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　MAWI ワーキンググループ

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・解析を行ってる。

1. 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出来るが、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 だからできる。
2. 多地点で見る
   インターネットのもう一つの大きな特徴は、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。
3. 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、一日や一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、ある日誰かが思い付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。

今回の報告書では、第 2 章において、昨年度に引続き集約型トラフィックプロファイラを使った国際線の計測について報告する。このツールは、WIDE バックボーンのトラフィックをリアルタイムかつ長期的にモニタリングする目的で作られ、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。今回は、WIDE 報告書の発行時期の変更のため 2002 年 4 月から 12 月の国際線トラフィックの傾向を報告する。

第 3 章では、GPS を用いた高精度の時刻情報を記録できるパケットキャプチャ装置の開発について報告する。本装置は 1Gbps までのイーサネットに対応し、専用 ASIC によるパケットフィルタリング機構を持つ。 キャプチャされたパケットは、GPS による高精度時刻情報を付加して別のイーサネットインターフェイスから記録用のマシンに転送され、そこでディスクに保存される。7 月の横浜 IETF では、この装置を用いて対外線のモニターを行ったので、そのトラフィック解析結果も含めて報告する。

第 4 章、第 5 章では、横浜 IETF での無線 LAN 利用について報告する。無線 LAN は急速に普及しているが、現在市場に出回ってる製品は小規模 LAN をターゲットにしているため、大規模な無線 LAN の

運用にはさまざまな問題がある。

ハイエンドとされる無線アクセスポイントでも 200 ユーザ程度の同時使用が限界であり、負荷が限度を越えるとリブートしてしまうのが現状である。そこで、アクセスポイントの配置や、使用無線チャネル、電波強度などを工夫して、負荷分散をはかる必要がある。

特に IETF では、1000 人以上の無線ユーザがいて、休憩時間に一斉に会場移動することや、会議室の扉の開閉やひとの移動によって室内外の電波状態が変動してユーザが接続先のアクセスポイントを自動的に切替えるなど、無線 LAN 運用が極めて難しい。 NOC では、ユーザの利用状況をリアルタイムにモニターして適切な対応を取れるようにした。

さらに、これだけの規模になると、ユーザの中には自分で気付かずに無線 LAN を ad-hoc モードにしたり、DHCP サーバーや経路広告をするなどの問題を発生する者がでてくるため、問題の検出と問題源となっているユーザの特定を速やかに行う必要がある。しかし、無線 LAN 環境ではユーザの特定は容易ではない。横浜 IETF では、各アクセスポイントを定期的にポーリングして、接続しているクライアントの MAC アドレスを取得してデータベースに登録した。これにより、MAC アドレスから接続しているアクセスポイントが特定できるため、ユーザの物理的な位置がある程度わかる。 また、問題が深刻な場合は該当 MAC アドレスの接続を強制的に切断できるような設計と運用を行った。

第 6 章では、ルートネームサーバならびに ccTLD サーバの応答時間計測について報告する。この研究は、2002 年度に ICANN RSSAC(DNS Root Server System Advisory Committee) に報告したものである。DNS はインターネットの基幹をなすインフラのひとつであり、とくにルートあるいはトップレベル DNS サーバの果たす役割は非常に重要である。しかしながら、これまでは経験則に頼って運用してきた部分が多く、十分な測定データがなかった。現在は、WIDE と CAIDA(the Cooperative Association for Internet Data Analysis) が中心となり、世界規模で DNS の性能計測に取り組んでいる。

2003 年度の予定としては、

- DNS 計測を中心とした CAIDA との共同研究
- 10G バックボーンの試験運用に伴う、広帯域トラフィックデータの取得に関する研究
- 集約型トラフィックプロファイラの機能拡張

などを計画している。

第 2 章　AGURI を用いた WIDE 国際線のトラフィック傾向

2.1 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/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[37] 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。

2.2 収集データ

WIDE プロジェクトでは以下に示す 3 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(18Mbps CAR on 100Mbps

表 2.3. 識別された IP アドレス

graph	IP アドレス	hostname
図 2.7, 2.17, 2.18, 2.19	133.27.0.0/16	keio.ac.jp
図 2.4, 2.7	150.65.0.0/16	jaist.ac.jp
図 2.17	203.178.136.0/21	wide.ad.jp
図 2.17,2.19	202.249.24.0/22	ai3.net
図 2.17	163.215.3.16/28	tec.ac.jp
図 2.17,2.18	192.244.23.1	IRC サーバ A
図 2.18	192.244.23.1	IRC サーバ B

link)

2. samplepoint2
   US-Japan line(US side 60Mbps POS)
3. samplepoint3
   US-Japan line(Japan side)

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の US 側 (samplepoint2)、日本側 (samplepoint3) でそれぞれデータ収集を行っている。samplepoint2 と samplepoint3 ではほぼ同一のデータを収集することができる。

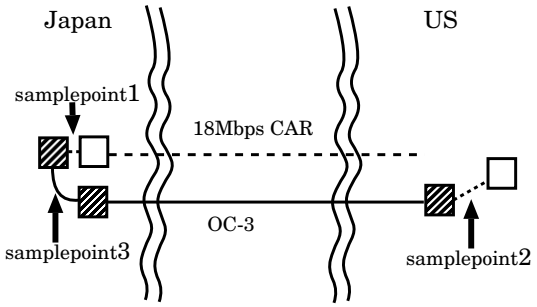


図 2.1. データ収集地点

2002 年度の WIDE 報告書では、samplepoint1 と samplepoint3 で収集した WIDE 国際線の年間トラフィック傾向を図 2.2 から図 2.25 に示す。図の出力は時期を四半期ごとに、対象を 1) 送信元 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 2.1、表 2.2)

表 2.1. トラフィック傾向一覧表 (samplepoint1)

	4月–6月	7月–9月	10月–12月
送信元IPアドレス	図2.2	図2.3	図2.4
宛先IPアドレ図2.5	図2.6	図2.7	
送信元ポート番号	図2.8	図2.9	図2.10
宛先ポート番図2.11	図2.12	図2.13	

表 2.2. トラフィック傾向一覧表 (samplepoint3)

	4月–6月	7月–9月	10月–12月
送信元IPアドレス	図2.14	図2.15	図2.16
宛先IPアドレ図2.17	図2.18	図2.19	
送信元ポート番号	図2.20	図2.21	図2.22
宛先ポート番図2.23	図2.24	図2.25	

図 2.2 から図 2.25 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 2.3, 表 2.4 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、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のため、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 IP アドレス空間を検出できた。
  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。

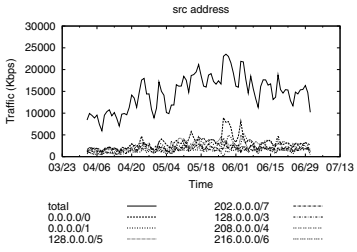


図 2.2.　送信元 IP アドレス (4 月–6 月)

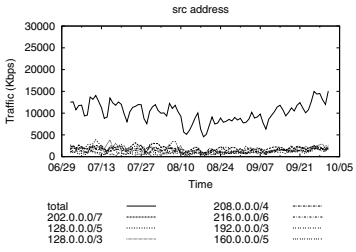


図 2.3.　送信元 IP アドレス (7 月–9 月)

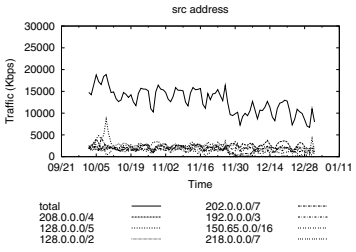


図 2.4.　送信元 IP アドレス (10 月–12 月)

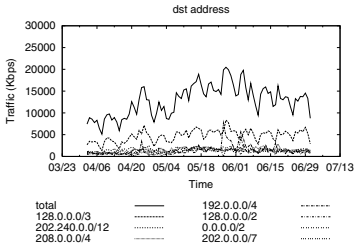


図 2.5.　宛先 IP アドレス (4 月–6 月)

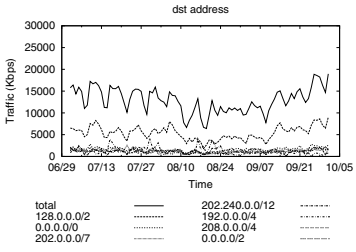


図 2.6.　宛先 IP アドレス (7 月–9 月)

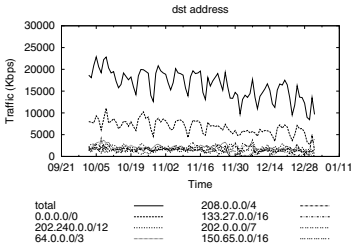


図 2.7.　宛先 IP アドレス (10 月–12 月)

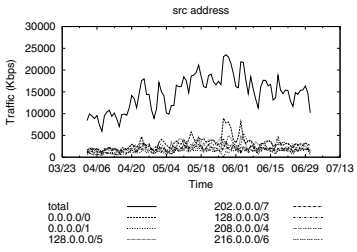


図 2.8.　送信元ポート番号 (4 月–6 月)

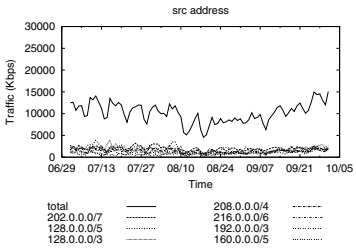


図 2.9.　送信元ポート番号 (7 月–9 月)

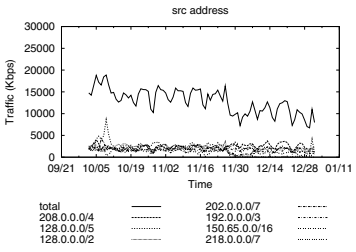


図 2.10.　送信元ポート番号 (10 月–12 月)

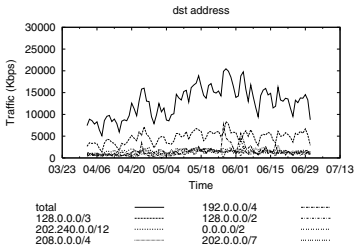


図 2.11.　宛先ポート番号 (4 月–6 月)

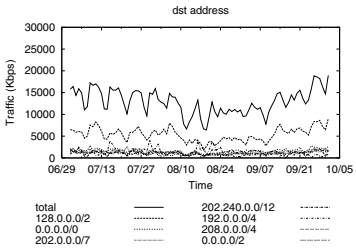


図 2.12.　宛先ポート番号 (7 月–9 月)

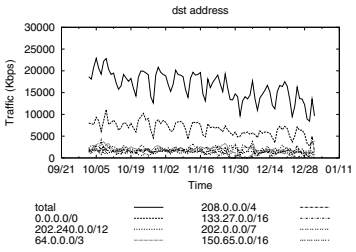


図 2.13.　宛先ポート番号 (10 月–12 月)

特に 2001 年度の WIDE 報告書と比較した場合、2001 年度に観測されなかった ai3.net を宛先としたトラフィック、tec.ac.jp を宛先としたトラフィック、IRC サーバを宛先としたトラフィックを抽出することができた。

また、今年度は WinMX や Napster といった P2P 型ファイル交換アプリケーションのリスト交換データである ‘4:6:6699’（IPv4:tcp:port6699 番）を samplepoint1 において抽出することができ、観測地点によってアプリケーションの流行が異なる結果となった。

