マルチメディア通信と分散処理 75－3
（1996. 2. 29）

インターネットにおけるトラッフィク収集と解析

串田高幸* 佐藤 卓由 山内長承

日本アイ・ビー・エム株式会社 東京基礎研究所
日本アイ・ビー・エム株式会社 コンテンツ事業開発部

インターネットにおいてトラフィックデータを収集解析することは、ネットワーク管理においてネットワークの状態を正確に把握するために重要な項目の一つである。本稿では、インターネットにおけるトラフィックの収集解析を行なうためのシステムについての概要を述べる。またトラフィック収集方式について述べ、収集したデータの解析方式について述べる。この方式を利用して実際にインターネットで収集したトラフィックデータを示す。最後にそのデータの結果からネットワークの状態について考察を行なう。

The Traffic Analysis on the Internet

Takayuki Kushida(kushida@trl.ibm.co.jp)
Takuyoshi Satoh
Nagatsugu Yamanouchi

IBM Research, Tokyo Research Laboratory
IBM Japan Ltd., Contents Solution Development

It's important for the network management to monitor the network traffic on the Internet. Because we have to know the network status to detect the network troubles and to expand the network facilities. This paper describes the overview of the system to gather and analyze the network traffic on the Internet. We describe a general concept of the sampling and analysis method for the traffic. As a example of the real traffic, we show several graphs of the traffic on the Internet backbone, and then describe the current network status by analyzing the traffic. This research is a part of the research project in the IM-NET(Inter-Ministry NETwork) by the Science and Technolgy Agency.

はじめに

インターネットにおいてトラフィックデータを収集解析することは、ネットワーク管理においてネットワークの状態を正確に把握するために重要な項目の一つである。インターネットでは、プロトコルとしてTCP/IPを使用している。そのためデータグラムのヘッダーだけを収集することによってインターネットのトラフィックを測定することができる。

インターネットのバックボーンでは、大量のトラフィックが常時流れている。そのトラフィックデータのすべてを収集解析するためには、測定装置が複雑になり、またそのため高価になってしまう。さらにトラフィックデータの解析にも時間がかかる。もしネットワークのトラフィック収集に対して、(1) 必要最低限度のパラメータを規定することができ、さらに (2) そのパラメータに基づいて部分的にトラフィック収集を行なうことができ、結果として (3) ネットワークの状況がわかれば、その測定は、ネットワーク管理に対して有効な方式となる。

本研究では、インターネットのトラフィックの収集解析における一般的な測定方式を確立することを目的としている。言い替えると、どのようなデータをどのくらいの時間収集して、またどのような種類のデータを出せば、ネットワークのトラフィック解析として有効であるか、またネットワーク管理において役立つのかを探ることを行なう。

本稿では、まずインターネットにおけるトラフィックの収集解析を行なうためのシステムについて概要を述べ、トラフィック収集方式について述べる。次にデータを使った解析方式について述べる。また、この方式を利用して実際にインターネットで収集したトラフィックデータを示す。最後にそのデータの結果からネットワークの状態についての考察を行なってゆく。

トラフィックの収集、蓄積及び解析装置

本研究では、インターネットのトラフィックを収集解析するためにネットワークのモニターとして汎用ワークステーションによるネットワーク収集解析システムを新規に開発した。このシステムでは、汎用ワークステーションによりネットワーク上に流れている全データを収集するためBPF(BSD Packet Filter)機能を利用している。また収集するデータは、データグラムのヘッダー情報だけにしている。その収集したデータを時間とともにディスクに順次、書き込んでゆく。これが、モニタープログラムである。

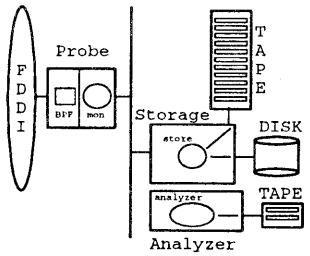

図 1: トラフィック収集解析システム

ディスクに書き込まれたデータは、ある一定時間 (例えば、1時間) ごとに一つの独立したファイルとして生成されてゆく。そして、このデータファイルは、さらに一定時間ごとに順次、磁気テープに書き込まれてゆく。書き込まれたファイルの入った磁気テープは、データが一杯になるか、あるいは一日一度、新しいテープを変更するようにしている。こうして最大連続20日間、ネットワークデータを無人で収集できるようになっている。本システムは、実際に省際ネットワークのオペレーションセンターへ持ち込まれ、現在収集が行なわれている。システムの設計及び内容の詳細については、[1][2] が参照できる。図1が、このシステムの概略図である。このシステムでは、FDDIバックボーンに流れている全トラフィックを一旦、収集し、Ethernetを経由してデータを連装型の自動切換の7GB磁気テープ装置に記録してゆくシステムになっている。

