第 IV 部

ネットワークトラフィック
統計情報の収集と解析

第4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　はじめに

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、また、データの保存と利用に関する活動を行っている。

MAWI WG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。

- 広域で行う

  インターネットの最大の特徴は、大規模な広域ネットワークにある。しかし、トラフィック情報には組織の機密保持やプライバシー保護の問題が伴うため、特に不特定多数のトラフィックを含む広域データを第三者が入手するのは困難である。

  ひとつの企業や組織内といった狭い範囲でデータを取ることは各組織で可能だが、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だからこそ可能である。

- 多地点で見る

  インターネットのもう1つの大きな特徴は、自律したネットワークが相互接続して経路制御を行い、エンド・エンドで通信制御を行う分散システムにある。したがって、ある地点でトラフィックを観測しても、ネットワーク全体の状態を捉えることは不可能である。観測者はあくまでその観測点から見たインターネット像が得られるだけで、別の観測点からはまったく別の世界が見えているかも知れない。

  MAWI WGでは、多地点で観測したデータを照らし合わせることによって、より広い範囲のネットワークの状態を把握する手法や、それを俯瞰で可視化することによって直観的に分かりやすく観測する手法について研究を行っている。

- 長期間行う

  ネットワークのトラフィックの挙動は、TCPの特定のアルゴリズムが関係するようなミクロなものから、過去10年間のトラフィック量の推移のようなマクロなものまで、幅広い時間スケールに渡っている。また、ミクロな挙動についても、インターネットのマクロなレベルの発展に伴って、次第に変わっていく。

  したがって、1日や1週間といった短期間の計測も重要だが、何年間という長いスパンでデータを取り続けることが非常に重要になる。しかし、長期的にデータを収集し、その蓄積を持つことは、ある日誰かが思いついてできるものではない。そこで、ワーキンググループとしてメンバーが協力して継続的なデータ収集を行っていくことが必要である。

今年度の報告書では、第2章において、CAIDA（the Cooperative Association for Internet Data Analysis）との共同研究について報告する。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。ここでは、2003年3月と11月に行った2度のワークショップの内容と、Huffaker氏を日本に招いて行ったIPv6の経路とトポロジに関する研究について報告する。また、2004年度もCAIDAとの共同研究を継続し、DNS計測、IPv6トポロジ計測、BGPの解析などを行っていく予定である。

第3章では、netmanチームが行っているBGPの経路情報収集について報告する。ここ数年、BGP経路情報を利用した研究が増えてきているが、WIDEでもそのノウハウを蓄積するために経路情報収集を始めた。今後はこのグローバルな経路情報を使ったWIDEらしい広域ネットワークの挙動解析に繋げていきたい。

第4章では、2003年11月に新しく設立されたnetviz WGについて報告する。netviz WGの目的は、ネットワークの視覚化およびその他の表現、そのためのツールやノウハウに関する議論と情報の共有を行う事である。従来は、MAWI WGなどで計測データの視覚化等の議論をしてきたが、研究目的を

持ったワーキンググループの中では、どうしても研究内容に話が行くため、参加者の間口が狭くなる。そこで、独立したワーキンググループとして、データ表現にフォーカスした netviz を設立することにした。

2004 年度の予定としては、

- DNS 計測、IPv6 トポロジ計測、BGP の解析などを中心とした CAIDA との共同研究
- 10 G バックボーンの本格運用に伴う広帯域トラフィックデータの取得に関する研究

などを計画している。

第 2 章　2003 年度 CAIDA との共同研究報告

2.1 はじめに

本章では、2003 年度の CAIDA との共同研究について報告する。CAIDA と WIDE は従来から関係をもっていたが、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。

その一貫として、2003 年 3 月と 11 月に 2 度のワークショップを行い、また、CAIDA の Huffaker 氏を 2 カ月間日本に招いた。

2.2 第 1 回 CAIDA/WIDE Measurement Workshop

第 1 回ワークショップは 2003 年 3 月 21 日に CAIDA の本拠地であるカリフォルニア州立大学サンディエゴ校で行った。参加者は、WIDE から 9 名、CAIDA から 9 名、ISC から 1 名の計 19 名であった。直前にイラク戦争が勃発したため、参加できなかった WIDE メンバーが 6 名いた。