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。

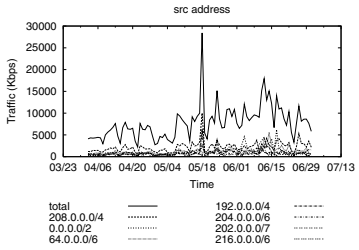


図 2.14. 送信元 IP アドレス (4 月–6 月)

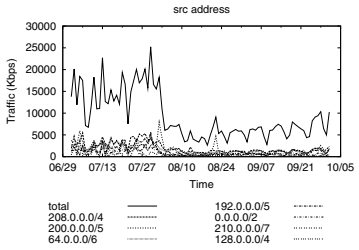


図 2.15. 送信元 IP アドレス (7 月–9 月)

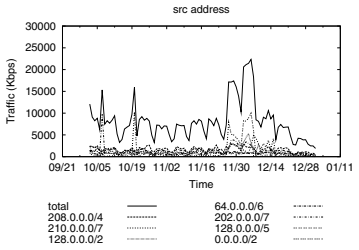


図 2.16. 送信元 IP アドレス (10 月–12 月)

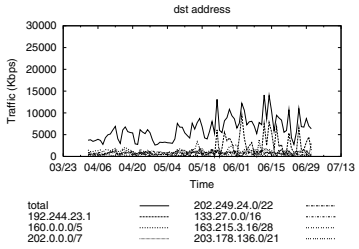


図 2.17. 宛先 IP アドレス (4 月–6 月)

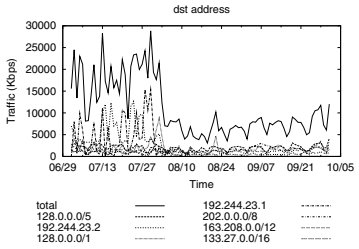


図 2.18. 宛先 IP アドレス (7 月–9 月)

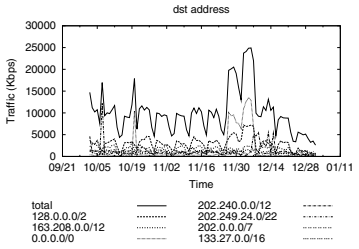


図 2.19. 宛先 IP アドレス (10 月–12 月)

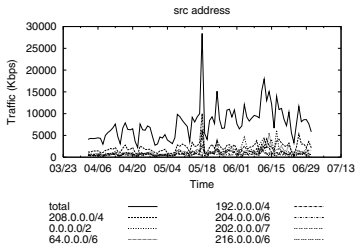


図 2.20. 送信元ポート番号 (4 月–6 月)

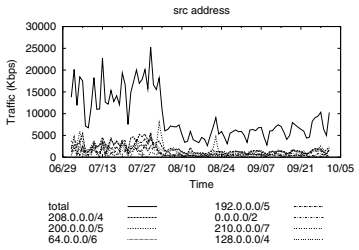


図 2.21. 送信元ポート番号 (7 月–9 月)

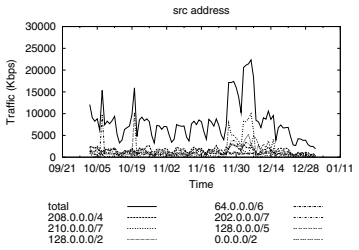


図 2.22. 送信元ポート番号 (10 月–12 月)

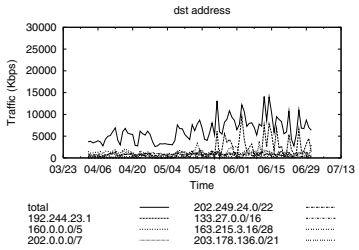


図 2.23. 宛先ポート番号 (4 月–6 月)

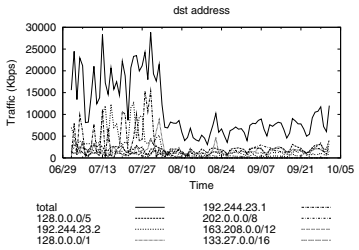


図 2.24. 宛先ポート番号 (7 月–9 月)

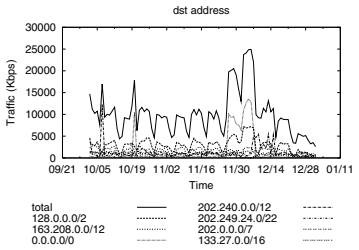


図 2.25. 宛先ポート番号 (10 月–12 月)

表 2.4. 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 2.8, 2.9, 2.10, 2.20, 2.21, 2.22	4:6:80	HTTP
図 2.8, 2.9, 2.10	4:6:6699	WinMX, Napster

2.3 結論

本章では、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/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 3 章　A High-speed IP meter and Its Experience at the 54th IETF

3.1 Introduction

As the number of Internet applications in critical situations increases, the quality assurance of the network infrastructure becomes more and more important. Various new services on the IP network, such as Video on demand and TV conferencing, also require the quality assurance of network services to improve customers satisfaction.

The operators of such networks have to maintain the network as reliable and available with sufficient performance. They are faced with various problems, which are becoming more and more difficult everyday. For example, long-established IP services such as mail, news and WWW, do not require real-time performance. However, real-time performance such as jitter in millisecond order is important for multimedia stream services. The new services evolving daily make the problem much more complex.

Although measurement technology is the key to maintaining the network condition and many important results have been obtained from enthusiastic research network engineers have to concede that there is room in the current measurement technology for improvement to meet the various requirements. Such requirements involve packet capturing from high-speed backbone networks, filtering techniques to select the target IP packet from others, accurate timing information gathering from world wide networks, etc.

To meet these requirements, a high-speed and accurate IP meter, *HIM*, has been developed. This paper presents the design policy and implementation of *HIM*. The experimental results with *HIM* have shown the current computer system's limitation, which cannot be analyzed without developing *HIM*-like equipment.

3.2 Related Works and Research Issues

Since the network characteristics are represented by various attributes, IETF's IPPM working group proposes a framework of IP performance metrics [159]. Their work is important for providing a baseline to compare the measured results by standardizing the attributes to be measured.

Surveyor [152] is a project to create a measurement infrastructure based on standard work being conducted in the IETF's IPPM WG. It measures the performance of the Internet paths among participating organizations. The project is also developing methodologies and tools to analyze the performance data. The WIDE MAWI working group [163] is making similar efforts. Their traffic data has been made open to the public. Their traffic repository is a good resource of various analyses.

Since the bandwidth of a backbone network increases rapidly, gathering information on network

behavior becomes difficult. Both high-speed data collection and large volume data require special technology and systems. NLANR [110] has a project to develop a large-scale data collection system as the infrastructure of various data analyses.

How to analyze the gathered data is also an important issue. CAIDA [36] is making various tools to analyze network data. Their visualization tools cover various analyses of network data.

Although the importance of this problem has attracted many researchers conducting work in the measurement area, their work tends to share a common characteristic: their data are gathered using a software-based approach. They tend to use PC hardware with special software designed to gather data from the network.

Most of the measurement software works are performed on the UNIX operating system or an operating system, which tends to have more overhead. Since the reliable timing unit of UNIX is 10 milli seconds, which is derived from the well-known clock tick rate of 100 Hz, a software-based approach tends to fail in gathering timing information of μ second order. Since some DV streams, for example, contain 125-μ second interval packets, a software-based approach cannot be used to evaluate the timing quality of DV data transmission.

A high-speed backbone network also creates a difficult problem for a software-based approach. Since the speed of a 1-Gbps backbone network is higher than the typical hard disk speed (less than 50 KByte/second, i.e., 800 Kbps), capturing data from such a high-speed network requires special system configuration. Even if the data required for analysis is a small subset of the traffic on the network, data selection raises another difficult issue for the software-based approach.

3.3 Design Policy and Implementation of IP meter *HIM*

Considering the above requirements and the limitations of current measurement technologies, a hardware-based IP meter, *HIM*, has been designed. The target specifications of *HIM* are:

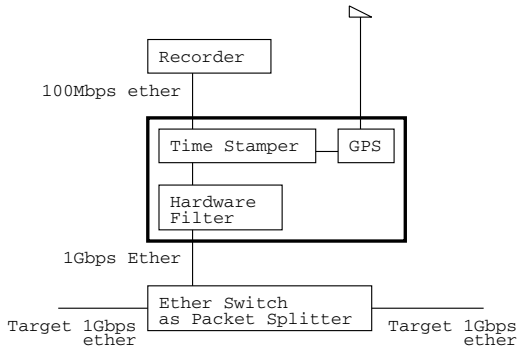


Fig. 3.1. Configuration of *HIM*

100-μ second order time stamp accuracy:
To measure the quality of multimedia services, the IP meter has to be able to record timing information in 100-μ second order.

Capturing speeds up to 1 gigabit per second:
To measure the backbone traffic, the IP meter has to be able to record data from a gigabit class network.

Wire speed filtering:
To select target data from backbone traffic, which involves unrelated data, the IP meter has to have a filtering mechanism, which does not affect the above specifications.

This section explains the implementation of *HIM*, which follows these specifications.

3.3.1 Over-All Structure

Figure 3.1 shows the configuration of *HIM*. It has 5 major components: i.e., packet splitter, hardware filter, time stamper, GPS timer and recorder. After the packet splitter duplicates all the packets from the target 1-gigabit Ethernet, the hardware filter selects the target packet. Then, the time stamper records the timing information using the GPS timer and the PC-based recorder records the observed data (original packets with timing information) for the latter analysis.

The processing time of the hardware filter (i.e., total processing time of Routing and Forwarding Application Specific Integrated Circuits, ASICs) is less than 5 μ seconds. The accuracy of the GPS

Table 3.1. Conditions to select IP packets

Name	Details
IP source address	IPv4 address with netmask
IP destination address	IPv4 address with netmask
Port number	Such as 80 for WWW, 25 for smtp

timer is also about 5 μ seconds (1 μ second for GPS accuracy+Time Generator processing time). To speed up development, software has to be used to implement the time stamper, which makes the final IP packet. The processing time of the time stamper is expected to be about 10 μ seconds. Thus, the total accuracy of timing information is expected to be less than 20 μ seconds. This accuracy has been evaluated in the experimentations described later in this paper.

The generated IP packet is sent to the PC-based packet recorder through 100-Mbps Ethernet. Since this 100-Mbps Ethernet and the hard disk of the PC is the bottleneck of *HIM*, the full capturing performance of current *HIM* is 100 Mbps.

It was possible to used a commercial Ethernet switch as the packet splitter, and a PC as the recorder. The hardware filter and time stamper are the key components of *HIM* and these two subcomponents can be used directly. For example, if the target network segment has a packet mirroring function, the packet splitter of *HIM* can be omitted.

3.3.2 Hardware Packet Filter

A hardware-based packet filter, which can select IP packets with a wire speed rate of gigabit Ethernet, has been developed. The hardware has two ASICs. One is the forwarding ASIC, which filters and forwards the captured packet to the time stamper. The other is the routing ASIC which decides the filtered packet.

When the forwarding ASIC captures a packet from a monitored network segment, this stores the packet in the packet buffer and asks the routing ASIC. The routing ASIC decides whether the packet is filtered or not. At this time, this ASIC refers to the filtering table. If the packet is forwarded, the forwarding ASIC sends it to the time stamper.

The filtering table column specifies the conjunction of the IP address and port numbers (See Table 3.1 for details). By specifying multi-table columns, the disjunctive conditions can be specified. A PC-packed packet recorder is used to set up the filtering table.

The packet forwarding ability of both the Filtering and Routing ASICs is 1 Mpps and the maximum total processing time of the Filtering and Routing ASIC is less than 5 μ seconds.