このシステムでは、全トラフィックのデータを収集することができるため、すべてのIPデータグラムのヘッダーに関して、解析を行なう

ことが可能である。そのため、今までのシステムのように事前に定義したデータを収集する場合とは異なり、後から何度でも別な解析方法を使ってデータを解析することができる。

トラフィックの解析プログラムは、磁気テープに記録されたデータをファイルとしてハードディスクに読み出す。そして磁気テープからハードディスクに記録されたデータをデータグラムのヘッダーに関するフィルターによってトラフィックのなかで必要な情報だけの加算を行なってゆく。この解析プログラムによって、全トラフィックの積算を行なうことができる。またフィルターの構成を変更することによって、トラフィックの加算する条件が変更される。

ネットワークとトラフィック

ネットワークの状況を把握するためにネットワークのトラフィックを収集解析する仕事は、以前から行なわれている。インターネットでは、NSFNET バックボーンにおけるトラフィックの収集と解析が知られている [3]。NSFNET では、専用の特別なルータを使用して、全米に散らばっているバックボーンノードからのトラフィックを収集し、そのデータを解析した。

一般にネットワークにおいてのトラフィックを調べることは、次のようなことである。

1. 現在のネットワークのトラフィック量、
2. 過去からの変化により将来の予測、
3. 日常のデータの解析から異常なデータの検出、
4. 短期間の変化から長期間の変化を検知、
5. 個別課金のために必要な情報の入手。

トラフィックを収集することによって、ある時間のトラフィックの量がわかる。このトラフィックの量によってネットワークの使用率がわかる。また各組織ごとの利用率も同じ方法を使って知ることができる。もし過去のデータの変化が関数化できれば、その現在の結果から将来の変動の予測も可能となる。例えば、過去からのトラフィックが単調増加であると仮定できれば、その関数を使って将来の量が予測できる。また日常のデータを蓄積しておくことによって、日常に起こらない特別の変化を検知することができる。また短期間の変化を詳細に解析しておき、それに基づいて長期間の変化を推測することができる。トラフィックには、IP アドレスが入っているため、IP アドレスごとのトラフィックによる個別課金が可能になる。IP アドレスごとの個別課金ができれば、ネットワークごとの使用頻度が算出できるので、この結果によって組織ごとの課金も可能になる。

また、ネットワークにおいて収集すべきトラフィックは、

1. 全体のトラフィック
2. プロトコルごとのトラフィック
3. アプリケーションごとのトラフィック
4. 組織ごとのトラフィック

というクラスに分かれる。これらのクラスは、それぞれパケット数とバイト数というの 2 つの単位を持つことができる。

これに加えて、必ず収集された時間が入っているため時間に対するクラスもある。一般に時間に対する変動は、通常のトラフィックであれば周期的になっている。そのため、時間のクラスは、さらに周期的な変動によって以下の項目に分けることができる。

- 昼夜のトラフィック変動
- 曜日 (平日と週末) による変動
- 月の間での変動
- 月ごと (季節) による変動

トラフィックの収集方式

広域ネットワークの FDDI バックボーンでは、流れているデータは、パケットヘッダーだ

けでも1日当たり数GB以上と膨大な量となる。そのため、その膨大なデータを解析し、その結果を出すためには、処理時間が非常にかかる。もし部分的な収集解析により正しくネットワーク全体の状態がわかるならば、全体のトラフィックデータを取得せずにすむのでファイルの記憶容量を少なくでき、さらに処理の時間も短くすることができる。そのために部分的な収集方法の規定することは、重要である。

統計においては、母集団を推定するために標本抽出を行なう場合、無作為抽出という方法が一般的に行なわている。この方式を用いれば、抽出されたデータが母集団での割合と同じなる。無作為によって抽出されるデータは、母集団のどの標本も同じ確率により選ばれる可能性があるからである。

一般にトラフィックデータの部分的な収集方法は、時間に対して周期的な部分を選択する方法と時間に対してランダムに選択する方法とそして収集データ量に対して常に一定量を選択する方法を適用することができる。さらに特定のトラフィックのイベントによって、トラフィックを収集することも可能である。以下に部分的な収集方式についてまとめる。

- 時間に対して周期的な収集
- 時間に対してランダムな収集
- 常に一定データ量の収集
- イベントによる収集

ネットワークトラフィックのデータ

実際のトラフィックとして1995年10月のうちの20日間をグラフとして示している。図2、図3、図4は、それぞれ全データ量、TCPのデータ量、UDPのデータ量を示している。図3は、24時間の周期的な変化と週の変化がある。図2において、最初の2日ぐらいの間、データ量が多いのは、図4にも示されているようにUDPのデータ量が多いためである。またUDPは、途中に台形となっている時間がある。このようにUDPの変化は時間や曜日のような時間軸に対する変化がなく、どちらかというとアプリケーションの連続動作に依存している。図5、図6、図7は、それぞれ全データのパケット数、TCPのパケット数、UDPのパケット数を示している。この場合もデータ量と同様にTCPには、時間による周期的な変化を持っているが、UDPには、時間的な変化がない。