前半は自己紹介を兼ねて、各参加者が自分の研究を紹介した。

- Brad Huffaker (CAIDA): skitter プロジェクトと、そのディスタンスメトリックス、PAM で発表予定の DNS ルートサーバーの配置解析についての報告。
- Nevil Brownlee (CAIDA): NeTraMet による DNS ルートと gTLD サーバの RTT のパッシブ計測と、web で公開している結果のグラフについての説明。
- Andre Broido (CAIDA): RFC1918 アドレスの漏洩に関して、AS112 に来る dynamic update を解析した結果の報告。
- Duane Wessels (CAIDA/Measurement Factory): F-root のパケットトレース解析を行い、ルートに来るクエリのうち正規のものは 2%に過ぎないという報告。また、実験室環境での BIND の挙動解析についての説明。
- Jun Murai (WIDE): WIDE プロジェクトの活動概要の説明。
- Yoichi Shinoda (WIDE): StarBED プロジェクトの説明。
- Suguru Yamaguchi (WIDE): AI3 プロジェクトと関連する活動についての説明。
- Hiroshi Esaki (WIDE): WIDE に関連するテストベッド・ネットワークについての説明。
- Hideki Sunahara (WIDE): Internet CAR に関連する研究プロジェクトの紹介。
- Osamu Nakamura (WIDE): NSPIXP の紹介。
- Akira Kato (WIDE): M ルートサーバの運用と、NSPIXP6 についての説明。
- Kenjiro Cho (WIDE): WIDE トラフィックリポジトリ、aguri (aggregagtoin based traffic profiler)、DNS のサーバ選択アルゴリズム解析に関する報告。
- Yuji Sekiya (WIDE): rootprobe というアクティブ計測ツールを用いたルートおよび gTLD サーバーの応答時間解析についての報告。
- Joao Damas (Internet Software Consortium): ISC から参加。F ルートの anycast 展開についての報告。
- kc claffy (CAIDA): CAIDA における計測活動の概要についての説明。

午後のセッションでは、今後の共同研究に関して議論を行った。共同研究の候補として、DNS 応答時間計測に関するアクティブとパッシブな手法の比較、地理的分布を表すサンプルとして特徴的な ccTLD サーバの抽出、AS112 のログ解析、BGP anycast の効果の評価方法などが議論された。

2.3 第 2 回 CAIDA/WIDE Measurement Workshop

第 2 回ワークショップは 2003 年 11 月 7 日に南カリフォルニア大学 ISI で行った。参加者は、WIDE から 12 名、CAIDA から 8 名、ワイカト大学から

2 名、RIPE、NLNET、ISI から各 1 名の計 25 名であった。

前回は自己紹介的な発表が中心であったが、今回は具体的な共同研究項目を中心に、3 つのセッションを行った。

1. DNS の計測とモデリング
   - Yuji Sekiya (WIDE): DNS 応答時間計測に関するアクティブとパッシブな手法の比較のため、WIDE 内に NeTraMet を設置したという報告。
   - Nevil Brownlee (CAIDA): NeTraMet データの自動収集と web を使ったデータ表示についての説明。
   - Henk Uijterwaal (RIPE): RIPE NCC で行っている DNS のモニタリング手法についての説明。
   - Kazunori Fujiwara (WIDE): ICANN CNNP test に基づいた DNS サーバのモニタリングに関する提案と問題点の提示。
   - Duane Wessels (CAIDA): 実験室環境でネームサーバの挙動解析を行う手法の説明。
   - Yasuhiro Morishita (WIDE): OpenBlocks を使って複数のバージョンのネームサーバを動作させて、評価を行う実験の紹介。
   - Marina Fomenkov (CAIDA): ローカル DNS レゾルバの挙動解析をパケットトレースから行う手法についての報告。

