インターネットのトラフィックを測定及び解析するためのツールの設計及び開発

串田 高幸
日本アイ・ビー・エム株式会社 東京基礎研究所

インターネットにおいてトラフィックを収集し解析することは、ネットワークの管理において重要なことである。本研究の目的は、インターネットにおいて、どのようにネットワークのトラッフィクの測定を行ない、またどのようにしてネットワーク全体の状況を調べればよいかについて探ることである。本稿では、まずトラフィック収集についての定義と方式についてまとめている。次にインターネット上に流れているトラフィックを収集し解析するためのシステムについて、その設計方針と開発方式について述べる。最後にこのシステムの経過について述べる。このようなネットワークトラフィック測定のためのツールは、ネットワークの運用上、今後、必要不可欠なものである。

1　はじめに

最近、インターネットは、情報基盤として広く使われるようになり、その重要性が増してきている。インターネットの上で新しいアプリケーションが次々と開発され、従来から使用されているアプリケーションとともに広く利用されてきている。このように情報基盤が一般に広く使われてくるとネットワークを運用してゆく上においてネットワーク上を流れているトラフィックを効率よく収集し、そこで収集したデータを系統的に解析することが重要になってくる。このようして行なわれたトラフィックの収集と解析は、現在のネットワークに対する改善個所を発見することや、あるいは次期のネットワークへの設計に対しての基礎データとしても不可欠なものとなる。

一般にネットワーク上を流れるデータを収集し解析する方式について、次の2つに大きくわけて定義する。

- ネットワークモニター (マクロの視点)
  長期にわたり連続的に日常のトラフィックデータを収集する。統計情報としてのネットワークにおけるデータ収集である。最終的には、それぞれの項目を加算したデータを使用してトラフィックの解析を行なう。

- ネットワークアナライザー (ミクロの視点)
  プロトコルの解析、イベントトリガーによる測定及び問題点の発見など、一時的にかつ局所的に使用される測定方法である。また新しいプロトコルの開発のテスト時のテスターとしても使用される。アナライザーは、特定のデータを集めて、そのデータを解析的に表現して、その問題を解決するための目的に使われる。

本研究では、ネットワークトラフィックの解析を行なうためのネットワークモニター機能について行なった。

ネットワークにおいては、人が実際に目で見てトラフィックを確かめることが不可能である。そのため実際にデータがネットワーク上に、どの程度流れていて、またそのトラフィックによってどの程度混雑しているのか全くわらない。またどのような種類のアプリケーションのデータが流れているかもまったく見当がつかない。そのため基幹となる個所にネットワークのトラフィックを収集するための装置を入れて、データを収集し、そのデータを解析する必要がある。

さらにネットワークでは、バックボーンと呼ばる基幹の個所において常にネットワーク中に大量のデータが流れている。そこでネットワーク全体として、どのようなデータがながれているかを調べて、ネットワーク全体の状況を知る必要がある。そのためにバックボーンの各サイトにトラフィックを収集する装置を配置し、データの流れと解析を行なうことが必要となってくる。

ここでは、ネットワークのトラフィックを収集し解析するの目的を以下のようにまとめた。

- 現在のネットワークのトラフィックの状況を調べる

Design and develop tools for gathering and analyzing the Internet traffic
Takayuki Kushida(kushida@trl.ibm.co.jp)
IBM Research, Tokyo Research Laboratory

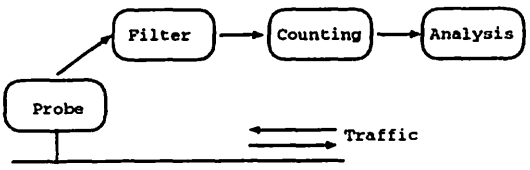


図 1: トラフィック収集解析のモデル

- 長期間収集することでネットワークの連続的な変化を調べる
- 将来のネットワークの新設や変更に対する基礎データを集める
- 長期間の測定によって経験的な平常値を求め、トラブル時の異常値を検出する

本研究では、まず一般的なネットワークの測定について、その概念をトラフィック収集と解析の観点からまとめる。次にインターネットにおけるネットワーク測定に対して、今までどのような仕事があったか述べる。そしてインターネットでのトラフィック収集をどのように行えばよいかの方法について述べる。それをもとにしてトラフィックを収集するシステムの設計開発に関して述べてゆく。また設計にあたって必要となった各データについても同時に述べる。最後にいままでのまとめと今後の方針について述べる。

