result, the traffic pattern still appears to be dominated by residential traffic, which is indicated by the peak hours and the differences between weekdays and weekends. However, we also observe office hour traffic (from 8:00 to 18:00) in the daytime on weekdays but traditional office commercial traffic appears to be smaller than residential customer traffic. Note that we cannot directly compare the traffic volume of (A2) with that of (A1) because (A2) was provided by only four of the seven ISPs.

The traffic patterns common to Figure 5.4 and 5.5 are quite different from well-known academic or business usage patterns. For example, Figure 5.6 shows the weekly traffic of ABILENE[2], an Internet2 backbone network for universities and research labs. From Figure 5.6, it is clear that office hour traffic is dominant; traffic peaks occur around noon, and there is less user activity on weekends.

5.3.2 External Traffic

The external traffic groups are used to understand the total traffic volume in Japan. Figure 5.7 shows traffic to and from the six major IXes (B1). It is apparent that the traffic behavior is strongly

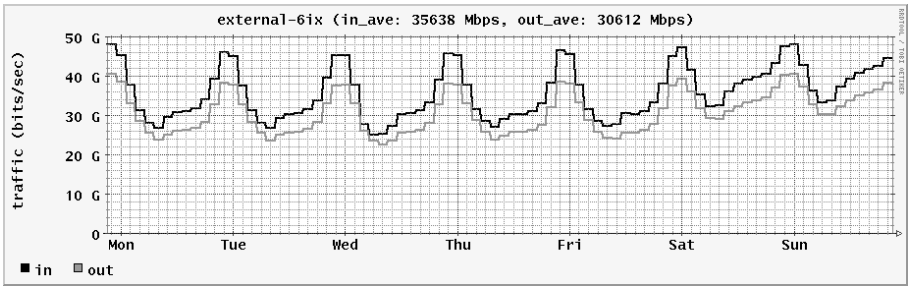

Fig. 5.7. Weekly external traffic to/from the 6 major IXes in September 2004.

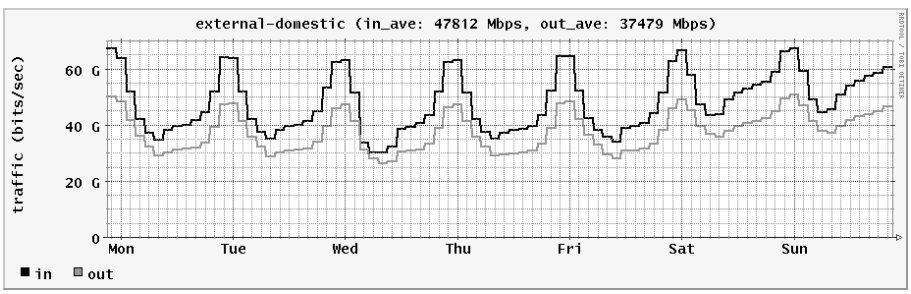

Fig. 5.8. Weekly other domestic external traffic in September 2004.

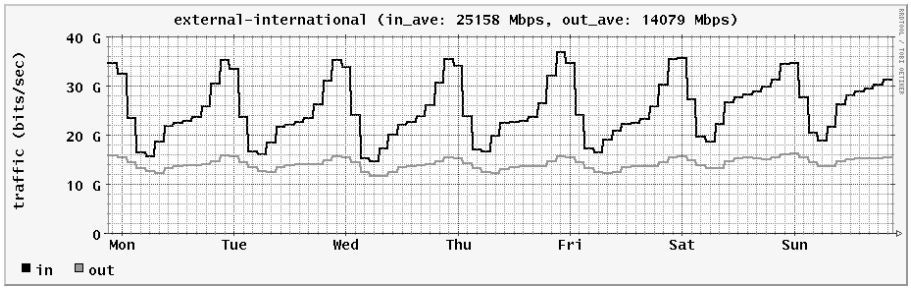

Fig. 5.9. Weekly international external traffic in September 2004.

affected by residential traffic.

Figure 5.8 shows the external domestic traffic (B2) including regional IXes, private peering and transit but not including traffic for the six major IXes. The traffic pattern is very similar to Figure 5.7.

Figure 5.9 shows international traffic (B3). The inbound traffic is much larger than the outbound, and the traffic pattern is clearly different from the domestic traffic. The peak hours are still in the evening, but outbound traffic volume fluctuates less than inbound traffic, suggesting that the traditional behavior of content downloading to Japan still dominates international traffic.