3.3.3 GPS Time Stamper

After the packet filter selects the target IP packet, the time stamper forms a new IP packet, which is constructed from the original IP packet and the timing information at the tail. To do this, the time stamper first checks the time information created by the GPS timer, then converts this time information into a format, which consists of the UTC time format and the time information of the nano-second order part.

The GPS timer creates time information using GPS. The GPS timer consists of the GPS receiver and time generator. The GPS receiver receives time information from the GPS satellite every 1-second and forwards it to the time generator. The time generator creates the UTC time using the information from the GPS receiver. It also creates the 100 nano-second order time information using its own crystal and register. When the time generator receives the information from the GPS receiver, it clears the value of the register, and it increments the value of the register every 100 nano

seconds. FPGA is used to increment the register value using the pulse from the crystal.

3.4 Performance Characteristics of developed IP meter

To evaluate the time stamp accuracy of *HIM*, 3 different tests have been conducted. The first test measured the time stamp accuracy by measuring the accurately sent packets. The second test measured the synchronization accuracy between *HIM* by comparing the time stamps of two *HIM* for a single packet. Finally, the accuracy of *HIM* was compared with a software-based IP meter.

3.4.1 Time Stamp Interval Accuracy

The first test measured the time stamp accuracy by measuring the accurately sent packets. A hardware-based packet generator is used to generate target packets (See Fig. 3.2). The hardware-based packet generator sends packets every 200 μ seconds. The accuracy of this interval was confirmed by observing the sent packets with a digital oscilloscope and the error of the interval was less than 1 μ second, which is the most detailed time resolution that can be used.

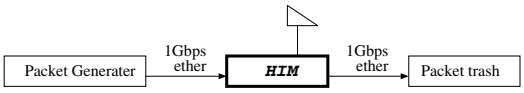


Fig. 3.2. Configuration for Time Stamp Interval Test

Figure 3.3 shows the time stamp interval observed by *HIM*. Since the target IP packets have an accurate interval, the jitter observed by this experimentation shows the time stamp accuracy generated by *HIM*. The average interval observed was 200 μ seconds and 99.5% of the observed interval had an error less than 20 μ seconds. This 20-μ second error comes from the software used in the time stamper and a fully-hardware-based *HIM* is being developed to eliminate this defect. But this new version of *HIM* is beyond the scope of this paper.

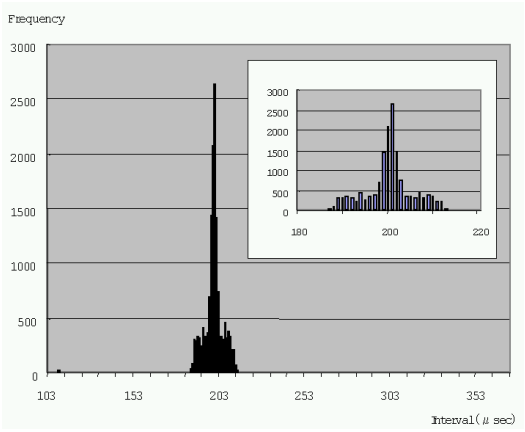


Fig. 3.3. Time Stamp Interval Accuracy

3.4.2 Synchronization Accuracy between *HIM*

The second test measured the synchronization accuracy between *HIM* by comparing the time stamps of two *HIM* for a single packet. Figure 3.4 shows the system configuration for this test. In this experiment, two *HIM* ware connected in a series. The packet generated by the time stamper of the first *HIM* is injected into the packet filter of the second *HIM*. By doing this, the first *HIM* adds the time stamp information at the tail of the original packet and forms a capsule packet. The second *HIM* adds the time stamp information again at the tail of the capsule. By calculating the difference of the two time stamps for the same original packets, it is possible to calculate the synchronization accuracy of two *HIM*.

Figure 3.5 shows the results. The average difference of the time stamps is 176 μ seconds and 99.5% of the observed interval has an error less

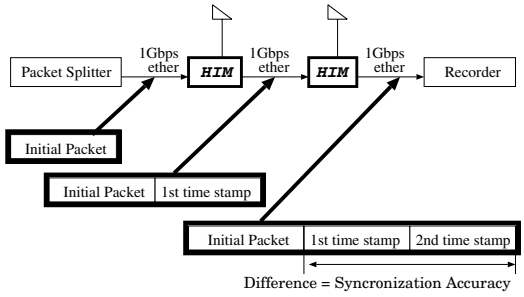


Fig. 3.4. Configuration for Synchronization Accuracy Test

than the 40 μ seconds. Since about 170 μ seconds is required to be processed by *HIM*, this result shows that the synchronization accuracy between *HIM* is within 50 μ seconds.

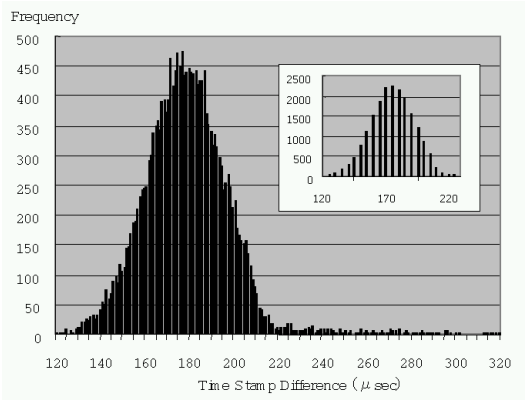


Fig. 3.5.　Synchronization Accuracy

3.4.3 Comparison with Software-based IP Meter

With *HIM*, video stream transmission was observed and a typical system configuration could send one digital video stream (30 megabits per second) with 100–125 μ second intervals within a 20-μ second jitter (Fig. 3.6, See next section for details).

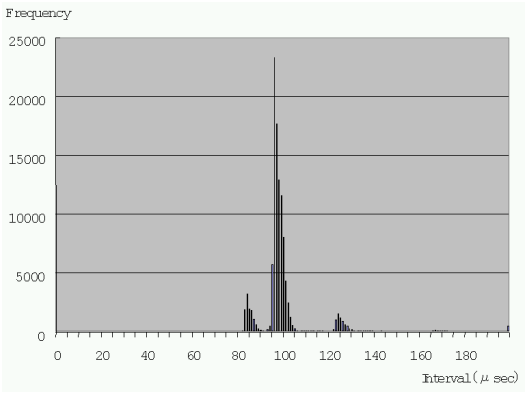


Fig. 3.6.　Result of *HIM*

Figure 3.7 shows the results for the same DV stream with the software-based IP meter. To observe 100 μ order timing data, the CLOCK TICK parameter in BSD UNIX was modified from 100 Hz to 10,000 Hz. However, the data by the software-based IP meter has two peaks, i.e. one for around 0 μ seconds and another for around 200 μ seconds. This is due to the feature of the software driver. The software-based IP meter uses the driver software of the operation system. And the driver used here handles every set of 2 packets at once. If the driver handles 2 packets at once, the software-based IP meter misinterprets the interval of these 2 packets as being 0 μ seconds. The interval of packets between other combinations is then misinterpreted as being 200 μ seconds. Even if the CLOCK TICK parameter was modified to make the basic timing unit inside UNIX to the μ second order, the rest of the huge software such as the driver requires enhancement so that it does not have this type of defect.

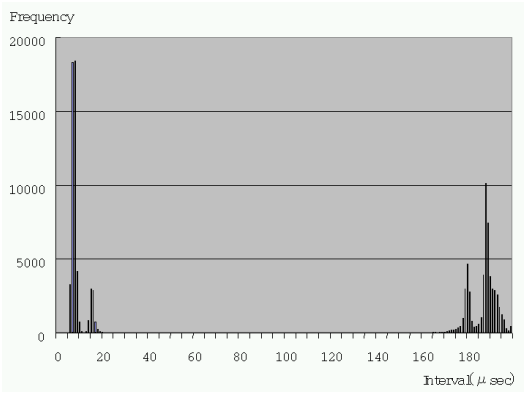


Fig. 3.7.　Result of BSD-Base IP Meter

In summary, these results show that the software-based IP meter cannot achieve the accuracy of 100 μ second order, which is necessary to evaluate the transmission quality of video stream.

3.5 Traffic Measurement and Analysis of 54th IETF Meeting at Yokohama

In the wide area network, there is much network traffic of various network users. Since each user uses the different network application, e.g. WWW, mail, streaming and so on, no one can identify the overall behavior of the network. However, it is important for the network operation to accurately measure their behaviors.

As an one step of measuring the backbone network traffic, we use the HiMs at an international meeting site (the 54th IETF meeting at Yokohama), which was held during one week. We choose this site as a typical site with high speed Internet backbone network. Several hundreds people attended the meeting and access the Internet. The meeting provided two backbone network lines (each has 100Mbps line speed) to access the Internet. Two HiMs collect the traffic in these lines. Each HiM monitors the traffics of each network line. After monitoring, each traffic data is merged so that we can analyze flows which are eventually forwarded using both 2 lines.

3.5.1 Outline of traffic

Average of traffic amount is about 1.28T bytes/day (Highest bit per second is almost 50M). Average of packet amount is about 260M packets/day (Highest packet per second is almost 11,000).

Figure 3.8 shows IPv4 Incoming packets and Figure 3.10 shows outgoing packets. Figure 3.9 and Figure 3.11 shows about IPv6. As shown in Figures, IPv4 still transfer major traffic.