2 ネットワークの測定

ネットワークのトラフィック収集と解析とは、ネットワーク上を流れているデータをすべて収集し、あらかじめ定義しておいた規則により必要な項目だけを抽出して分類することである。そして収集したデータを一定の期間だけ、それぞれの項目別に加算していく。その収集を行なうために、ネットワーク上に流れているデータを一旦、すべて捕獲し、その中から必要なデータだけを抽出する機能を持つ必要がある。またデータの抽出に際しては、適用する規則を用意しておく必要もある。そして、その規則に合致したデータだけを取り出して記録する。このように事前に不必要なデータを排除しておくことによって、目的のデータだけを記録することができる。その結果、保管しておくデータ量を少なくすることができる。一般にネットワークトラフィックの収集と解析は、図1のようなモデルであると規定できる。モデルにおけるそれぞれの機能は、以下の項目にまとめられる。

1. 全データの収集 (Probe)
   測定すべきネットワークに流れているすべてのデータを収集する。また収集するデータに取りこぼしがないように収集する装置では、パフォーマンス等を十分に考慮しなければならない。
2. 規則による取り出し (Filter)
   規則にもとづいてデータの取り出しを行なう。ネットワーク上を流れているデータは、その量が多い。そのため、すべてのデータを収集する必要がない場合には、フィルターを入れて必要なデータだけを選択する規則を入れる。選択を行なうことによって、保存するデータ量を収集したデータ量に比べ少なくすることができる。また規則は、各項目に対する And/Or の条件の組み合わせによって作られている。
3. 時間あるいは規則による加算結果 (Counting)
   データ収集に必要な項目をあらかじめ決めておき、その項目だけを選択的に一定の時間で順次加えてゆく。その加算された和が基礎データとなる。
4. 加算結果を解析 (Analysis)
   加算した結果をパラメータとしてグラフや表にして表す。この結果を調べることによりネットワークのトラフィックの状況がわかる。

トラフィック収集解析方式

一般的にトラフィックを収集し解析する方式は、時間的な観点から大きく次の2つに分かれる。

- 実時間で解析する方式
  実時間で解析する方式では、データを収集するときに、その収集されるデータだけに対してあらかじめ規則を設けておき、その規則に合致したデータだけを捕獲し、そのデータのみを加算し算出を行なう。この方式では、蓄積すべきデータが、加算後の結果だけのデータだけになる。そのため保存すべきデータ量が、少量ですむのでデータの取扱いが容易である。結果として必要なデータが、あらかじめわかっている場合には、データが順次、解析されていくので、すぐにデータを入手することができる。特にこの方式では、実時間での解析を行なって結果が即時に欲しい場合や、常に同じ項目のデータだけを必要するような固定的な解析を行なう場合に都合のよい方式である。しかし、例えば、データの積算を日変化から曜日変化に変更して解析したい場合や、時間とアプリケーションの関係を今までと別な方法で解析したい場合には、実時間で解析する方法ではできない。この方法では、今までに行なっていなかった方法で解析

したい場合には、解析前のもとのデータを記録していないので、過去に遡ってデータを改めて別な観点から解析することができない。

- バッチで解析する方式
  ネットワーク上に流れているデータをすべて収集し、そのデータを順次、大容量装置に蓄積しておく。そして、その蓄積されたデータを後で取り出して解析する。これがバッチで解析する方式である。この方式であれば、生のデータを全て記録しておけるので、後になって解析方法を変更して何度でも解析を行なうことできる。しかし、すべてのデータを記録し保管しなければならないため、記録すべきデータの絶対量が多くなってしまう。そのためシステムを構築するときに大容量の蓄積機能をもった磁気テープ装置が必要となる。この方式は、解析を何度も自由に行なうことができる点では、優れているが、大量のデータを記録する必要がある場合においては、実時間で解析する方式に比べて不利となる。

トラフィック収集解析のツール

今までに開発され、よく使われているトラフィックの収集や解析に使用されるツールとして以下のようなものがある。

- tcpdump
  tcpdump は、ネットワークのトラフィックを実時間で解析し、それをダンプするプログラムである。ネットワークに流れているデータを BPF(BSD Packet Filter)[2] から読み出し、MAC 層の解析を行ない、その後、TCP/IP のトランスポートまで含めたプロトコルの解析を行なう。さらに TCP/IP の場合には、トランスポート層の TCP や UDP ばかりではなく、それ以外の OSPF や IGMP などのプロトコルも解析可能である。また TCP/IP 以外の他のプロトコルに関しても解析を行なうことができる。