5.3.3 Prefectural Traffic

In order to investigate regional differences (i.e., between metropolitan and rural areas), we collected regional traffic rates of the 47 prefectures. Figure 5.10 illustrates aggregated traffic of one metropolitan prefecture (top graph) and of one rural prefecture (bottom graph). Both graphs exhibit similar temporal patterns such as peak positions and weekday/weekend behavior. In addition, about 70% of the average traffic is constant regardless of the traffic volume. These characteristics are common to other prefectures. One noticeable difference found is that metropolitan prefectures experience larger volumes of office hour traffic, probably due to heavy business usage.

Figure 5.11 is a scatter plot of traffic and population for the 47 prefectures. We found that a prefecture's traffic is roughly proportional to the population of the prefecture. We obtained similar

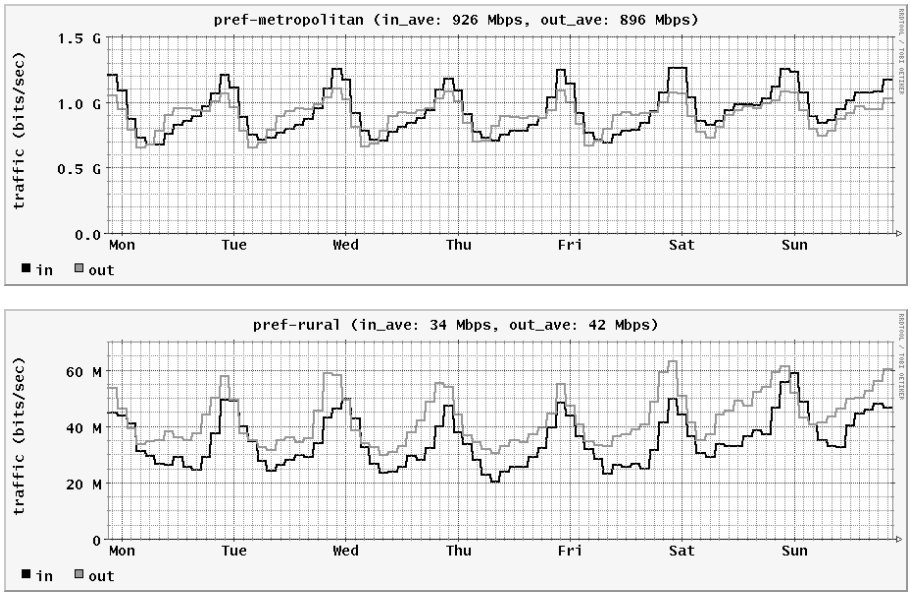

Fig. 5.10. Example prefectural traffic: a metropolitan prefecture (top) and rural prefecture (bottom).

results when the number of Internet users found in [303] is used instead of the population. The result indicates that there is no clear regional concentration of heavy hitters of the Internet. That is, the probability of finding a heavy hitter in a given population is constant.

In order to analyze the scaling property of traffic volume — to find a typical size of prefectural traffic volume, we show the (complementary) cumulative distribution of prefectural traffic on a log-log scale in Figure 5.12. The plot conforms to a power law distribution with a cutoff point at 700 Mbps, meaning that there is no typical size of prefectural traffic volume. In other words, most prefectures generate a small amount of traffic, still prefectures with high traffic volume are observable with a certain probability. It is also observed that the plots for the top 5 largest prefectures deviate from the power law. To investigate this power law decay, we show the cumulative distribution of prefectural populations in the sub-panel. The plots reveal that the power law appearing in traffic volume is derived from the power law decay of prefectural populations, as can be inferred from the linear relationship between traffic and populations in Figure 5.11. Thus, we can conclude that the probability of finding a heavy hitter in a given population is constant and the distribution of aggregated traffic volume directly depends on the population.

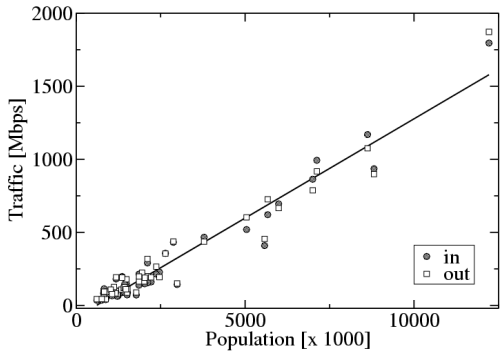

Fig. 5.11. Relationship between population and traffic for prefectures.