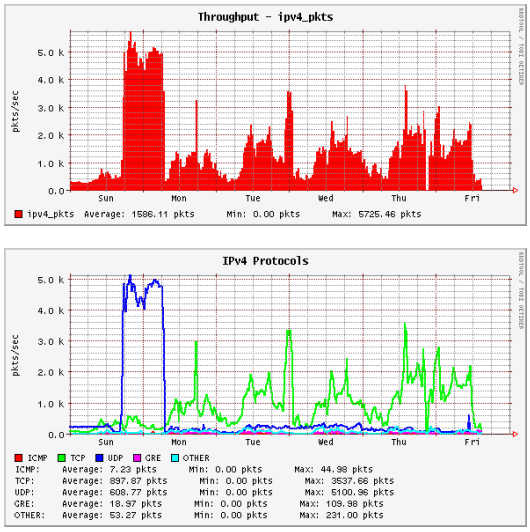


Fig. 3.8. IPv4 Incoming packets

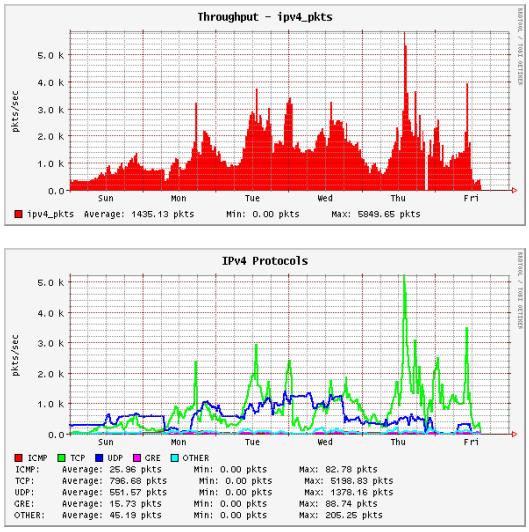


Fig. 3.10. IPv4 Outgoing packets

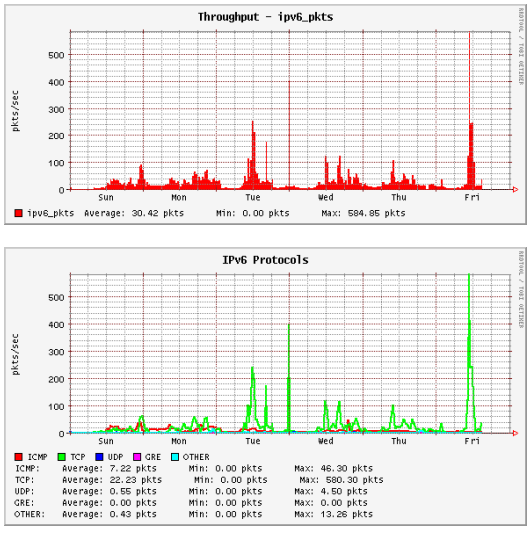


Fig. 3.9. IPv6 Incoming packets

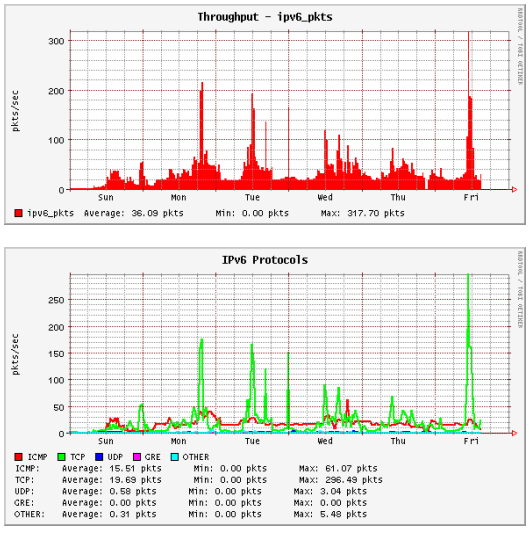


Fig. 3.11. IPv6 Outgoing packets

3.5.2 Individual flow behaviors

For the analysis, we follow the definition of the flow in RFC2722. Further, we select flows with more than 100 packets. The time interval to identify flow end was set to 3 seconds. The result data contains 30,043 flows in total.

Table 3.2 shows the 10 mostly used port and the number of flows of that port. Figure 3.12–3.15 show the examples of individual flow behaviors. Figure 3.12 is an example of HTTP flow which is mostly used during the experience. SSH (Figure 3.12) and POP3 (Figure 3.12) are commonly used also.

Table 3.2. 10 mostly used port

Port	# of flows
http	7793
ssh	6276
9875	1768
1214	957
pop3	813
10000	636
1031	452
639	404
microsoft-ds	382
6346	368

As shown in the data, the HTTP flow contains the request from the client in the beginning, and the succeeding HTML data from the server. POP3 client sends multiple requests in a single flow, i.e., in a single connection, and receives corresponding mail bodies. Since SSH is used to transfer data through tunnel, their independent behavior is different. Example shown in Figure 3.13 contains data from both end nodes.

Though we have found many unknown UDP flows, most of them seem to do one-way bulk transfer data (See Figure 3.12 for an example).

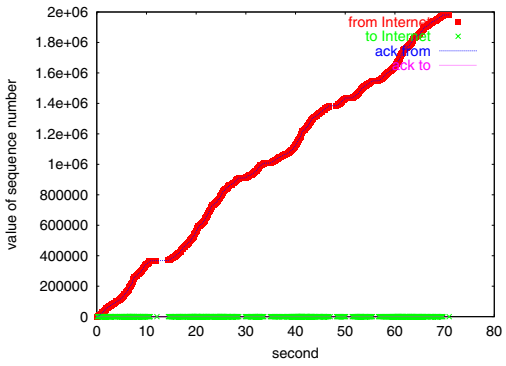


Fig. 3.12. Example flow of HTTP

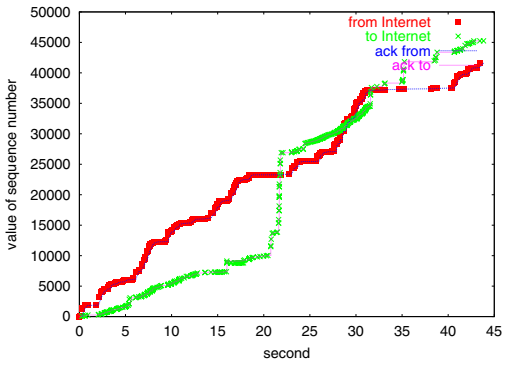


Fig. 3.13. Example flow of SSH

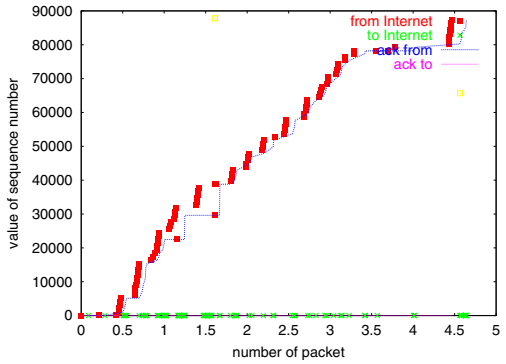


Fig. 3.14. Example flow of POP

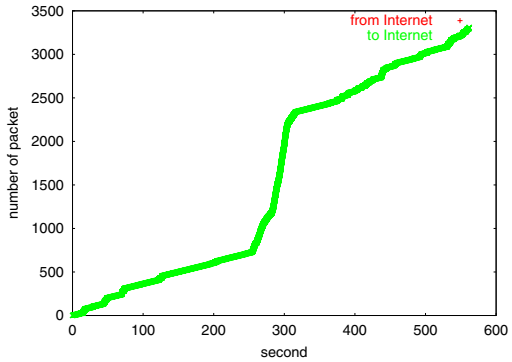


Fig. 3.15. Example flow of unknown UDP application

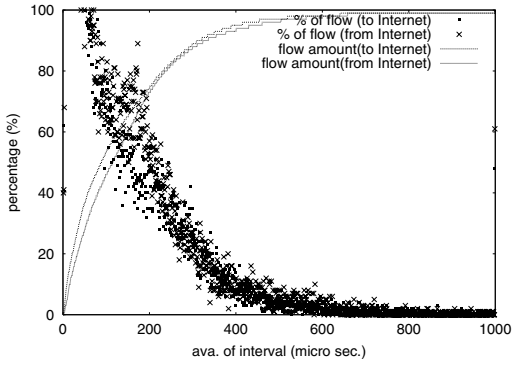


Fig. 3.16. Interval of TCP

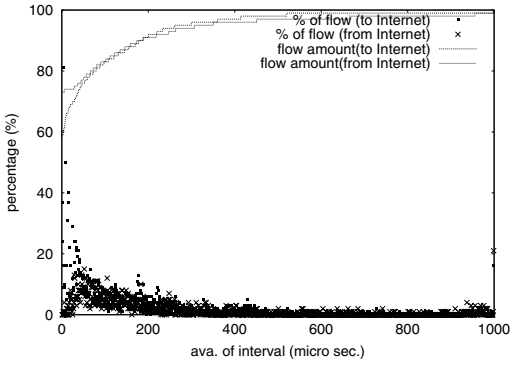


Fig. 3.18. Interval of UDP

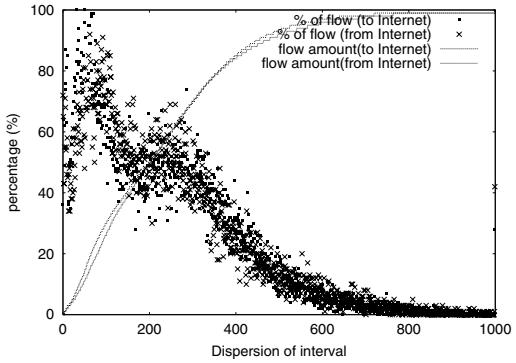


Fig. 3.17. Jitter of TCP

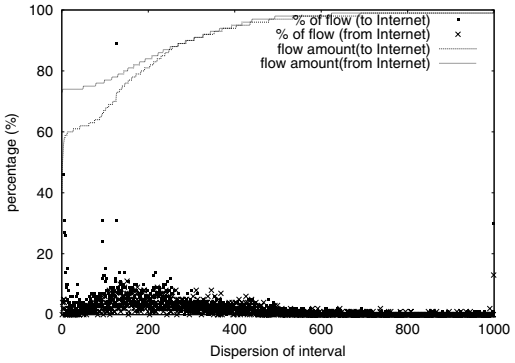


Fig. 3.19. Jitter of UDP

3.5.3 Aggregated flow behaviors

Figure 3.16–3.19 are the results of aggregated flow analysis. These figures show the average packet interval and interval jitter of TCP and UDP flows.

Figure 3.16 and 3.17 is the result of TCP flow analysis. Figure 3.18 and 3.19 is the result of UDP flow. Figure 3.16 and 3.18 show the average of interval time between two packets in the flow. X-axis is the average of time interval between two packets in flow. Y-axis is the percentage of flows. Dot marks shows the percentage of packets with the corresponding interval of 1 mill. second unit. Lines shows the accumulated numbers.

+ mark and upper line show the results for incoming data from the internet. ∗ mark and lower line show the results for outgoing data to the internet.

Figure 3.18 and 3.19 show the corresponding results for UDP data.

3.5.4 Characteristics of High speed backbone Network

Each application has its own transfer behavior. Although we can use HiM to analyze such behaviors, we could not find any general rule except following two phenomena

- Both average packet interval and interval jitter of TCP flows are longer than those of UDP flows.

  This seems to indicate the characteristics of TCP flow control. Since TCP flow is controlled to avoid internet congestion, their burstness is weaker than UDP flows.
- As for the UDP packets, the average packet interval and interval jitter of the incoming packets from the internet is shorter than those outgoing packets to the internet.

  As shown in Figure 3.18 and 3.19, UDP flows are burst data transfer in general. But burstness is stronger in the incoming packets

from the internet, i.e., 73% of UDP incoming packet intervals are less than 1 msec and 58% of UDP outgoing packet intervals are less than 1 msec.

The reason of the second phenomenon still requires further studies. But we suspect the existence of backbone routers which treat multiple packets transfer in bulk. If there exists a router or switch which transfer multiple packets in bulk, the average packet interval and interval jitter of TCP flows becomes shorter because such bulk transfer has shorter packet interval.

3.6 Conclusion

HIM, a high-speed and accurate measurement equipment for IP data packets, was developed to examine the network and its service qualities. The characteristics of this hardware-based IP meter *HIM* include:

- High-speed data capturing function up to 1 gigabit per second of data stream.
- Accurate time stamp function with 20-μ second accuracy.
- Hardware-based filtering function, which can select target IP packets without decreasing the above 2 performances.

The experimental results with *HIM* showed: In high speed backbone Network, both average packet interval and interval jitter of TCP flows are longer than those of UDP flows. And as for the UDP packets, the average packet interval and interval jitter of the incoming packets from the internet is shorter than those outgoing packets to the internet.

Although, we need further studies, TCP flow controll seems to weaken their burstness than UDP flows, and the burstness of UDP flow is made stronger by the characteristics of routers in wide area network.