- NNStat
  NNStat は、事前に規則を記述し設定しておき、それにもとづいて解析をするネットワークの統計収集パッケージである。NNStat では、それぞれのネットワークのセグメントでトラフィックを収集するソフトウェアと収集されたソフトウェアを一カ所でまとめて集計するソフトウェアからできている。

- SNMP によるトラフィック収集 [3]
  インターネットにおけるネットワーク管理では、SNMP を使ってトラフィックデータの収集を行なうことができる。この方法では、MIB(Management Information Base) のなかにパケットのデータ量をカウントしている情報があり、その MIB の情報を SNMP を経由してアクセスする。この方法でアクセスを行ないトラフィック収集する。実際に MIB は、Interface Group にある ifInOctets や ifOutOctets であり、この値を調べることで各インターフェイスごとの入出力データの値がわかることになる。このデータをある一定時間ごとに収集することでインターフェイスごとのデータ量を算出することができる。しかし、アプリケーションデータの比率とか IP アドレスごとの集計とか、さらに解析を必要とする詳細なデータの収集には、実際にプロトコルデータの内容をチェックする必要があるために、この方式だけでは測定をすることができない。

インターネットにおけるトラフィック収集

インターネットにおけるトラフィック収集に関する仕事としては、NSFNET における T1 バックボーンのトラフィック収集がよく知られている [1]。この仕事は、NSFNET T1(1.5Mbps) のバックボーンのサイトにルータとして設置されたワークステーション RT/PC のうち 1 台を専用のトラフィック収集装置として使用し、そこでトラフィック収集を行なった。このバックボーンの各サイトに分かれている結果を一か所に集め、バックボーン全体のトラフィックとして集計した結果が、NSFNET の統計情報として公開されている。またトラフィック収集と平行して SNMP(Simple Network Management Protocol) による各インターフェイスのパケット数とバックボーンの間で ICMP による RTT(Round Trip Time) の測定も同時に行なわれていた。

NSFNET の場合、専用に開発されたハードウェアによって NSF バックボーンのすべてのトラフィックの収集を行なうことができた。専用のハードウェアであれば、設計時にトラフィックを収集するための機能を組み込み、それを使ってトラフィックを収集することができている。

ネットワークでのトラフィック収集の限界

インターネットのトラッフィク収集は、各バックボーンの Ethernet や FDDI といった LAN に対して行なう。しかしルータにおいて複数の回線が収容されている場

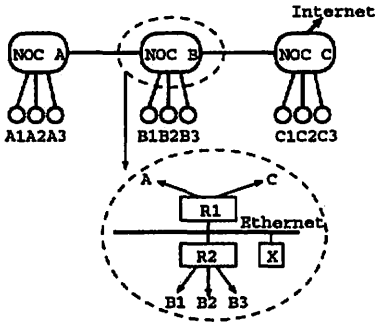


図 2: バックボーンネットワーク

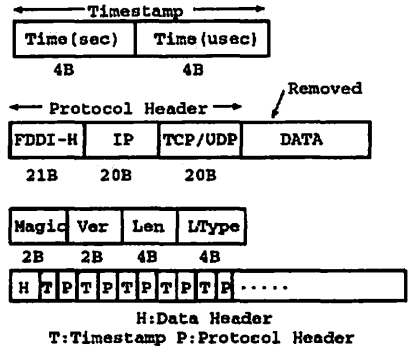


図 3: データの構造

合には、ルータの中でデータの交換が行なわれてしまうため、測定している LAN 側にトラフィックが流れなてこない場合がある。また各地域にネットワークを集線する NOC(Network Operation) がある場合には、NOC だけでの局所的なトラフィックも存在し、そのローカルの LAN それぞれに収集装置がないと、測定することができない。

図 2は、典型的なバックボーンネットワークの構成である。この図のバックボーンでは、NOC が A, B, C と 3 箇所に設置されている。NOC にそれぞれ参加組織が、$\{A1, A2, A3\}, \{B1, B2, B3\}, \{C1, C2, C3\}$ というように接続されている。ここで NOC Bの Ethernet に X というトラフィック測定装置を設置することにする。この X において測定できるのは、NOC Bの参加組織から A, Cと Internet へ向かうトラフィックと、そして A, C と Internet から NOC Bの参加組織へ向かうトラフィックだけである。NOC Bの参加組織間のトラフィックは、ルータ R2 の中で転送されるために X においては収集することができない。また A, Cの間のトラフィックも、ルータ R1 のなかで転送されてしまうために、このトラフィックも X で収集することができない。このようにルータの中の転送については、簡単な収集方法がないためネットワーク全体としてのトラフィックを収集することができないことになる。たとえ、A, B, Cのそれぞれの NOC に測定装置を設置したとしても NOC の組織の間におけるトラフィックは、ルータ内で転送されてしまうために計測することができない。