   応答時間計測の手法の比較や、実験室環境でのネームサーバの挙動解析に関して、いろいろ互いに協力できる部分があることが確認され、共同して進める事になった。
2. IPv6 計測
   - Matthew Luckie (WAND): 彼の書いた scamper というトポロジ取得のためのツールと、それを利用した IPv6 トポロジの解析についての報告。
   - Kenjiro Cho (WIDE): デュアルスタックホストに対して、IPv6 と IPv4 のパスの比較を行う手法と試験データの紹介。
   - Bradley Huffaker (CAIDA): IPv4 と IPv6 のトポロジ計測に関して、現状の技術的問題点の報告。

   今後とも、CAIDA、WIDE、ワイカト大学が協力して、IPv6 のトポロジに関する計測を行っていく事が合意された。
3. BGP 計測
   - Kengo Nagahashi (WIDE): StarBED を使った BGP シミュレーションの計画と、簡単なトポロジによる評価実験の報告。
   - Patrick Verkaik (CAIDA): Atoms による BGP 情報の集約に関する報告。

今後、Atoms の評価を StarBED 上で行う可能性について議論された。

2.4 夏期共同研究

CAIDA の Bradley Huffaker 氏を 7、8 月の 2 カ月間 WIDE に招いて共同研究を行った。ここでは、その間に行った研究について報告する。

2.4.1 IP landscape of the digital divide

図 2.1 は、国、大陸別の IPv4 と IPv6 の AS 数、プレフィックス数、アドレス数比較をグラフにしたものである。これらのデータは、オレゴン大学の Route-Views プロジェクトが提供する BGP データを解析したものである。アドレスブロックがどの国に属するかは、CAIDA の NetGeo ツールを用いて調べた。こ