第 4 章　第 54 回 IETF における無線利用モニタリング

4.1 はじめに

本章では、横浜で開催された第 5 回 IETF の会場で運用された無線ネットワークの計測結果について述べる。

無線基地局は会場のすべてをカバーするように配置され、また、会場に付属するホテルにも無線基地局を配置した。

第 52 回 IETF では 1900 人が参加し、そのうちの 87% にあたる 1650 人が無線を利用していた。これほどの人数が密集した環境で無線を利用することは今までになく、今後 Hot Spot などのサービスが一般化していく上で、重要な先行事例となる。

4.2 利用状況の視覚化

今回、基地局の利用状況をリアルタイムで視覚化するアプリケーションを開発し、監視に用いた。図 4.1 はそのスクリーンショットである。

図中に見える円が基地局である。ユーザー数の増加にしたがって、円の大きさが増え、利用状況が直感的に把握できるようになっている。

ユーザー数の取得は SNMP を利用して行ったが、基地局とリンクしているユーザー数などの 802.11 に特化した情報は、標準化されている MIB にない。これらの情報は enterprise MIB(ベンダが固有に定義できる MIB) 上で定義されており、そのために対象とする基地局の種類ごとに情報収集の方法を変えなければならなかった。

今回、基地局は Cisco 製と富士通製 (Lucent OEM) の 2 種類を使った。それぞれ取得方法が違うため、それらの違いを吸収できるように収集部と集計部を別々にしてモジュール化をしている。

ユーザー数の情報はアプリケーションから要求を出すたびに取得し、そのときの状態をリアルタイムに視覚化する。

4.3 集計情報

ユーザー数の状況は、アプリケーション上でオンデマンドに取得するものに加えて、15 分ごとにスナッ

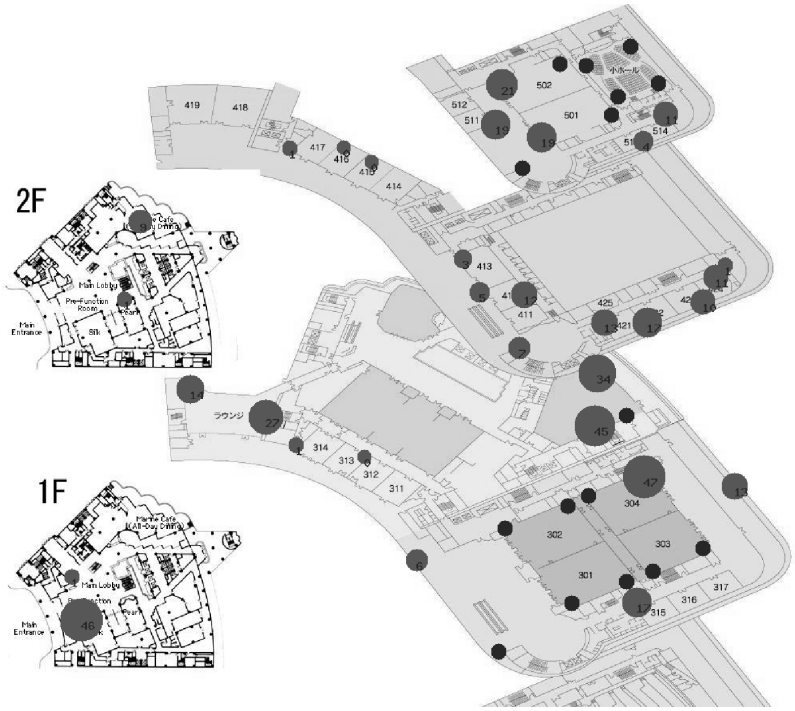


図 4.1. 管理用ツール

プショットを取って保存した。それらの集計結果を以下に示す。

会場などでは、会期中は比較的ホスト数は安定していたが、ホテルのロビーや端末室などは、無線デバイスつきノート PC を持ったまま出入りする人が多かったため、数の変動が多い。

また、全体集計を見るとで 7/17 18:00 ごろに異常な結果 (約 2800) が得られている。この時間は、plenary(全員出席の会議）があり、人が集中したため、アクセスポイントが高負荷によって異常動作を起こしていた。そのため、複数の無線基地局で重複してユーザー数がカウントされていた。この時間中は無線によるアクセスは非常に不安定な状態となっていた。

また、7/16 から 7/17 にかけてホテル 2F のユーザーが 0 となっているが、これは基地局へのネットワークの配線が不調となって、基地局にアクセスできなかったからである。

4.4 結論

本章では、横浜 IETF において運用された無線基地局の計測について述べた。

大規模な無線環境を構築するにあたっては、まずその利用状態について把握し、それに対応して機器の設置場所の選定などを含めた設計を行う必要がある。

今回の計測では利用状態に関する情報を収集し、その傾向を明らかにした。この情報は、今後このような無線環境を構築する際に有益なものであると考えられる。

第 5 章　IETF54 会議における無線 LAN 運用結果から得られた問題点

本章では、高負荷環境における 802.11b の問題点を報告する。

我々は、神奈川県横浜市において開催された第 54 回 IETF 会議における無線ネットワーク運用を行った。本無線 LAN ネットワークは、1000 人以上のユーザが、サービスエリア内において、シームレスに安定して利用できることを目標に設計された。それには、会場と隣接するホテル内において、単一のレイヤー 2 セグメントを IEEE802.11b 技術をベースに提供すること、802.11b の電波チャネルの干渉を最小限となる基地局の設置、各基地局における利用ユーザ数の把握と接続管理の 3 つの要件を満たす必要があった。

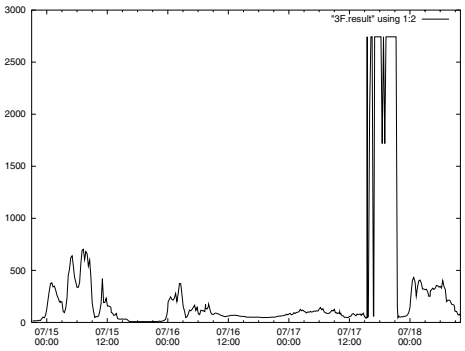


図 4.2.　会場 3F

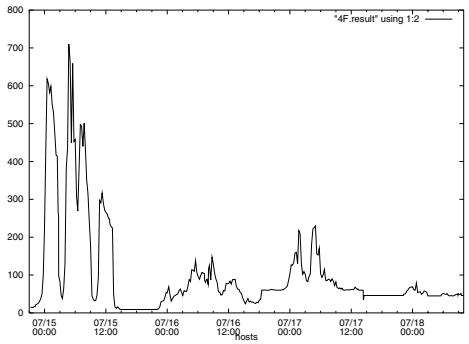


図 4.3.　会場 4F

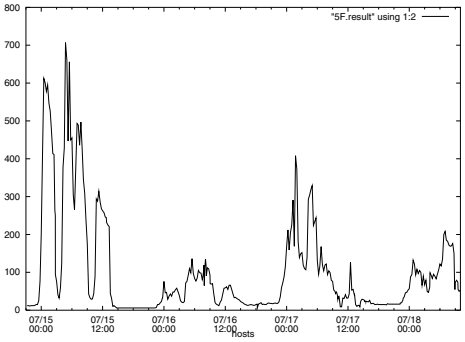


図 4.4.　会場 5F

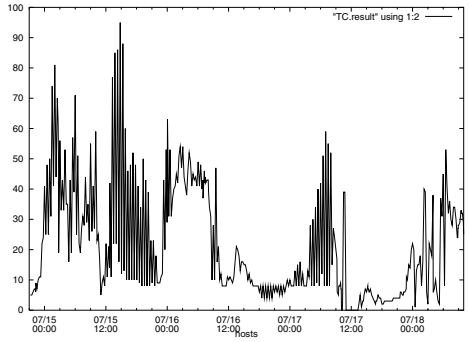


図 4.5.　端末室

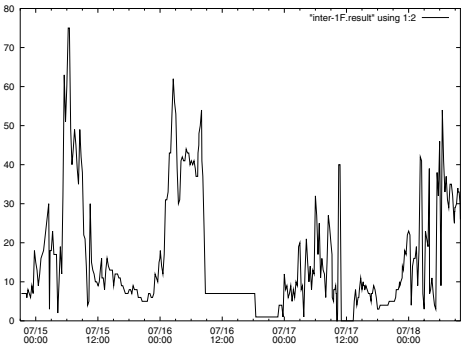


図 4.6.　ホテル 1F ロビー

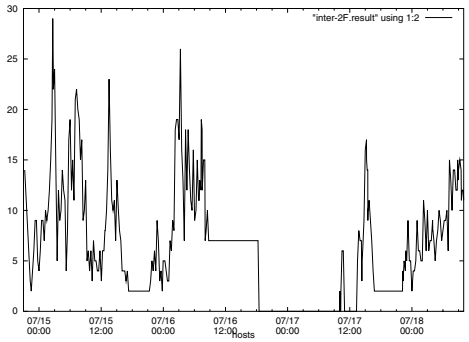


図 4.7.　ホテル 2F ロビー

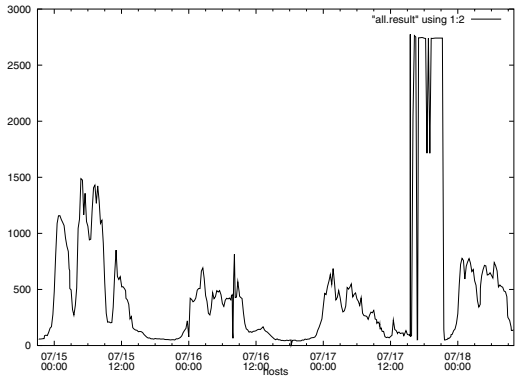


図 4.8.　全体の集計

そこで、本運用では、802.1Q(VLAN) と xDSL を利用した会場・ホテル内のレイヤー 2 ネットワークの構築、会場内実測と干渉測定に基づく基地局の配置、RADIUS を利用したユーザ管理の 3 点を実行することによって、先の要件を満たしたネットワークを運用した。

その結果、プレナリセッションなど、利用ユーザ数が 500 以上と極端に集中するエリアにおいて、ネットワーク性能が急激に低下し、場合によっては、基地局のハングアップなどの不測の事態が発生した。

5.1 はじめに

IEEE 勧告の 802.11b は、各ベンダー製品間で互換性の高い無線技術となった結果、低価格化がすすみ爆発的に普及している。したがって、昨今の国際会議やキャンパス・ネットワークなどのネットワークへの接続メディアとして利用される機会が多くなっている。そして、IETF 会議においても、Ethernet 等有線によるインターネット提供よりも、無線ネットワークを利用したインターネット提供がおおきなウエートを占めるようになった。

802.11b は、周波数帯域として 2.4GHz 帯を使用し、同帯域上で、1〜14 チャネルを割り当てている。各国において利用できるチャネルは異なっており、日本では、1〜14ch、アメリカ・カナダ (FCC) では、1〜11ch までが利用可能となっている。

そして、各チャネルは、お互いに干渉しており、独立した帯域を確保するためには、4 チャネル分の帯域をあける必要がある。従って、干渉のなく同時に使えるチャネルは、1-6-11ch か、2-7-12ch か、3-8-13ch か、4-9-14ch となる。日本においても FCC の認可にあわせて、一部の製品は、1〜11ch までしか利用できないデバイスがあることや、国際的なユーザが利用する事を考えれば、FCC 認可に沿って、802.11b で利用するチャネルは、1-6-11ch の組み合わせのみしか利用できない事となる。

802.11b は、ノードが基地局を利用する際に、アソシエーションを行う。このアソシエーションを経て、ノードから基地局を経由した有線間の通信が可能となる。本運用で用いたすべての基地局は、このアソシエーションをノードの MAC アドレスを用いた RADIUS 認証の可否結果に同期させることが可能となっている。本運用では、各利用ノードの MAC アドレスを認証し、記録することによって各基地局での利用数や、不正ノード (RA の不正送出や、不正な DHCP サーバ等) の排除が可能なシステムを運用した。