例えば、参加組織とネットワークプロバイダー側との境界が、シリアル線の上にあると仮定する。その境界においてネットワークプロバイダーに入ってくるトラフィックと、ネットワークプロバイダーから出てゆくトラフィックに対して正確に捕捉する必要がある。現在の LAN におけるモニターでの収集方式では、常にネットワークの一部のトラフィックを調べていることに限定されてしまうことになる。

3 システムの設計及び開発

本システムは、FDDI をバックボーンとしたネットワークに対してネットワークモニターを実時間とバッチの方式の両方で行なう機能を持つように設計した。またこのシステムによって、実際のネットワークで測定及び解析の行なうことで、その結果からさらにシステムを拡張することも計画している。

本システムの基本方針を以下に列記した。

- 測定ネットワークは、100Mbps の FDDI とする。
- 実際に FDDI をバックボーンとして使用している広域ネットワークへの適用を考える。
- FDDI ネットワーク上を流れているデータのうち TCP/IP プロトコルに関連している IP と ARP/RARP のデータだけを収集する。
- FDDI に流れている全データを時間とともに記録する。
- 大きな容量の記憶装置を使う。(7GB の磁気テープ)
- データ解析は、実時間での概要解析とバッチによる詳細解析の 2 つの機能を持たせることにした。
- システムは、普通に使われている汎用のハードウェアの上にソフトウェアを開発する。

各収集されたデータについている時間の生のデータは、秒が 4 バイトで、マイクロ秒が 4 バイトの合計 8 バ

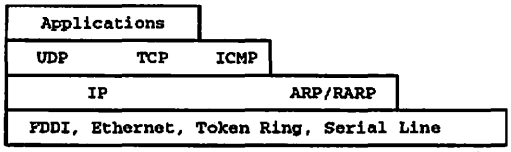


図 4: プロトコル

イトのデータとなっている。これを以下のような項目にわけて解析を行なう。

- そのままの時間データ (4Byte+4Byte)
- 年 (西暦 199n 年)
- 月 (1 月-12 月)
- 日 (1-28/30/31)
- 曜日 (日ー土)
- 時、分、秒 (0:00.00-23:59.00)

TCP/IP では、ネットワーク層に IP を使用し、トランスポート層に TCP と UDP を使用している。このプロトコルヘッダーのデータを次のように分割し、それを統計的な手法で解析を行なってゆく。

実際に収集されているプロトコルヘッダーは、次のような形式となっている。

- LLC/SNAP ヘッダー
  TYPE
- IP ヘッダー

```
VERS(4bit), HLEN(4bit)
SERVICE TYPE(8bit), TOTAL LEN(16bit)
IDENTIFICAION(16bit)
FLAGS(3bit), FRAGMENT OFFSET(13bit)
TIME TO LIVE(8bit), PROTOCOL(8bit)
HEADER CHECKSUM(16bit)
SOURCE IP ADDRESS(32bit)
DESTINATION IP ADDRESS(32bit)
```

- ARP/RARP ヘッダー

```
PROTOCOL TYPE
```

- UDP ヘッダー

```
SOURCE PORT(16bit), DEST PORT(16bit)
MESSAGE LENGTH(16bit), CHECKSUM(16bit)
```

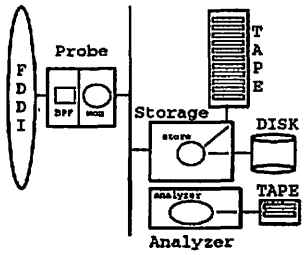


図 5: システムの構成

- TCP ヘッダー

```
SOURCE PORT(16bit), DEST PORT(16bit)
SEQUENCE NUMBER(32bit)
ACKNOWLEDGEMENT NUMBER(32bit)
HLEN(3bit), CODE BITS(6bit)
WINDOW(16bit), CHECKSUM(16bit)
URGENT POINTER(16bit)
```

- ICMP ヘッダー

```
TYPE FIELD(8bit)
```

システムの構成

このシステムは、図5のように構成上、大きくわけて (1) データ収集部及び実時間解析部、(2) データ記録部、(3) バッチデータ解析部の 3 つから成り立つように設計された。以下に、それぞれの機能について、それぞれ概要を説明する。