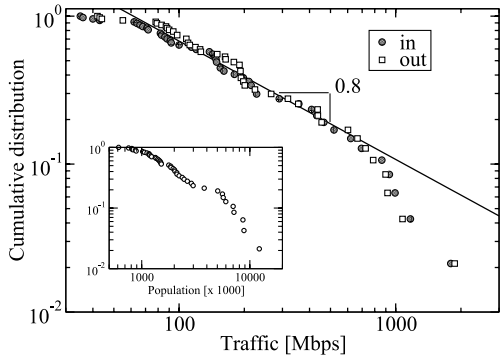

Fig. 5.12. Cumulative distribution of prefectural traffic. Sub-panel indicates the cumulative distribution of populations for comparison.

5.3.4 Summary of Traffic

The monthly average rates in bits/second of the traffic groups are shown in Tables 5.2 through 5.5.

Table 5.2 is the average rates of aggregated customer traffic. As explained before, the non-RBB customer traffic was obtained only from the four ISPs so that it is difficult to directly compare (A1) with (A2). Thus, we estimated the ratio of the RBB customer traffic (A1) to the total customer traffic (A) from only four ISPs' data with both (A1) and (A2). The estimated ratio (A1)/(A1 + A2) is 65% for inbound and 67% for outbound.

Table 5.3 summarizes the average rates of aggregated external traffic. We observe that the total volume of external domestic traffic (B2), mainly private peering, exceeds the volume for the six

Table 5.2. Average rates of aggregated customer traffic.

	(A1) customer-RBB (7 ISPs)		(A2) customer-non-RBB (4 ISPs)	
	inbound	outbound	inbound	outbound
Sep	98.1 G	111.8 G	14.0 G	13.6 G
Oct	108.3 G	124.9 G	15.0 G	14.9 G

Table 5.3. Average rates of aggregated external traffic.

	(B1) ext-6ix (7 ISPs)		(B2) ext-dom (7 ISPs)		(B3) ext-intl (7 ISPs)	
	in	out	in	out	in	out
Sep	35.9 G	30.9 G	48.2 G	37.8 G	25.3 G	14.1 G
Oct	36.3 G	31.8 G	53.1 G	41.6 G	27.7 G	15.4 G

Table 5.4. Average rates of total customer traffic and total external traffic.

	(A) customer (A1 + A2)		(B) external (B1 + B2 + B3)	
	inbound	outbound	inbound	outbound
Sep	112.1 G	125.4 G	109.4 G	82.8 G
Oct	123.3 G	139.8 G	117.1 G	88.8 G

Table 5.5. IX traffic observed from ISPs and from IXes.

	(B1) ext-6ix	traffic observed by IXes
	outbound	inbound
Sep	30.9 G	74.5 G
Oct	31.8 G	77.1 G

major IXes (B1). From this result, it can be concluded that simply relying on data from IXes to estimate and understand nation-wide traffic may be misleading, because a considerable amount of traffic is exchanged by private peering. At the same time, it is possible that the volume of private peering is larger in our measurement than the rest of the Japanese ISPs because private peering is usually exercised only between large ISPs. The ratio of international traffic to the total external traffic is 23% for inbound and 17% for outbound.

There is a relationship between the total customer traffic (A1 + A2) and the total external traffic (B1 + B2 + B3) in Table 5.4. If we assume all inbound traffic from other ISPs is destined to customers, the inbound traffic volume for the total external traffic (B) should be close to the outbound traffic volume for the total customer traffic (A). Similarly, the outbound traffic volume of (B) should be close to the inbound traffic volume of (A). However, the non-RBB customer data is provided by only 4 ISPs. If we interpolate the missing ISPs in the non-RBB customer traffic using the ratio from the four reporting ISPs, the total inbound customer traffic is estimated to be 152.1 Gbps, and that outbound to be 167.8 Gbps. Though these volumes are higher than those for the total external traffic, this is probably because the total customer traffic contains traffic whose source and destination belong to the same ISP.

Lastly, we examined the relationship between our IX traffic data (B1) and the total input rate of the six major IXes, as obtained directly from these IXes[1]. In comparison with the published total incoming traffic of these IXes, our data represent 41% of the total traffic as shown in Table 5.5. If we assume this ratio to be the traffic share of the seven ISPs, the total amount of residential broadband traffic in Japan is roughly estimated to be 250 Gbps.