5.2 RADIUS を用いた観測・管理システム

本システムは、各基地局のノード数把握、不正ノード (DHCP サーバや、RA を送出することによって、ネットワークに悪影響をあたえる可能性のある) の遮断、不正アクセスを行ったノードの追跡が可能である。

このシステムは、3 つのコンポーネットで構成されている。

1. 基地局の RADIUS 認証機能
   基地局 (Lucent 社 AP1000/2000、Cisco 社 AP12000) は、アソシエーションする際に、ノードの MAC アドレスの認証を RADIUS サーバへ要求をする。RADIUS サーバで認証が行われば、アソシエーションが行われ基地局が利用可能となる。
   なお、基地局の RADIUS 認証のキャッシュ保管期間は 15 分とした。 従って、各基地局において一旦認証か完了した MAC アドレスは、再度アソシエートする際、キャッシュ保管期間が満了するまで、RADIUS サーバへの問い合わせを行わない。
2. RADIUS サーバとデータ管理部
   基地局からの MAC アドレスを認証する。不正ノードは、その MAC アドレスをサーバ上で登録することにより、認証を拒否するできる。これより、不正ノードは本ネットワーク全域においてネットワークを利用することは出来なくなる。
3. データ集計部
   RADIUS の認証データを解析し、ノード数や各ノードの移動記録を行う。その結果はデータベースとして管理され、データベースを元に HTML 化を行い利用ユーザの把握などか可能なインターフェースを提供する。

以上のシステムを FreeBSD 4.6R 上で実装した。

IETF 会議上の無線ネットワーク上で、本システムの運用を 2002 年 7 月 15 日午前 0:00 から 2002 年 7 月 20 日午前 6:00 までの間おこなった。

5.3 運用結果

本システムから得られた全基地局

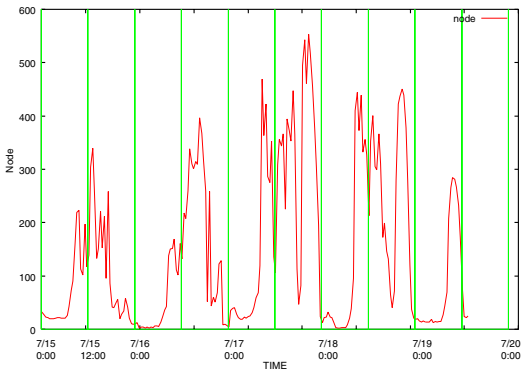


図 5.1.　全基地局におけるノード数の変化

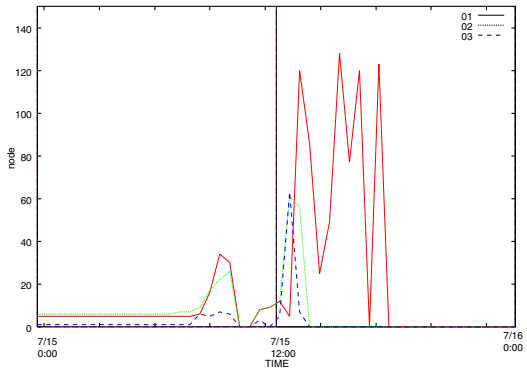


図 5.2.　過負荷状態におけるノード数の変化

(AP1000/2000/Cisco) におけるノード数の変化を図 5.1 に示す。

図 5.1 の横軸は日時、縦軸は、観測を行った全基地局におけるノード数の総和を示す。このノード数の推移は、IETF のスケジュールに合わせたノード数の変化を記録している。

午前 7 時位からノード数が増加し、セッション中は増加傾向である。そして、昼食・夕食といったセッションの合間にはノード数は減少する。これは、食事などのために、無線ネットワークサービスエリア外へと移動を行ったからといえる。

一方、セッションの合間におけるノード数の変化は、昼食・夕食時に比べて少ない。これは、ノードが基地局をシームレスにハンドオーバし、ネットワークをそのまま利用できる環境が実現出来ていたことから、ユーザがノードを停止することなく移動した結果だといえる。

このように、このグラフの変化は、シームレスなネットワーク環境を実現がユーザの行動に反映されているといえる。

ユーザー数のピークを迎えたのは、7/17 PM6:00–7/18 AM0:00 の間 (IESG プレナリ) である。その数は、550 ノードであった。各基地局に 100 ノードを超える接続があり、各基地局において、過負荷状態となった。そして、いくつかの基地局はハングアップした。これは、図 5.1 において、山の変動があるのは、基地局がハングアップしたからである。

5.4 基地局性能と波状破綻

15 日午後。ノード集中が会場 301–304 において発生した。この結果、基地局が過負荷状態となり再起動を繰り返した。このため、会場 301–304 では、無線ネットワークを利用したインターネットへの接続が出来なくなる状態が続いた。

図 5.2 は、その際に記録したノード数の変化である。なお、負荷の軽減のために、01 番基地局をのぞいて、該当するエリアの基地局管理を中止した。

図 5.2 の激しいグラフの変動は、基地局が過負荷状態となって、ハングアップし、何度も再起動したことを示している。02 番と 03 番基地局が 15 日午後セッション開始直後に過負荷状態となり再起動した結果、02 番と 03 番でアソシエートしていたノードがすべてほぼ同時に 01 番基地局へ切り替わった。その結果、01 番基地局も過負荷状態となり、再起動が発生した。

このように、ノードが集中した部屋においては、1 基地局の破綻がその部屋をカバーする他基地局に、アソシエーションの影響を与える結果となった。場合よっては、すべての基地局が再起動する波状破綻状態となる。そして、本運用においては、セッション終了まで不安定な状態を繰り返した。

波状破綻が発生した 15 日における部屋 301–304 の基地局は、Lucent AP-1000 を利用した運用を行っていた。ここで、AP1000 の基地局の最大許容ノード数について考察を行う。

15 日午前中において、40 程度のノード数では問題は発生しなかった。15 日午後のセッションにおいて、ノード数が 60 を超えた時点で頻繁に再起動が発生するようになった。01 番基地局での最大ノード数が 120 であったから、ユーザ数が 60〜120 の間が AP-1000 の最大許容ノード数であることがわかる。

部屋 301–304 における各基地局のノード数の分散は、15 日午前においては非常にスムーズであった。よって、AP-1000 の場合、50 ノード/基地局程度の

使用内で収まるのであれば、破綻することはない。ただし、パケットロスがどの程度発生していたのかは、不明である。

次に、波状破綻を防ぐために、部屋 301–304 の基地局を AP1000 から、より性能の高い AP-2000 へ変更した。

そして、AP-2000 による運用結果は次のようになった。

図 5.1 から、17 日 (ピークを迎えた IESG プレナリの開催時間帯を含む) の会場 301–304 に配置した基地局における結果を図 5.3 に示す。

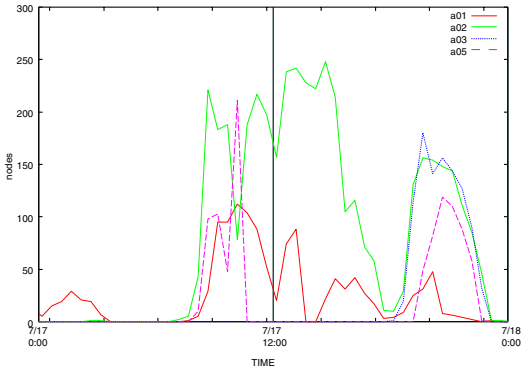


図 5.3. AP-2000 過負荷状態におけるノード数の変化

これには、基地局 (AP-2000) が再起動などによって、ユーザのハンドオーバが行われた際の各基地局でのノード数の変化が記録されている。

17 日午前中の A02 番基地局を再起動を行った。この際に、A02 番基地局のユーザが、同一部屋内に設置された A05 基地局へ一斉にハンドオーバがかかったことが観測されている。

午後は、A05 基地局が不安定になり、一切の観測ができていない。A02/A03 基地局のみが稼働している状態となった。観測はできていないが、ノードのアソシエーションは可能であった。したがって、これは、AP-2000 の RADIUS 認証に関して不都合が発生してたのでは、ないかと推測している。

IESG プレナリ開始当初は、順調に増えるユーザ数に対して、負荷が各基地局に分散し、安定した運用が行うことができていた。しかし、各基地局におけるノード数が 150 を超えたあたりで、A03 基地局が再起動した。したがって、150 ノード位が AP-2000 の利用可能なノード数の限界値と言うことができる。

5.5 ノード数とパケットロスの関係

7/18 の IAB プレナリ中におけるノード数の変化とパケットドロップ率を記録したグラフを図 5.4 に示す。パケットドロップ率は、各基地局配下のノードに対して、有線側のネットワークから、ICMP ECHO/REPLY の結果を基にしている。

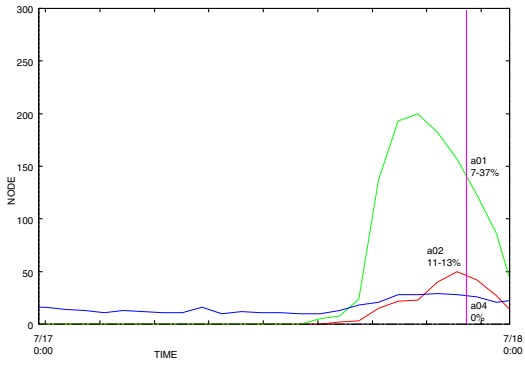


図 5.4. ユーザ数とエラーの関係

145 ノード程度 (A01 基地局) では、7–37%と非常にばらついた結果となった。この程度のパケットロスの環境では、インターネットを快適につかえる環境とはいえない。

50 ノード程度 (A02 基地局) では、パケットロス率は、11–13%となり、そのばらつきは少ないが、ノード数が少ない割にドロップレートが高い。この原因の調査のためには、バースト転送しているノードがいるかどうかといったトラフィック観測が必要であったといえる。また、145 ノードの場合とおなじく、インターネットを快適につかえる環境とはいえない。

そして、25 ノード程度 (A04 基地局) において、パケットロスの問題はなかった。

従って、AP-2000 において、150 ノード程度までは、アソシエートの維持が可能である。よって、過負荷による再起動は発生しない。しかしながら、快適な通信環境を実現するためには、1 基地局あたり 25 ノード程度となる。また、干渉のない 1-6-11ch の 3ch を利用して、75 ノード程度の接続とすれば、利用者に対して、快適な 802.11b 環境を提供することができる。