(1) データ収集部及び実時間解析部

データの収集部は、専用のワークステーションと収集用のソフトウェアからできている。ここで使用される OS には、BPF が装備されている。この BPF のフィルターを使って、ネットワーク上を流れている全データのうちプロトコルヘッダーだけを収集するようになっている。

ネットワークからデータを収集する場合に、ワークステーションで収集するパフォーマンスが重要となってくる。BPF では、データを収集するために UNIX カーネルのネットワークインターフェイス部分に効率よくデータを捕獲するための'network tap' と呼ばれている機能が組み込まれている。この機能によって、ネットワーク上にあるデータを収集してゆく。本実験では、アプリケーションのデータの内容すべてを必要としないため、

BPFのフィルター機能によって、トランスポート層までのプロトコルヘッダーを収集するように設定している。

高速ワークテステーションであれば、たとえFDDIであっても、すべてのデータを収集することが可能なように設計している。またここでは、収集すると同時に実時間で解析可能な項目については、その解析と表示を行なうようになっている。そのために実時間解析機能ととレポート機能を持っている。

(2) バッチタイプでのデータ記録部

データを記録するためのワークステーションと、その専用ソフトウェアからできている。ネットワークの上を流れている大量のデータを記録するために必要な磁気テープ装置と、その磁気テープをプログラムによって取り出し可能なロボットを使っている。この装置は、7GBの磁気テープを20本装着ことができるので、人の手を介さずに最大140GBのデータを連続的に記録することができる。

図5は、テープに記録されるデータの構造である。時間は、8バイトで表現される。プロトコルヘッダーは、FDDIのヘッダーが20バイト、IPヘッダーが20バイト、TCPヘッダーが20バイト、UDPヘッダーが8バイトとなる。またICMPヘッダーは、タイプにより異なるが、8バイトから20バイトである。ファイルとして記録するために、各ファイルの先頭にはファイルヘッダーが12バイトが必ず付加される。

(3) データ解析部

データ解析部は、磁気テープに記録された生のデータを読み出し、解析条件と規則を設定して、データの解析を行ない、その結果を出力しする機能を持っている。収集されたデータに対して統計的な手法を使用して解析を行なうが、そのデータが大量であり、複数の磁気テープにわたって記録されているために、データを連続的に読み出して解析してゆくことを行なわなければならない。

図6が、その解析の方式である。プロトコルの構成で解析を行なうときは、まずプロトコル階層をそのまま木構造にして、それぞれのノードに項目を入れるようなデータ構造をとる。また時間データは、プロトコルヘッダーのデータと並行した概念として存在しているので、内部では、時間データ、年、月、日、曜日、時、分、秒と、それぞれにわけて解析に使えるようにしている。

計時機能とTCP/IPプロトコルの測定項目

プロトコルを詳細にかつ解析的に処理する必要がある場合には、マイクロ秒以下の単位でパケットの時間記

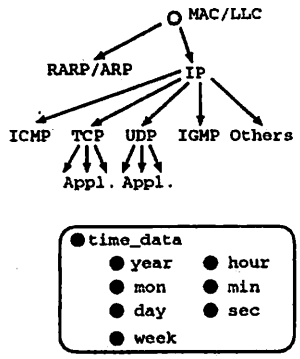


図6: 解析方式

録を行なわなければならない。しかし、トラフィックを測定する場合のタイムスタンプは、最終的には、長期間の統計的な結果となるので、時間に関して多少その粒度が荒くても問題ない。例えば、ミリ秒単位の時間が、収集されたパケットとともに記録されていれば時間として十分である。このような計時測定能力は、一般のワークステーションにおいても十分に計測可能な時間である。そのため本実験では、時計に関しては、ミリ秒単位までの正確さと粒度を保つようにしている。

測定するデータの見積りについて

システムを構築する際にあらかじめ収集されるデータ量を見積ることが必要である。この見積りによって、収集されるデータの総量が決まってくる。ここでは、システムとしてバックボーンを使用することを仮定する。例として、100MbpsのFDDIからの収集データの見積りを行なった。