図8は、全体量とWWWのポートであるTCP80番のデータ量の比率をグラフに示したものである。これをみてわかるようにWWWは、アプリケーションのなかでも40%から50%と比率が高い。図8に最初と途中に2箇所ほど、比率が下がるところがあるが、これは、UDPのトラフィックが上がったため相対的にWWWの比率が下がったためでWWWのデータの絶対量が減少したためではない。

図9は、ICMPのパケット数を示している。ICMPは、ネットワークの制御やネットワークの状況を知らせるために使われている。図9では、何箇所か局所的に急激なパケットの増加が見られる。この増加は、なんらかのトラブルによってICMPが平常時よりも多く送出されたためと考えられる。そのため、ICMPのパケットのタイプと時間変化をさらに詳しく解析することによって、ネットワークでの異常を検知手段となる。

図10は、(1) 全データ量の10分ごとに集めた1時間ごとの平均値、(2) 全データ量の各時間の最初の10分間ごとの値、(3) 全データ量の各時間のランダムな10分間ごとの値、に関して (2)-(1) と (3)-(1) をそれぞれグラフに示したものである。それぞれの標準偏差 s は、$s = 2410297, s = 604053$ となった。この標準偏差から考えると (2) の収集方式に比べ、(3) の収集方式がより平均に近い値となっている。

結論と今後

本稿では、トラフィックの解析において以下のような結論を導きだせる。

- 全データ量とTCPのデータ量は、24時間周期及び7日周期の変動がある。

- UDP のデータ量には、周期的な変動がない。
- WWW のデータ量は、全データ量と比べると 35%-50%ぐらいになっている。
- ICMP のパケットの調査は、ネットワークの異常の検出に利用することができる。
- 部分的な収集方式に関しては、周期的な方式に比べ一定の計測区間のランダムな選択方式の方が、より有効となる。

本研究を行なうにあたって省際ネットワークの運用管理を行なっている NTT 大手町省際ネットワーク NOC の方々及び NTT ソフトウェア研究所の方々には、FDDI バックボーンでのトラフィック測定に関して色々とお世話になった。ここにお礼を申し上げる。また本研究は、科学技術庁の「省際研究情報ネットワークの調査研究」として行なわた。

参考文献

1. 串田高幸,「ネットワーク解析ツールの設計」. 平成 7 年情報処理学会第 51 回全国大会. 1995 年 9 月.
2. 串田高幸,「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」. 情報処理学会マルチメディアと分散処理ワークショップ. 1995 年 10 月.
3. K. C. Claffy, H. Braun, and G. C. Polyzos. *Tracking Long-Term Growth of the NSFNET.* CACM Vol.37 No.8 1994.
4. 初等統計原書第 4 版, P.G. ホエール著, 共立出版,1981.
5. K. C. Claffy, G. C. Polyzos, H. Braun. *Application of Sampling Methodologies to Network Traffic Characterization.* In Proc. ACM SIGCOMM'93 San Francisco, CA, Oct 1993.