To check consistency, we collected the September results and the October results separately in October and November respectively. These results are consistent so that we are fairly confident about their accuracy. However, the traffic increase from September to October was higher than our projection; the traffic of the six IXes increased by only about 3% but the other groups increased by about 10%. We suspect that some links missing in the September measurements could have been added later for the October measurements, and we expect further measurement will shed light on this issue.

5.3.5 Distribution of per-customer traffic

In order to verify our assumption that the distribution of heavy hitters is similar across different regions, we obtained per-customer traffic information for October 2004 from one of the participating ISPs. The inbound/outbound traffic volumes of residential broadband customers for each prefecture were collected by means of sampled NetFlow[38] and matching customer IDs with the assigned IP addresses. Although this data set is from only one ISP, the results appear to be consistent with the aggregated results. The results are also consistent with earlier measurements on peer-to-peer traffic by Sen and Wang[250]; peer-to-peer traffic is extremely variable and highly

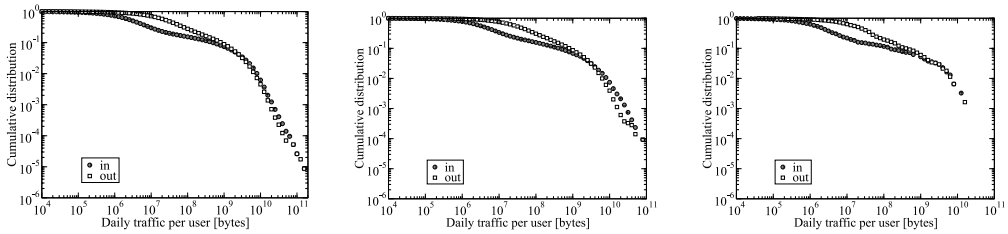

Fig. 5.13. Cumulative distribution of daily traffic per user: all prefectures (left), a metropolitan prefecture (middle) and a rural prefecture (right).

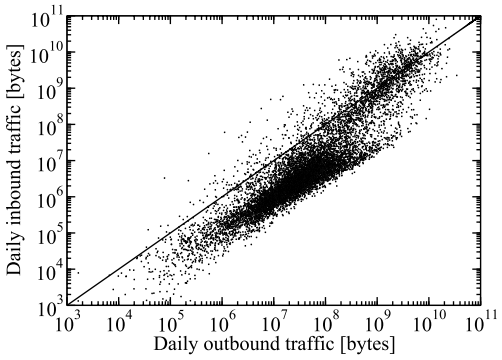

Fig. 5.14. Correlation of inbound and outbound traffic volumes in one metropolitan prefecture.

skewed among participating nodes.

Figure 5.13 shows the (complementary) cumulative distribution of daily traffic per customer on a log-log scale, and compares all the prefectures (left) with one metropolitan prefecture (middle) and one rural prefecture (right). The daily traffic volume is the average of the month, and the distribution is computed independently for inbound and outbound traffic. It is common to the three plots that about 4% of the customers use more than 2.5 GB/day (or 230 kbits/sec) and, beyond this point, the slope of the distribution changes. Thus, heavy hitters can be statistically identified as customers using more than 2.5 GB/day. The distribution also shows that outbound traffic is dominant for most customers but it does not hold for heavy hitters. These trends are consistent across different prefectures, and the differences are only in the tail length affected by the number of customers, which confirms that the distribution of heavy hitters is similar across different regions.

Figure 5.14 is a log-log scatter plot to show the correlation between inbound and outbound traffic volumes for each customer. There is a positive correlation as expected, and the highest density cluster is below and parallel to the unity line where the volume of outbound (down-streaming for customers) is about ten times larger than that of inbound. This is probably due not only to application characteristics but also to the restriction of asymmetric access lines. In a higher volume region, a different cluster appears to exist around the unity line. The slope of the cluster seems to be slightly larger than 1, which explains the inversion of inbound and outbound traffic volumes in Figure 5.13. A plausible interpretation of excess upstream traffic of heavy hitters is that symmetric high bandwidth of FTTH access lines complements the shortage of upstream bandwidth of DSL lines. It can be also observed that, across the entire traffic volume range, the inbound/outbound traffic ratio varies greatly, up to 4 orders of a magnitude. This plot is taken from a metropolitan prefecture but the correlation is common to all prefectures.