FDDIは、物理速度が100Mbpsであるので換算すると

$$100(Mbps) = 12.5(MByte/s)$$

となる。1秒間に12.5MBであるので、

$$12.5(MByte/s) \times 3600(sec) = 45(GB/hour)$$

$$45(GB/hour) \times 24(hour) = 1080(GB/day)$$

$$Max.1080(GB/day)$$

1日最大、1080GBのトラフィックが流れることになる。実際には、LANの接続されたバックボーンではなく、広域の64Kbps-3Mbpsぐらいの専用線がFDDIを経由して相互接続された構成を仮定している。そのためそのときの平均のトラフィックを物理速度の最小1%から最

大3%ぐらいであると見積ると、

$$10.8(GB/day) \leq x \leq 32.4(GB/day)$$

となる。もしデータヘッダーが、全データ長さの$\frac{1}{10}$であれば、

$$1.08(GB/day) \leq x \leq 3.24(GB/day)$$

となる。結果として、1日、約1～3GBの容量を持つ外部記憶装置が、必要となってくる。

この見積りによるとFDDIのネットワークでは、1日に7GBのテープ1本を用意しておけば十分に間に合うという計算になる。もし20本の自動テープの切替え装置があれば、20日に1度テープの交換を行なえばよいことになる。この見積りは、一応の目安であり、パケット数が増加すれば、平均トラフィックが物理速度の3%以上となるので、そのことにより収集されるデータ量も当然ながら増加する。

トラフィックの収集に必要な項目

インターネットのトラフィックを収集する際には、最低限必要とする項目をまとめこれをもとにして解析を行なう必要がある。このような共通にトラフィックを集める項目を定義し、その項目が各機器に共通に実装されれば、特別なハードウェアを使うことなく統計データを作るための基礎データを測定することができる。

インターネットのトラフィック収集について最低限必要な項目は、プロトコルヘッダーのうち次の項目である。

- ARP/RARP
  OPERATION(ARP/RARP)
- IP
  VER, TOTAL LEN, TOS, IDEN, TTL, PROTOCOL, SRC/DST addr
- UDP
  SRC/DST port, LENGTH
- TCP
  SRC/DST port, HLEN, URGENT POINTER
- ICMP
  TYPE

また時間は、秒を4byte、マイクロ秒4byteの合計8byteで記録すればよい。

またトラフィックの集計では、上記の項目に加えて、

- 各インターフェイスごとの入出力バイトの記録
- ICMP ECHOのRTTの実測値

も同時に計測することが望ましいと考えられる。

実験環境

本システムは、基本となる部分をC言語で、各種の附属ツールや周辺部分をperl言語により開発した。

また機器は、データ収集装置にRS/6000 model 41T、データ記録装置にRS/6000 model C10と7GB磁気テープ20連装の装置、データ解析装置は、RS/6000 model 3ATを使用した。

このシステムは、省際ネットワーク[4][5]の東京NOCのバックボーンであるFDDIに設置されており、本格的にデータ収集が開始されている。現在、そのデータを収集しはじめ、解析をはじめているところである。

4 まとめ

本稿では、トラフィックを収集し、解析するためのツールについての研究結果についてまとめた。

今後は、さらにこの結果をふまえて、以下のことを行なっていくことを計画している。

- 収集されたデータを解析しその検討を行なうこと
  データをいくつかの方法で解析し、その結果を議論検討する。
- インターネットにおけるトラフィック収集の標準化作業
  インターネットにおけるトラフィック収集について、さらに標準化のような可能性について検討をする

最後に、本研究は、科学技術庁の「省際研究情報ネットワーク」の研究の一貫として行なわれた。また本研究を行なうにあたり東京基礎研究所の山内担当には、大変お世話になった。また省際ネットワークを運用しているNTTソフトウェア研究所の方々及びNTT大手町NOCの方々には、FDDIバックボーンでの測定を行なうにあたり、測定機器導入には、大変お世話になった。ここにお礼を申し上げる。

参考文献

[1] K. C. Claffy, H. Braun, and G. C. Polyzos. *Tracking Long-Term Growth of the NSFNET.* CACM Vol.37 No.8 1994.

[2] S. McCanne and V. Jacobson. *The BSD Packet Filter: A New Architecture for User-level Packet Capture.* 1993 Winter USENIX conference.

[3] D. E. Comer. *Internetworking with TCP/IP (third edition) Volume I Principles, Protocols, and Architecture*, 1995 Prentice-Hall.

[4] 福田、鈴木、三上.「省際研究情報ネットワークの接続形態」. 第51回情報処理学会全国大会 2E-8, Sep. 1995.

[5] 鈴木、福田、三上.「省際研究情報ネットワークの構成について」. 第51回情報処理学会全国大会 2E-9, Sep. 1995.