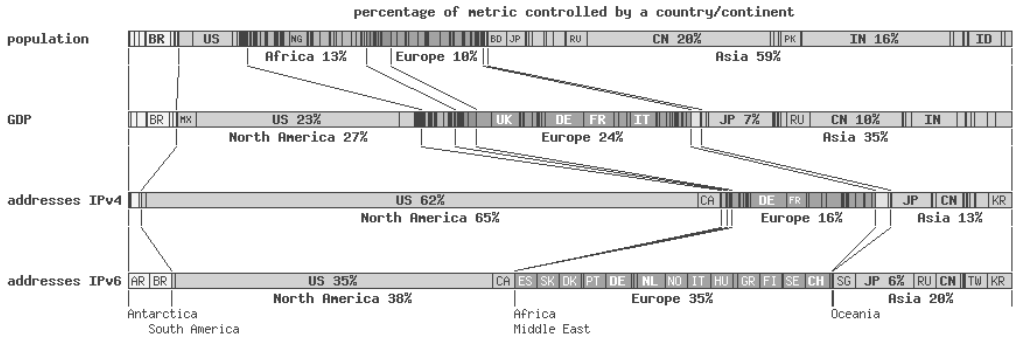

図 2.1. BGP データからみた IP アドレスの分配状況

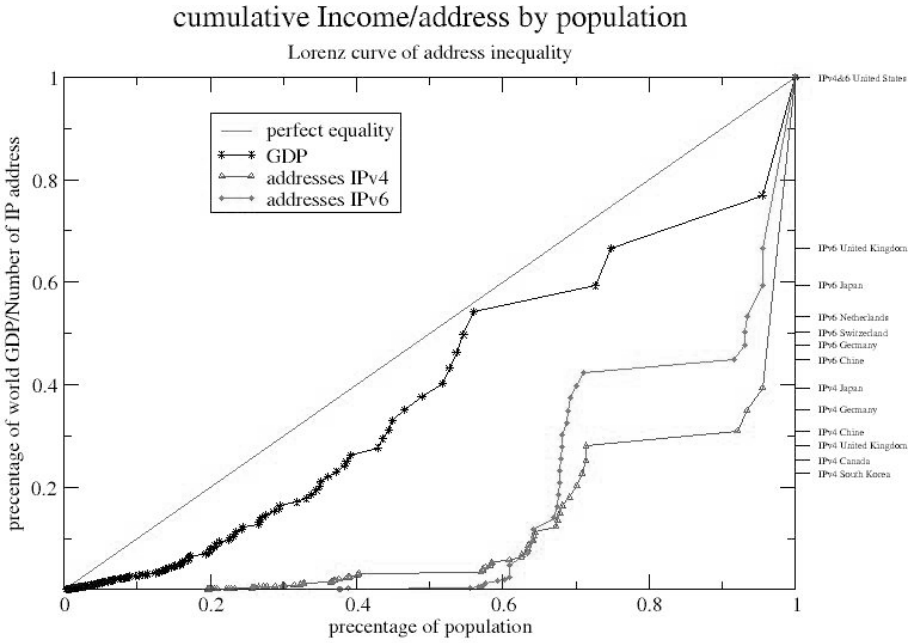

図 2.2.　ロレンツ曲線による IP アドレス分配

こでは、BGP でアナウンスされた同一プレフィックスないのアドレスは全て同一国内に存在するものとして扱っている。また、CIA ファクトブックから得られた人口、GDP とも対比することによって、IP アドレス分配に関するデジタルデバイドを示している。

人口と GDP の配分を見ると、北米とヨーロッパを中心とした先進国が人口に対して大きな GDP を持ち、良く知られるように富の配分に格差があることを示している。

インターネットのアドレス空間の配分は、実は GDP よりはるかに大きな格差がある。ネットワークのインフラへ投資するためには経済力が欠かせないのは事実だが、US が圧倒的に大きな IPv4 アドレス空間を持つのは、それよりも歴史的な理由による。つまり、ARPAnet および初期のインターネットの開発は米国政府による財源で行われたため、初期のインターネットのインフラやそのためのアドレス割り当ては、米国の関連組織間で排他的に行われていた影響が残っている。

IPv6 に関しては、先進国間では IPv4 より公平な配分となっていることが分かる。しかしながら、IPv6 の採用はまだ発展途上国に十分浸透していないため、現状では、先進国と発展途上国間の格差は IPv6 の方が大きい結果となっている。ただし、IPv4 はアドレス空間の枯渇から事態の改善が望めないのに対し、IPv6 のアドレス空間は広大なので、普及に出遅れた国でも将来アドレスを取得するのに不利はなく、IPv6 の浸透に伴いより公平な分配へと移行すると予想できる。

図 2.2 は IPv4 と IPv6 の国別アドレス分配を、経済学で富の分配の公平性を示すために使われるロレンツ曲線にしたものを示す。ここでは、人口に対する資源の累積配分を資源のサイズ順に国をソートしてプロットしている。もし、資源が公平に分配されていれば、$y = x$ の直線になり、この直線から離れるにしたがって不公平が強まる。

この図からも、IPv6 アドレスの分配が、全体として IPv4 より公平性に欠ける結果となっている。我々は、この図を使って今後の変化を観測することで、インターネット上のデジタルデバイドを示すひとつの指標を与えることができると考えている。また、ここでは国ごとの比較を行ったが、同様の手法で、国内の地域間のデジタルデバイドを示すことや、さらに職業別グループによる比較などにも応用が考えられる。

2.4.2 A view into the IPv6 AS-level topology

図 2.3 は、WIDE からみた IPv6 の AS レベルのトポロジを示している。測定は、2003 年 6 月 6 日に行った。

図 2.3. WIDE から見た IPv6 の AS レベルのトポロジ

まず、オレゴン大学の RouteViews プロジェクトが公開している BGP テーブルから広告されているプレフィックス毎にひとつの IPv6 アドレスを選び、計 4,173 の IPv6 アドレスをターゲットとして抽出した。次に、多数の IPv4/IPv6 ターゲットに対する traceroute を効率よく実行する scamper ツールで全てのターゲットへのフォワーディングパスを測定した。最後にパス上のアドレスを BGP テーブルのオリジン AS と最長マッチングする AS 番号に変換して、さらに集約しグラフ化を行っている。AS は地域ごとに分類しているが、Verio、IIJ、GBLX など多国間にまたがる AS に関しては、本社の所在地をもとに分類を行っている。フォワーディングパスのトポロジは、たとえ同一 AS 内であっても、異なる地点から測定すれば大きく異なる可能性がある。したがって、図 2.3 は、正確には WIDE（AS2500）内のある地点から観測されたものに過ぎず、WIDE の持つ全てのピアリングが表されているわけではない。ピアリングが存在しても、観測地点からターゲットへのパスに含まれていないために、グラフに現れてこない場合もある。したがって、図 2.3 は、IPv6 全体の AS 間トポロジではなく、あくまでも測定地点から観測可能な AS 間のピアリングを示しているに過ぎない。