Figure 5.15 shows the cumulative distribution of traffic volume of all of the prefectures with heavy hitters in decreasing order of volume. Again, the distribution is computed independently for inbound and outbound traffic. The graph reveals skewed traffic distribution among customers; the top N% of heavy hitters use X% of the total traffic. For example, the top 4% of customers at the knee point of the distribution in Figure 5.13 use 75% of the total inbound traffic, and 60% of the outbound.

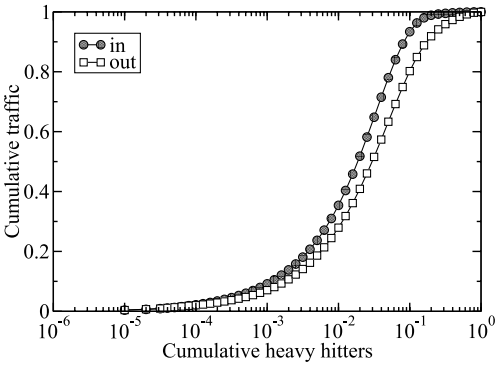

Fig. 5.15. Cumulative distribution of traffic volume with heavy hitters in decreasing order of volume.

5.4 Conclusion

The widespread deployment of residential broadband access has tremendous implications to our lives. Although its effects to the Internet infrastructure are difficult to predict, it is essential for ISPs to prepare for the future to accommodate innovations brought by empowered end-users.

Residential broadband traffic has already contributed to a significant increase in commercial backbone traffic. In our study, residential broadband traffic accounts for two thirds of the ISP backbone traffic, which should have a significant impact on the pricing and cost structures of the ISP business.

The properties of residential broadband traffic differ considerably from those of academic or office traffic often seen in literature. The constant portion of daily traffic fluctuations is about 70%, much larger than ones found in earlier reports[81, 87]. Research results obtained from campus or other academic networks may no longer apply to commercial traffic. More research efforts should be directed to measurement and analysis of residential broadband traffic.

The inbound/outbound rates are roughly equal throughout our data sets. Many access technologies employ asymmetric line speed for inbound and outbound based on the assumption that content-downloading is dominant for normal users. However, this assumption does not hold in our measurements.

Our measurements also suggest that a large amount of traffic is exchanged by private peering so that data from IXes may not be an appropriate index of nation-wide traffic volume.

The prefectural results show that traffic volume is roughly proportional to regional population. It indicates a unique characteristic of the cyberworld in which activities are not bound by time and place. If this is the case, it would affect the design of capacity planning for the future Internet.

For future work, we will continue collecting aggregated traffic logs from ISPs. We are also planning to do more detailed analysis of residential broadband traffic by selecting a few sampling points.

Acknowledgments

We are grateful to the following ISPs for their support and assistance with our data collection; IIJ, Japan Telecom, K-Opticom, KDDI, NTT Communications, POWEREDCOM, and SOFTBANK BB. We would like to thank the Ministry of Internal Affairs and Communications of Japan for their support in coordinating our study, and for providing the statistics of broadband subscribers. We would also like to thank Akira Kato of the University of Tokyo, Atsushi Shionozaki of Sony CSL, Randy Bush of IIJ America, and anonymous reviewers for their valuable input.

第 6 章　第 5 回 CAIDA/WIDE 計測ワークショップ報告

6.1 概要

2005 年 3 月 11 日から 12 日にわたり、WIDE プロジェクトと CAIDA による計測ワークショップが、ロサンゼルス・マリナ・デル・レイにある南カリフォルニア大学の Information Sciences Institute で行われた。以下は、そこでの発表について、簡単にまとめたものである。

本件のサイトは、http://www.caida.org/projects/wide/0408/ となっており、プレゼンテーション資料が公開されているので、詳細については参照されたい。

6.2 プログラムとプレゼンテーションの概要

6.2.1 Activity Review — WIDE and CAIDA

WIDEプロジェクト代表村井から、最近のアクティビティ、トピックについての紹介が行われた。以下のようなキーワードがあげられる。内容は、2004年度あるいは2005年度の報告書に書かれている。

- Unwired Networking（InetnetCAR, 電車でのインターネットアクセス、MANET）
- RFID関連アクティビティ
- IPv6関連アクティビティ（ゲームでの活用、携帯電話とルーティング）
- Realtime HD over IP
  （University of Washington–SFC間）
- 大容量コミュニケーション
  （U-Tokyo–CERN/7.5 Gbps TCP）
- Lambda network（T-LEX/IEEAF Status）
- Networkの遅延について