5.6 今後の課題

802.11b によって、ネットワーク接続を提供出来

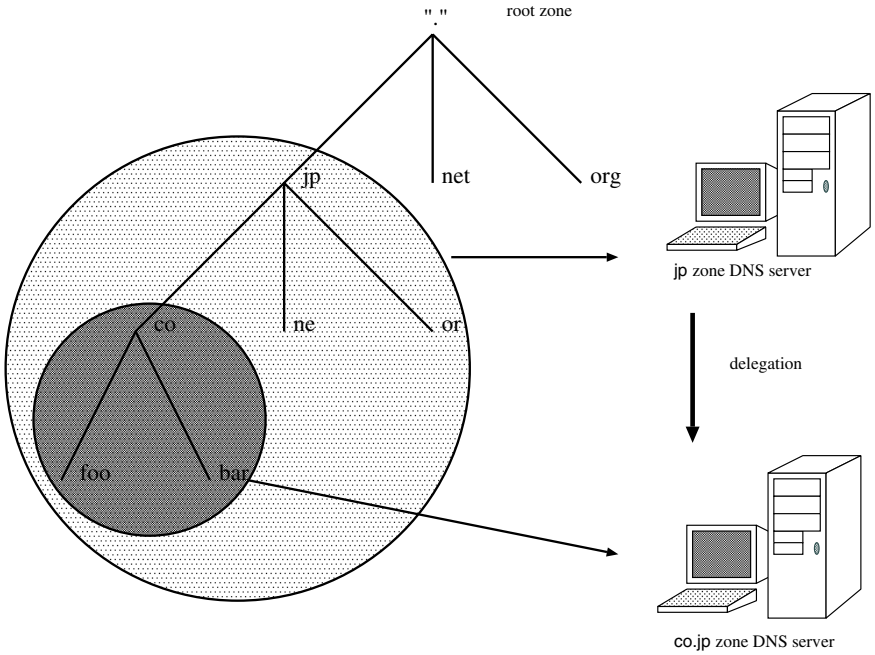


図 6.1.　DNS の木構造

るノード数は、基地局の性能や利用可能なチャネル数によって確定する。快適な利用を考えるのであれば、一基地局あたり 25 ノード程度である。このため、部屋などの閉じた空間においては、干渉のない利用を考えると、3ch 分の基地局設置となり、最大で 450 人程度の利用となる。

その許容ノード数の限界点は、IETF 会議の IESG プレナリといったセッションにおけるユーザ数を考えれば、低いと言わざるを得ない。したがって、このような結果を基に、無線ネットワークの規格制定部会へのフィードバックや、無線ネットワークの収納の限界数と利用ユーザの予測を十分に考えた上での、基地局の設置と管理を行わなければならないといえる。

十分なノード管理が出来ない場合は、過負荷状態において、十分なネットワークサービスを提供出来なくなるので、無線ネットワークのみではなく、有線での提供も十分考慮した設計を行わなければならないことが判明した。

第 6 章　ルートネームサーバならびに ccTLD サーバの応答時間計測

本章では、dnsprobe というツールを用いて、ルートネームサーバならびに ccTLD ネームサーバの計測を行った結果について報告する。

6.1 背景

インターネットの基盤を担うサービスとなっているのが DNS である。DNS は図 6.1 に示す通り、木構造のデータベースとなっており、その根本に存在する DNS サーバはルートネームサーバと呼ばれる。現在、ルートネームサーバは全世界に 13 台存在し、地理的に分散して存在している。

ルートネームサーバは、6 台が米国の東海岸に、4 台が西海岸に、2 台がヨーロッパ方面に、そして 1 台が日本に存在する。

また、ccTLD ネームサーバは、各国のトップレベルドメイン情報を保持するネームサーバである。日本ならば、「JP」というドメインを保持する DNS サーバが ccTLD サーバとなる。この ccTLD サーバは、全世界で 600 台以上存在する。

これらの DNS サーバは、DNS のサービスにおいて重要な役割を担っている。ルートネームサーバは、DNS データベースの始点となる DNS サーバであり、名前解決の起点となるサーバである。全てのドメイン名はこのサーバから始まることとなる。もし、なんらかの障害によって、ルートネームサーバに到達できない、もしくはルートネームサーバのサービスを受けることができない場合には、最悪名前解決ができない場合も発生する。また、ccTLD サーバは国別に存在しており、その国のドメイン名の起点となるサーバである。したがって、ある国の ccTLD サーバに障害が発生すると、最悪その国のドメインに属するドメイン名は、名前解決できないという事態が発生する。すなわち、ルートネームサーバならびに ccTLD サーバはともに、DNS のサービスを構築する基盤となっているサーバであると言える。

6.2 本計測の目的

本計測では、ルートネームサーバと ccTLD サーバへの到達性を計測した。前述した通り、ルートネームサーバと ccTLD サーバは DNS において重要な役割を担っている。

そこで、これらのサーバが全世界からどの程度の到達性を持っているかを測定し、現在どの地域からの到達性が乏しいのかという事を把握するのが第一の目的である。

また、今後ルートネームサーバや ccTLD サーバを拡張していく際に、どの地点に増設するべきか等の指針を得るのが第二の目的である。

そのため計測を行う際は、ある定点からのみ行うのではなく、なるべく多くの地点からの計測を行った。

6.3 dnsprobe

本計測では、dnsprobe というツールを用いて行った。このツールは、C 言語で記述され、各種 UNIX OS ならびに Windows 上にて動作するツールである。本ツールは http://mawi.wide.ad.jp/mawi/dnsprobe/rootprobe.tar.gz から入手することができる。本ツールは、ルートネームサーバならびに ccTLD サーバに対して DNS の query を送信し、返答があるまでの RTT を計測することが可能である。また、本ツールの特徴として、一般ユーザが気軽に手元で計測できるという点が挙げられる。管理者権限を持たなくとも dnsprobe を走らせることができるため、世界中のあらゆる地点から、分散して計測することが可能となる。本ツールによる計測例を、図 6.2 に示す。この例は、133.93.XX.1 から 13 個のルートネームサーバへの到達性を示した結果である。このようなレポートを定期的に一定時間行う。結果はメールもしくはファイルに出力することが可能である。

```
1026688779 133.93.XX.1 eth1 A: rtt 210 ms
1026688785 133.93.XX.1 eth1 B: rtt 159 ms
1026688790 133.93.XX.1 eth1 C: rtt 250 ms
1026688793 133.93.XX.1 eth1 D: rtt 180 ms
1026688798 133.93.XX.1 eth1 E: rtt 110 ms
1026688803 133.93.XX.1 eth1 F: rtt 130 ms
1026688808 133.93.XX.1 eth1 G: rtt 228 ms
1026688812 133.93.XX.1 eth1 H: rtt 190 ms
1026688816 133.93.XX.1 eth1 I: rtt 268 ms
1026688820 133.93.XX.1 eth1 J: rtt 210 ms
1026688827 133.93.XX.1 eth1 K: rtt 264 ms
1026688830 133.93.XX.1 eth1 L: rtt 130 ms
1026688837 133.93.XX.1 eth1 M: rtt 6 ms
```

図 6.2. dnsprobe の計測結果例

6.4 ダイヤルアップによる計測

本計測は、多地点から DNS サーバの計測を行うことを目的としている。多くの地点から計測するためには、多地点に dnsprobe のための計測ホストを設置するか、もしくは dnsprobe を走らせることのできるホストにアカウントを有しているユーザに依頼して、dnsprobe を走らせてもらうことによって計測が可能となる。しかし、実際に計測を依頼するためには、個々のホストのユーザへの交渉が必要であり、容易に実現できない場合もある。そこで本計測では、ダイヤルアップによる計測によってデータを補うことにした。

具体的には、遠隔地にあるダイヤルアップポイントにダイヤルアップ PPP を行い、接続された状態で dnsprobe を動作させ、計測した。しかし、ダイヤルアップの場合電話回線を利用して遠隔地に接続しているため、遠隔地に設置していあるホストにて dnsprobe を動作させる場合に比べて、電話回線を利用しているだけの遅延が発生する。本計測ではこの遅延を補正して、直接遠隔地で計測したデータと同

等に扱い、分析を行った。

まず、ダイヤルアップによる計測が、遠隔地で直接測定した結果と同じ傾向を示すことを確認した。図 6.3 は、慶應義塾大学藤沢キャンパス (CNS) 内に存在するホストから直接ルートネームサーバに対して dnsprobe を実行した結果と、日本国内から同キャンパスへダイヤルアップ接続を行って dnsprobe を実行した結果の比較である。さらに、図 6.4 は WIDE Project LosAngeles NOC 内に存在するホストから dnsprobe を行った結果と、日本国内から同 NOC にダイヤルアップ接続して実行した結果である。

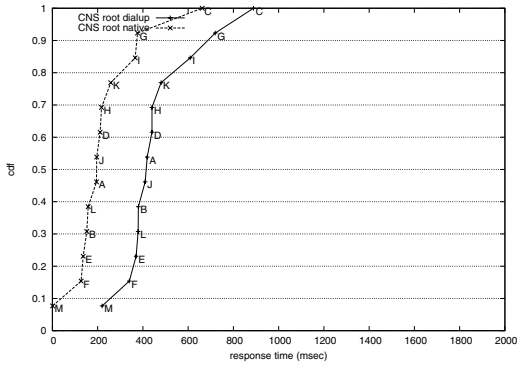


図 6.3.　CNS からの dnsprobe 結果 CDF グラフ

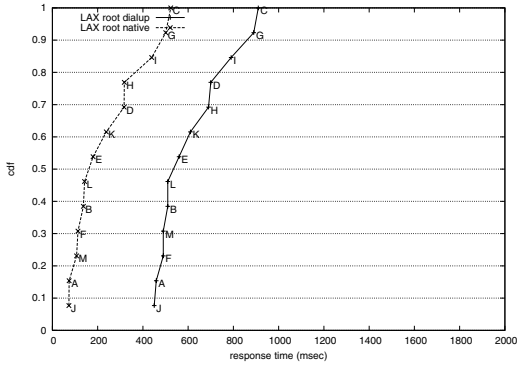


図 6.4.　LosAngeles NOC からの dnsprobe 結果 CDF グラフ

これらの計測結果から、ダイヤルアップによる計測は、遠隔地にあるホストからの直接の計測と同じ傾向を示すことがわかった。

6.5 ダイヤルアップデータの補正

次に、ダイヤルアップにて測定したデータを補正する際、どのような補正値を用いるかについて検討した。本計測では、ダイヤルアップ接続をした地点から最も近い ccTLD サーバまでの遅延を、計測した値から差し引くことによって、データの補正を行う方法を採用した。

各国の ccTLD サーバは、通常複数のサーバによって構成されており、その中の少なくとも 1 つのサーバは自国もしくは自国からネットワーク的に近い位置に存在していると仮定した。ダイヤルアップ地点から最も近い ccTLD サーバへの遅延を除くことによって、ダイヤルアップ地点に最も近いバックボーンからの、DNS サーバへの到達時間を算出することができると考えたからである。

また、本計測では ccTLD サーバの分布に関しても調査した。ccTLD サーバが物理的にどの位置に存在しているかを、サーバの IP アドレスと、whois 結果、ならびに traceroute の結果から判断し、まとめたのが図 6.5 である。

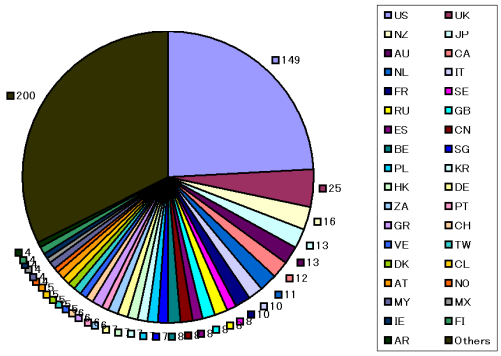


図 6.5.　ccTLD サーバの物理的配置

計測時点 (2002 年 3 月) で 601 台存在した ccTLD サーバのうち、149 台 (24.2%) が米国内に存在し、25 台 (4.1%) が英国内に存在した。大半は米国と欧州に存在しているが、200 台 (32.5%) のサーバは米国と欧州以外の国に存在していた。この結果から、ルートネームサーバに比べて、ccTLD サーバの方が世界各地に分散していることがわかった。この調査結果から、ccTLD サーバは世界中に分散して存在しており、ダイヤルアップ地点からもっとも近い ccTLD サーバを指標として利用することが可能だと考えた。

