

東工大で受発信する HTTP トラフィック量の解析

02004950 東京工業大学情報理工学研究科
 01506790 東京工業大学情報理工学研究科
 東京工業大学総合情報処理センター
 東京工業大学総合情報処理センター

*小上 賢一 OGAMI Kenichi
 藤本 衡 FUJIMOTO Kou
 合田 憲人 AIDA Kento
 櫻井 成一郎 SAKURAI Seiichiro

1 はじめに

本研究では、東工大 LAN (Titanet: Tokyo Institute of Technology AcademicNetwork) と外部ネットワークの出入口 (対外接続リング) に流れるトラフィック量を測定し、人間の行動を考慮してモデル化することを目的とする。

2 Titanet データの測定

Titanet は、主に WIDE (Widely Integrated Distributed Environments) ネットワーク、SINET (学術情報ネットワーク) の 2 つの外部ネットワークに接続されている。これらはすべて FDDI である対外接続リングにつながれていて、Titanet と外部ネットワークとのデータのやりとりはすべてこのリング上を通過している (現在は Ethernet に変わっている)。本実験ではこのリングに測定機器を接続したため、東工大と外部の全てのデータのやりとりを把握したことになる。

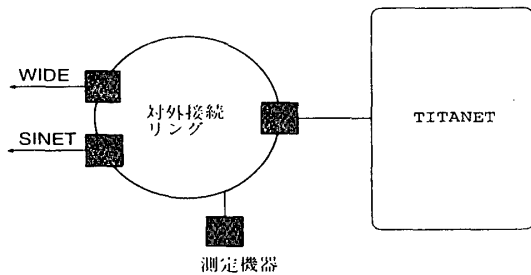


図 1: 対外接続リングと Titanet

本研究では、2 種類の測定データを用いて分析を行った。1 つは、1 分単位のトラフィック量を 10 ヶ月間にわたって測定したものである。もう 1 つはトラフィックをパケット単位である特定の 1 時間について測定したものである。この 2 種類のデータの詳細は次の通りである。

10ヶ月間における 1 分単位の測定

期間 1998 年 5 月 14 日から 1999 年 3 月 31 日

内容 1 分単位のデータ量 (バイト数) を測定
 ポート番号別、東工大出入別、TCP・UDP 別

パケット単位の測定

期間 1999 年 12 月 22 日 (水) 17:00 から 18:00

内容 パケット到着時刻 (0.001 秒単位)、ポート番号、FDDI ヘッダ、フラグ、MSS、シーケンス番号、データ量 (バイト数)。

このデータ測定はプライバシーを守ることが絶対条件であるので、発信元ホストと受信先ホストの識別子である IP アドレスを見ることは許されない。よって測定する際には IP アドレスを出力しないようにした。

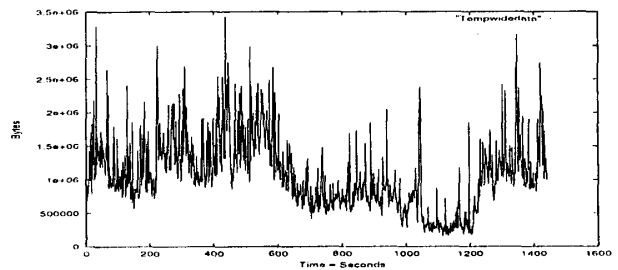


図 2: 1 日に東工大へ着信する WWW データ量

表 1: トラフィックにおける各アプリケーションの割合

アプリケーション	東工大からの 発信	東工大への 受信
FTP DATA	20.2 %	11.0 %
E-MAIL	3.6 %	3.6 %
DNS	9.1 %	4.8 %
WWW	45.8 %	66.7 %
NNTP	3.0 %	8.5 %
その他	18.3 %	5.4 %