続いて、Kim Claffyから、CAIDAのアクティビティの紹介があった。スライドに記載の事項以外に、以下のようなコメントがあった。

- status
  - WISP (Workshop on Internet Signal Processing) conferenceをホストした
  - skitterの結果分析——予算の都合で中止になる可能性あり
  - 新しいデータページが出来、データ種類、ACポリシなどで整理されている
    （http://www.caida.org/data/）
  - Internet Measurement Data Catalog（IMDC）を2005年6月にアルファリリース
  - Shark: 1万人が利用しているsshのログをどう監視するか

6.2.2 DNS measurements-I

- DNS anycast stability: some initial results
  IIJのRandy Bushから、DNSのルートサーバのAnycastとルーティングの関係についての初期的な調査結果が報告された。調査は、ボランティアによって、数百のホスト上で、2秒毎にUDP/TCP双方のクエリを発行するかけるスクリプトを数日間走らせることで行われた。
- Case studies of root server abuse
  Duane Wesselsによる、DNSサーバへの不正アクセスについてのケーススタディが紹介された。4件のアタックについての分析と、対策についての説明となっている。

6.2.3 Routing I

- IPv6 AS topology
  Bradley Huffakerによる、IPv6コアネットワークの状況の解析結果についての報告。いままでも解析されているv4の状況との比較が提示された。IPv4のAS数/リンク数に12517/35334に対して、IPv6は333/1304である。
- Active measurements of IPv6 topology: update on scamper project
  Matthew Luckieによる、scamperプロジェクトの進捗についての報告があった。
- Degree correlations and topology generators
  Dima Krioukovによる、確度の高いトポロジグラフ生成アルゴリズムについての考察。多くのトポロジー特性を反映できるように、物理学の最大エントロピー原理に倣った手法を用いており、低次（2K）までの定義の提案と実際のネットワークとの比較で一致していることを提示している。

6.2.4 DNS Measurements-II

- Passive and active DNS measurement: update
  関谷勇司による、dnsprobeによるDNS Measurementについての続報。
- RFC1918 updates on servers co-located with M and F roots
  Andre BroidoによるM及びFルートサーバにおける、RFC1918（Address Allocation for Private Internets, 1996）アドレスに対する、アップデートリクエストに関連した調査についての報告。

6.2.5 Security

- Current events in the Network Telescope project
  Colleen Shannonによる、Network Telescope

Project についての報告。

- Internet Threat Monitors: Are they safe?
  篠田陽一による、Internet Threat Monitor の安全性についての議論。

6.2.6 Routing and simlulations-II

- Analysis of route reflector performance in I-BGP
  長橋賢吾による I-BGP のルートリフレクタにおける可用性向上の提案と、そのスケーラビリティについての考察
- Advancements in the inference of AS relationships
  Xenofontas Dimitropoulos による、AS 間の関係推測アルゴリズムについての考察
- Comparison of server selection algorithms
  宮地利幸による、シミュレーションによる 3 種類のサーバ選択アルゴリズム（Best、Uniform（Random）、Reciprocal）の比較。

6.2.7 トラフィックモニタリング

- The Impact of Residential Broadband Traffic on Japanese ISP Backbones
  福田健介と長健二朗による、日本国内の一般家庭向のブロードバンドトラフィックの国内バックボーンへの影響の分析。IIJ、Japan Telecom、KDDI、K-Opticom、NTT-Communications、PoweredCom、Yahoo BB の 1 ヶ月分のデータを利用。
- Capturing dragonflies — how to measure very short-lived streams
  Nevil Brownlee による、インターネット上のストリームの持続時間についての観測についての報告。
- High-speed network measurement
  中村修と江崎浩による、10 Gbps のネットワークの計測についての簡単な報告

6.2.8 その他

- Measuring IPv6 Network Quality
  長健二朗による、IPv6 のネットワーク品質測定についての提案と進捗報告

6.3 今後の予定

次回のワークショップは、2006 年 3 月 17 日–18 日に、USC/ISI で行われる予定である。

第 7 章　まとめ

インターネットの研究において、計測はますます重要視されてきている。そのような状況のなかで、WIDE の計測活動は、グローバルな視点を持った継続的な計測活動として国際的にも認知されてきている。2006 年度には、従来の計測活動を継続しながら、さらに計測分野での国際協調が広がる予定もあり、新しい視点からの研究にも取り組んでいくつもりである。