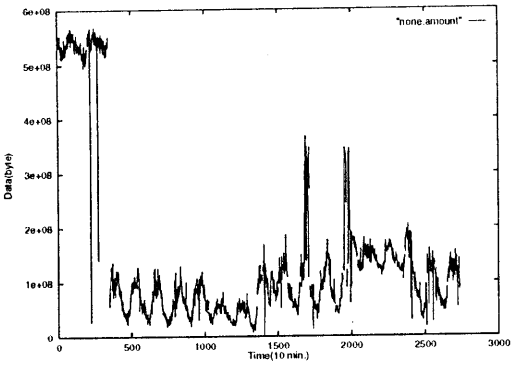

図 2: 全データのデータ量

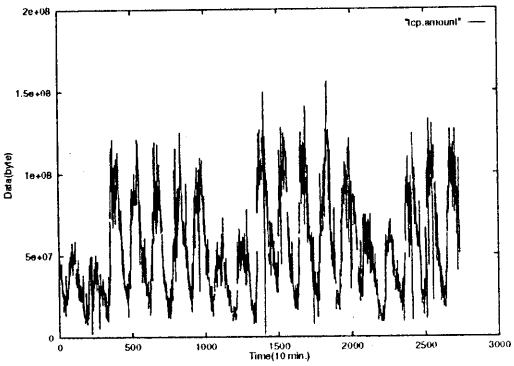

図 3: TCP のデータ量

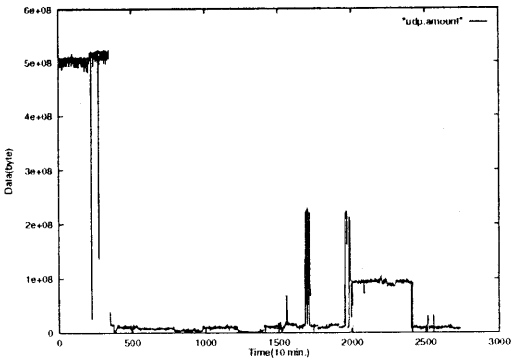

図 4: UDP のデータ量

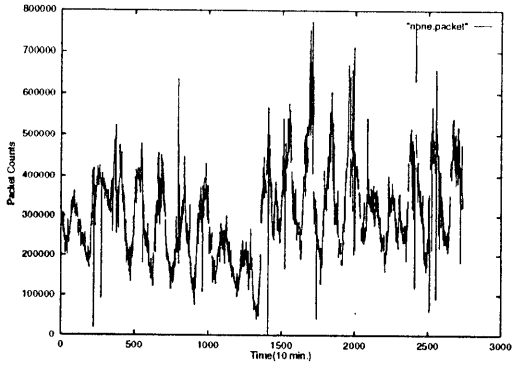

図 5: 全データのパケット数

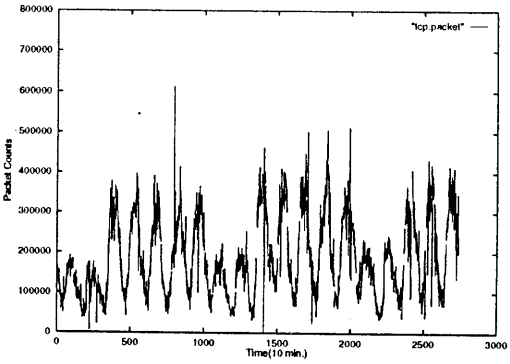

図 6: TCP のパケット数

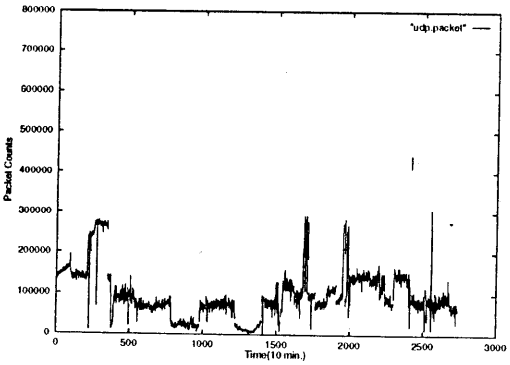

図 7: UDP のパケット数

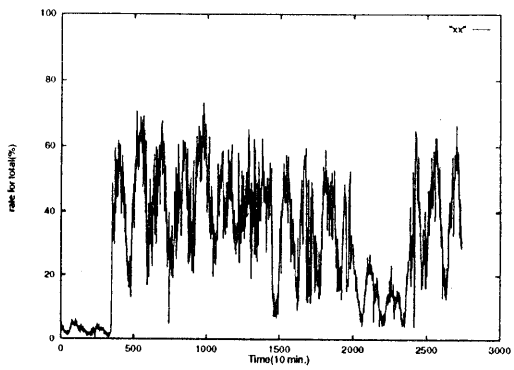

図 8: 全体量と TCP 80 番のデータ量の比

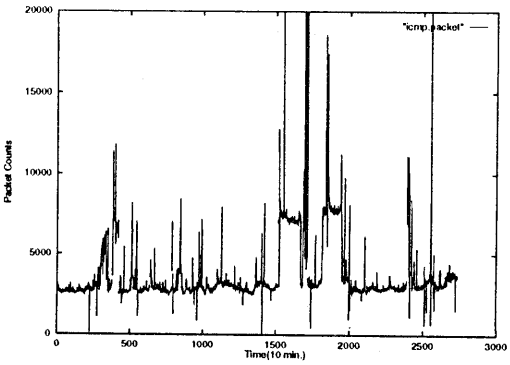

図 9: ICMP のパケット数

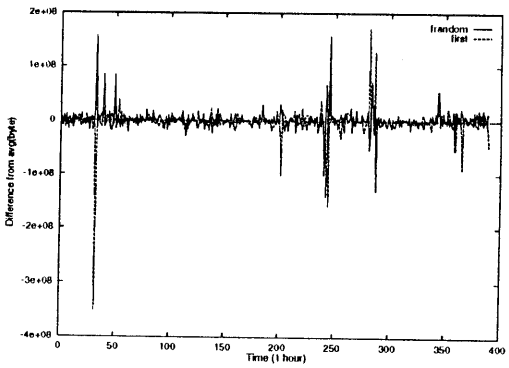

図 10: 収集方式と平均値との差