6.6 計測結果

dnsprobe にて、27 地点からルートネームサーバを計測した結果を表 6.1 に示す。この表の値は、継続的に計測した値の中間値である。なお、* マークのついている計測地点は、ダイヤルアップによる計

表 6.1. ルートネームサーバ応答時間の中間値 (msec)

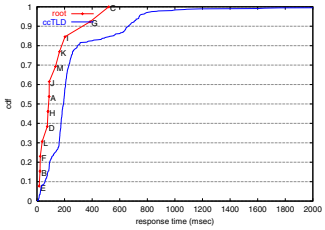
measurement	root servers												
point	A	B	C	D	E	F	G	H	I	J	K	L	M
OR	88	22	520	75	16	24	385	80	203	89	163	38	134
CA	79	21	545	67	2	2	374	72	183	79	152	24	123
CA	72	135	521	315	178	111	499	316	437	71	236	140	105
PA	2	70	430	6	64	76	315	4	116	3	79	75	192
MD	4	67	477	1	70	82	275	5	135	2	89	92	189
MA	22	76	449	9	70	82	200	15	131	23	93	94	192
ca*	140	200	570	140	371	181	461	160	220	120	191	200	330
mx*	110	91	101	100	131	100	290	90	200	81	170	100	211
uk	190	179	542	105	170	170	310	114	57	110	72	184	254
fr	116	188	540	108	193	148	397	152	32	112	32	179	251
ch	96	178	514	112	163	158	258	115	58	96	27	199	300
it*	200	251	630	150	270	220	347	160	100	170	70	220	331
pl*	170	220	660	140	361	200	361	150	90	150	80	230	356
ua*	180	501	620	440	270	250	620	451	350	160	350	500	590
cn*	280	401	930	220	551	400	591	470	371	480	351	151	421
cn*	750	670	1190	720	250	360	910	720	820	710	521	660	540
kr*	310	220	980	291	281	201	671	290	400	291	360	231	220
jp	178	140	614	169	102	100	430	170	270	170	230	137	1
nz	209	137	648	202	146	135	434	206	307	201	270	150	160
au*	360	270	800	381	390	250	705	320	480	321	440	250	200
za*	348	388	808	308	489	378	498	298	338	308	378	389	508
ke*	329	359	489	250	-	340	480	369	399	350	360	330	490
dz*	210	280	630	181	250	250	351	180	140	180	100	280	350
br*	140	161	541	111	161	151	101	101	211	101	181	181	251
br	140	198	555	149	190	194	327	125	248	141	216	196	303
ar	171	203	613	163	222	220	364	167	270	163	243	203	322
cl*	140	220	571	140	210	180	481	140	250	140	220	181	310

測である。

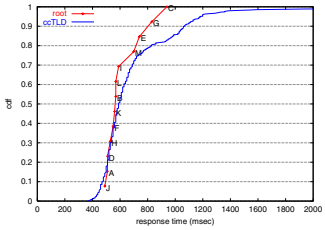
この結果から、オセアニアやアフリカ、南米、オーストラリアといった、ルートネームサーバのない地域からは、100 ms 以内で応答を得られるルートネームサーバが存在しないことがわかった。また、北米はルートネームサーバが複数存在するため、良好な応答時間でルートネームサーバからの応答を得られることがわかった。アジア地域では、日本に存在する M ルートネームサーバからの応答時間が一番短く、それ以外のサーバへの応答時間は 100 ms を超えるものが多いことがわかった。本計測は主に 2002 年 3 月〜10 月にかけて行われたものであり、主にルートネームサーバに対する分析のみとなっている。本計測とその分析は 2003 年度も引き続き行っていく予定である。

6.7 dnsprobe 計測結果 CDF グラフ

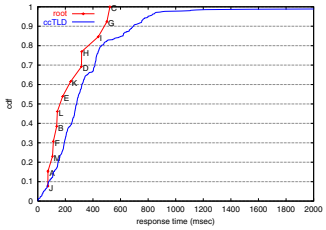
図 6.6 と図 6.7 に、27 拠点から行ったルートネームサーバならびに ccTLD サーバに対する dnsprobe の結果を CDF グラフとして示す。



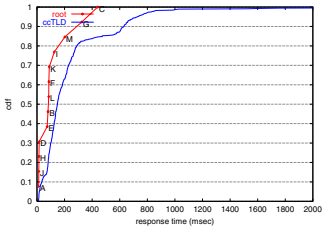
university, Oregon, USA



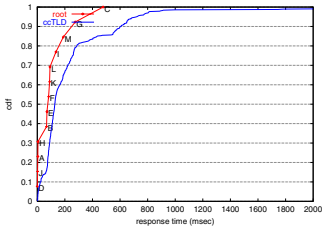
university, Palo Alto, USA



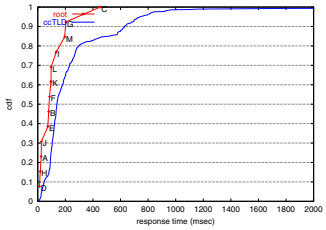
data center, Los Angeles, USA



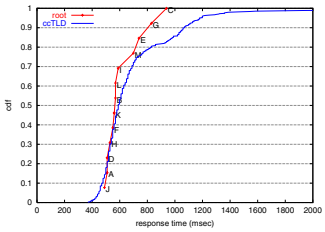
university, Pittsburgh, USA



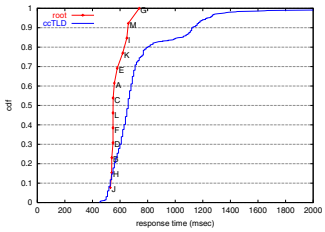
university, Maryland, USA



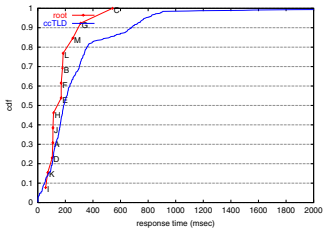
university, Cambridge, USA



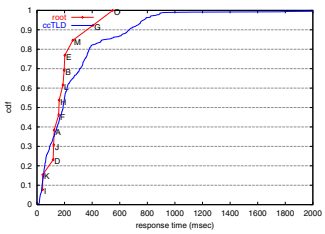
dialup, Ottawa, Canada



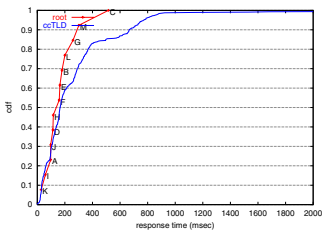
dialup, Cordoba, Mexico



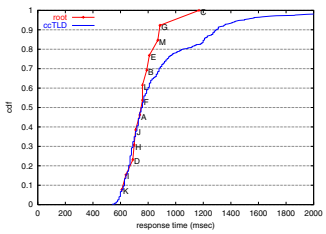
isp, London, UK



isp, Paris, France

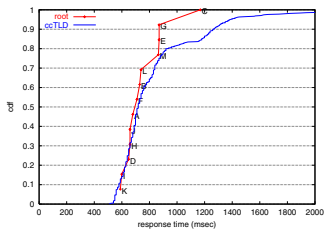


university, Zurich, Switzerland

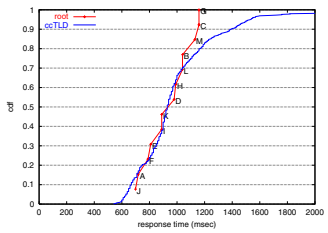


dialup, Parma, Italy

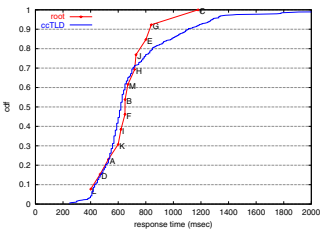
図 6.6.　median response time of root and ccTLD servers (1/2)



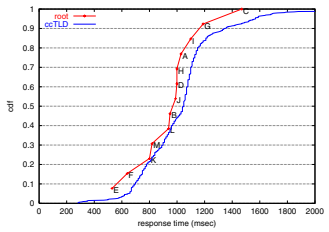
dialup, Torun, Poland



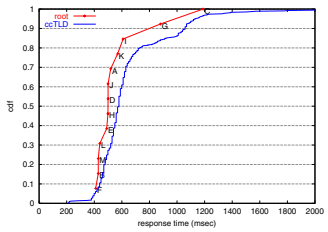
dialup, Ukraine



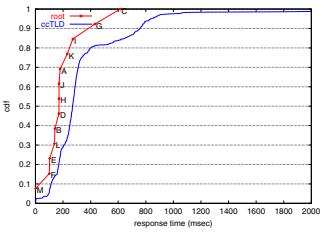
dialup, Shanghai, China



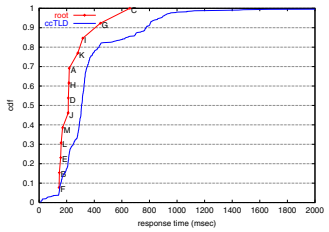
dialup, Beijing, China



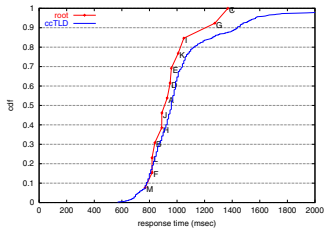
dialup, Seoul, Korea



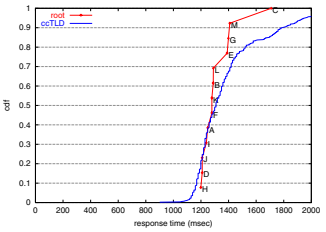
home, Tokyo, Japan



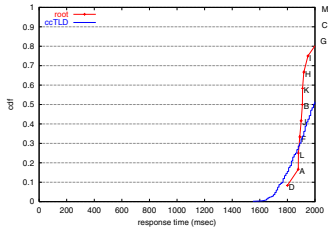
university, Hamilton, New Zealand



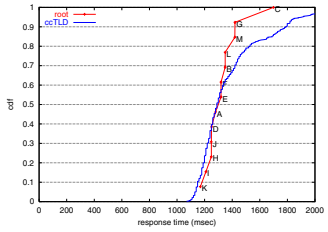
dialup, Canberra, Australia



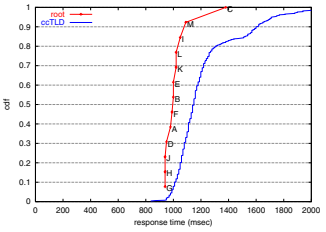
dialup, Cape Town, South Africa



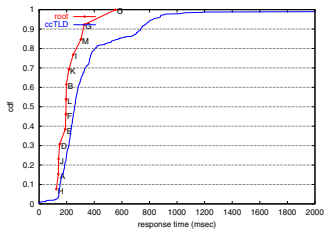
dialup, Eldoret, Kenya



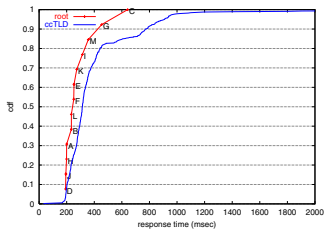
dialup, Algiers, Algeria



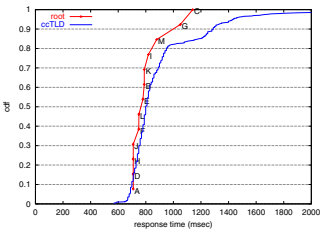
dialup, Salvador, Brazil



data center, Sao Paulo, Brazil



home, Buenos Aires, Argentina



dialup, Talca, Chile

図 6.7. median response time of root and ccTLD servers (2/2)