2.5 今後の予定

2004 年度も CAIDA との共同研究を継続する予定である。共同研究の主なテーマとして、DNS 計測、IPv6 トポロジ計測、BGP の解析などが挙げられる。昨年同様、ワークショップの開催や、人材交換を行っていく。また、IPv6 トポロジ計測に絡んで、ワイカト大学も共同研究に加わる事になっている。

第 3 章　BGP ルーティング情報の収集

本章は、仙台 NOC の wide-netman チームが行っているネットワーク情報収集活動において、ネットワーク情報の分析をより高度に行うために BGP プロトコルのルーティング情報を活用するため、仙台 NOC で 2003/8/27 より開始した BGP 情報の収集について報告するものである。

3.1 活動の背景

wide-netman チーム（仙台 NOC）では、パッシブ型モニタを用いたトラフィック情報の収集を継続して行っている。パッシブモニタリングでは収集されるデータ量は膨大になるため、適切なサマライゼーションや可視化を行う必要がある。我々はこれまで、ネットワーク地図を用いたネットワークの可視化の研究を行ってきた。ネットワーク地図とトラフィック情報を組み合わせることで、より直観的にネットワークを把握することができ、ネットワークの障害検出なども容易になると考えられる。ネットワークマップのような情報はネットワーク情報の「コンテキスト情報」と考えられるが、このような情報源として BGP プロトコルのもつ情報が有用であると考え、データ収集を行うこととした。

3.2 BGP 情報の有用性

BGP（Border Gateway Protocol）は、AS（Autonomous System）の概念に基づいたルーティングプロトコルである。各 BGP ルータは自身が広告するネットワークプレフィックスを、自身の AS 番号とともに BGP peer に対して広告する。BGP は AS パスの長さに基づいて、ネットワークプレフィックスの次ホップを決定する。BGP ルータは基本的には隣接する BGP ルータと peer を確立するため、ルーティング情報伝播の過程で生成される AS パス情報はネットワークマップとしての性格を持つ。また同時に、ルーティングのためのネットワークプレフィックスは現実に存在するネットワークアドレスである場合が多く、その情報はネットワーク構成情報としての性格も持つ。すなわち BGP の内部状態情報は、ネットワーク情報のコンテキスト情報や情報集約のためのネットワーク構成情報として有用である。

しかし、通常は BGP ルータのみがこれらの情報を持つことができるため、特にリアルタイムなネットワークモニタリングへの活用には問題もある。BGP 情報の収集を行っているプロジェクトとしては Routeviews Project（http://www.routeviews.org/）があるが、モニタリングポイントのネットワークに存在する BGP ルータから情報を取得しなければ正しいルーティング情報は得られない。

我々が開発を行っているパッシブモニタリング手法はネットワークに負荷を与えないため、詳細なト

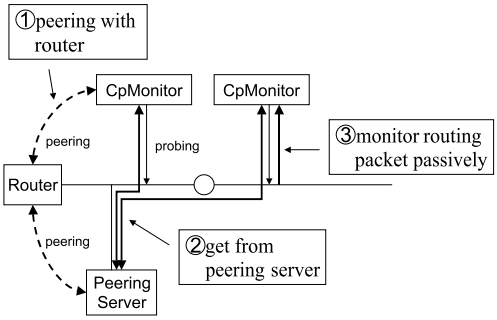

図 3.1. Injecting context information

ラフィックの分析が可能である。パッシブモニタが BGP の情報を取得する形態として以下の 3 つが考えられる（図 3.1）。

1. BGP ルータと private peering を確立し、BGP ルータから情報を転送する。
2. BGP ルータと private peering する peering server を設置し、peering server より SNMP や BGP-MIB 等を用いて情報を取得する。
3. BGP ルータが peering しているリンクをパッシブモニタリングし、BGP パケットをリアセンブルして情報を取得する。