3 測定データの解析

ここでは主としてパケット単位の測定データを用いて、パケット到着の挙動を表す確率過程モデルを提案するのが目的である。

モデルのアイデア トラフィックの到着モデルは、解析のしやすさからポアソン過程でしばしばモデル化されてきた。しかし、近年の通信トラフィックの発生過程はパケットレベルで見た場合、その複雑さからポアソン過程ではモデル化できないとされている [1]。

表1の1分単位のデータより、対外接続リングではHTTPトラフィック量は全体のトラフィック量の約半分を占め、最も重要なアプリケーションであることがわかる。本研究ではこのHTTPトラフィックに着目した。

高速道路を通過する車や一連の電話の呼など、人の動きに即したものはポアソン過程でよくモデル化される。その理由は、人やものの流れのランダム性を最もよく表した数学モデルだからである。HTTPにおいても、人間が関与する部分はポアソン過程であると考えられる。これはクライアントがあるWebページを見るという動作と考えるのが自然であろう、[3]にもWWWにおけるモデムへのアクセスの仕方がポアソンの挙動をすることが記されている。しかしこの動作に対応して複数のHTTPコネクションが束(HTTPセッション)となって発生する。これが解析を難しくしている原因である。

解析 この考えに基づいてデータ解析を行った。このHTTPセッションをデータから直接とらえられるとよいのであるが、プライバシー保護のために、IPアドレスを見ることは許されない。そこでクライアント側がポート番号を割り当てる特徴から、コネクションのまとめ方についてある仮定をおき、HTTPセッションを推定したうえで分析した。束の発生がポアソン分布

表 2: 1 時間におけるトラフィック量

総パケット数	9,677,360
HTTP パケット数	5,788,643
HTTP コネクションの個数	200,786
HTTP セッションの個数	47,911

に従うという仮説に対する χ^2 適合度検定を行なったところ、 χ^2 統計量が 35.17 未満である確率が 0.95 であるのに対し、実際の χ^2 の値は 31.638 となり棄却さ

れないことが確認された。このことから人があるページを見るという動作がポアソン過程でモデル化できることを示していると期待される。

次に測定期間における HTTP セッション中の HTTP コネクション数をヒストグラム化し、べき関数的であるパレート分布であることを示した。また、1つのコネクションによってデータ量およびペイロードがある(データを積んでいる)パケット数がどれくらい転送されるかを示し、同様にパレート分布で近似できることが確認できた。

4 まとめと今後の課題

本研究が示したところは、ユーザーがWWWのページを見るという行動の発生はポアソン過程に従い、その行動によって複数のコネクションが発生することである。そのコネクションの発生の仕方はパレート分布で近似できる。また1つのコネクションで送られるデータ量、ペイロードがあるパケット数も同様にパレート分布で近似できる。HTTPトラフィックが、これまで長期依存性を持つといわれていたのは、このパレート分布のようなべき関数的な分布が要因になっている。

しかしパケットの到着間隔などに関して、Time Out などのために分布の裾の部分が完全にパレート分布に従うとは考えにくい部分もある。それらの点も含め、今後はHTTPの挙動をもっと考慮に入れたモデル化などが課題となる。

なお本研究はデータ取得の際、東京工業大学総合情報処理センターネットワークシステム部門の皆様にとくさんの助言をいただいた。この場を借りて感謝したい。

参考文献

- [1] Toshihisa Ozawa, Fumiaki Machihara, Keisuke Ishibashi, "トラフィックの自己相似性とその周辺", 電子情報通信学会誌, 81, 5, pp.506-515 1998.
- [2] Vern Paxson and Sally Floyd, "Wide Area Traffic: The Failure of Poisson Modeling", IEEE/ACM Transactions on Networking, Vol.3, No.3, June 1995.
- [3] A.Feldmann, A.C. Gilbert, W.Willinger and T.G. Kurtz, "The Changing Nature of Network Traffic: Scaling Phenomena", ACM Computer Communication Review, vol. 28, pp. 5-29, Apr. 1998.