1. の手法はもっとも理想的であるが、パッシブモニタの数が多数になる場合は有効ではない。また現実には BGP ルータとの peering はルータのポリシーの問題から難しい。

2. はパッシブモニタの数が多い場合に有効であるが、peering server へのポーリングのタイミングによって情報の更新にディレイが生じる。また、1. と同様のポリシーの問題がある。

3. はルータのポリシーに関係なく情報を取得でき、仕組みももっとも簡単である。しかし BGP ルータは情報を更新分だけ流すため、情報全体を取得するまでに時間がかかることと、情報が流れるリンクが限定されることが欠点である。

3.3 BGP 情報の収集

実験を行っている仙台 NOC には BGP ルータが存在しないため、1. および 3. のアプローチをとることはそのままでは出来ない。そのため、2. の方式を取る事とし、WIDE の海外線に存在する CISCO ルータと ebgp multihop によって private peering を確立することとした。

これによって同時に 3. のアプローチを実験する環境も構築する事ができ、また同時に BGP の情報を収集、蓄積することで、オフラインでの解析や BGP データを用いたその他の解析にも役立てる事が出来る。

仙台 NOC において peering を行う PC の諸元は以下の通りである。

- Name/IP: pc10.sendai.wide.ad.jp（203.178.138.26）
- OS: FreeBSD 4.9-STABLE（2003/12/20 現在）
- Disk Space: 20 GB + 120 GB（RAID）+ 700 GB（RAID）
- BGP daemon: zebra-0.94（2003/12/20 現在）（zebra + bgpd）

peering 先は、

- cisco1.LosAngeles.wide.ad.jp（203.178.136.20）

である。Full Route 情報を Mirroring する設定としている。

データの蓄積は、Routeviews Project が蓄積している、BGP パケットの Full dump、UPDATE パケットの dump、および RIB データの 3 種類のデータを 2 時間毎に蓄積しているのに加えて、管理コンソールから（>show ip bgp）を実行して得られる output も 1 時間毎に収集、蓄積している。

データ蓄積は、2003 年 8 月 27 日から現在まで継続して行っている（PC のメンテナンス等のため、データの存在しない期間も若干ある）。現在、総データ量は 36.8 GB 程度となっている。

3.4 今後の予定

現状では BGP ルータとの peering および情報の蓄積を始めた段階であり、peering server からパッシブモニタへの BGP 情報の注入機能の具体的な実装がこれからの課題である。

第 4 章 netviz WG の設立について

netviz ワーキンググループが 2003 年 11 月に設立された。本ワーキンググループの目的は、ネットワークの視覚化およびその他の表現、そのためのツールやノウハウに関する議論と情報の共有を行う事である。

現在 WIDE には、ビジュアライゼーション自体を研究テーマにしている人はいないが、目に見えない

インターネット技術を説明するため、また、説得力のあるプレゼンテーションのために、道具としてビジュアライゼーションを使いこなす事はすべての研究者に必要になっている。

従来は、MAWI WG などで計測データの視覚化等の議論をしてきたが、研究目的を持ったワーキンググループの中では、どうしても研究内容に話が行くため、参加者の間口が狭くなる。そこで、独立したワーキンググループとして、データ表現にフォーカスした netviz を設立することにした。

netviz では、ネットワークと表現に関わるさまざまな話題を取り扱う。現在のところ、参加者の関心は以下の 3 つの領域に分けられる。

1. データ表現、データ抽出
   ネットワークトポロジやシステムログの視覚化、音声や光などアンビエントメディアを用いたネットワーク状態の表現など、データ表現の手法に関する領域。
2. ツール学
   グラフィックツール、文章作成やプレゼンテーション作成ツール、AV 機器の使い方のノウハウに関する領域。ここには、色や音声の有効な使い方、アクセシビリティやユーザビリティも含まれる。
3. 社会学、文化学
   ネットワークとデータ表現を広くとらえ、社会学的あるいは文化学的な議論も、データ表現に対する意識を高めるのに役立つと考えている。

前述したように、現状はこれらを主研究テーマとする者はいないため、当面は情報交換とノウハウの蓄積を行っていく予定である。将来的には、WIDE 内にもこのような分野の研究者が出てくる事を期待したい。