

インターネットトラフィックの分析とその応用



研究ノート

村田 正 幸*

Analysis of the Internet Traffic and Its Applications

Key Words : Traffic Analysis, Queueing Theory, The Internet, ATM, Capacity Dimensioning

1. はじめに

電話の発明に始まるいわゆる電気通信網は既に100年以上の歴史を持ち、近年、その中核技術としてATM技術が確立されつつある。一方、1960年代の米国におけるARPANETに始まるコンピュータネットワークは、現在、インターネットとして急速な成長を遂げているのは周知のとおりである。

これら2つの通信技術は別々の歴史を持ち、インターネットがコンピューター間データ通信、ATMが電話や動画像通信などのような実時間通信を指向してきたために、インターネットとATMはまったく異なる通信の仕組みを持つ¹⁾。しかし、最近、それら2つの通信技術の融合も進みつつあり、そのひとつの実現形態が、大阪大学学内通信網であるODINSである。すなわち、ODINSは、研究室や構内ではLAN(コンピュータネットワーク)、バックボーンネットワークはATM(電気通信網)を用いて構築さ

れている。また、家庭からインターネットにアクセスする場合には、ISDN(ATMの一世代前の通信技術)→インターネットサービスプロバイダ(インターネット技術)→公衆通信網(最近ではATM技術が使われていることも多い)、といった形態がとられ、ネットワーク文化が市民生活に浸透する原動力となっている。

もちろん課題も数多く残されており、その中で筆者は以下の問題に取り組んでいる。

1. ATMは電話網、ISDN網の延長線上で開発されてきたため、データ通信に弱い。ATMでデータ通信を効率よく実現するにはどうすればよいか？
2. 逆に、インターネットはデータ通信を目的として開発されてきたため、実時間通信に弱い。実時間通信をいかに実現するか？
3. インターネットで用いられるTCPは通信技術としては陳腐化しつつあり、光通信技術など高速通信技術への対応をどうするか？
4. さらに、情報通信の基盤技術として、ATMの次の通信技術はどうあるべきか？光通信技術が有望であるが、それをどのようなかたちで実現すべきか？
5. インターネットが急速に進展してきている中で、インターネット設計論をいかに確立するか？(後述)
6. 無線技術の発展によりラストホップは無線になりつつある。有線通信網との融合をどう実現すべきか？

以下では、5番目のインターネットにおけるネットワーク設計論の確立を目指した研究につ

*Masayuki MURATA
1959年7月28日生
1984年大阪大学大学院基礎工学研究科博士前期課程修了
現在、大阪大学大学院基礎工学研究科、情報数理系専攻、助教授、工学博士、情報通信工学
TEL 06-850-6586
FAX 06-850-6589
E-mail murata@ics.es.
osaka-u.ac.jp
URL : <http://asteroid.ics.es.osaka-u.ac.jp/~murata/>



いて述べる。

2. インターネットトラヒックの分析とその応用

従来の電話網では次のようなサイクルを経て、ネットワーク構築が行われてきた。

- (1) 基準品質の決定
- (2) トラヒック予測
- (3) 必要な設備量の決定
- (4) 投資コストの評価とサービス価格の決定
- (5) 品質測定により、設備量の見直しへ

上記のフィードバックループにおける諸理論/技術は確立されており、特に3番目のステップで重要となるのが、システムの「モデル化と性能評価」であり、その基礎理論となるのがトラヒック理論/待ち行列理論である²⁾。一方、インターネットは近年になって急成長を遂げたこともあり、上記のような成長サイクルが正常に機能せず、サービスプロバイダが提供する通信品質の劣悪さが問題になっている。競走市場であるがゆえに、まず最初に市場をにらんだ料金設定が行われ、ネットワークが混んでいる(ような気がする)ので回線容量を増やす、あるいは、通信品質に関する苦情をユーザから受けてはじめて設備投資を行う、といった場当たり的な解決策がとられているのが現状であり、それが適切に行われずにユーザの離反を招くといった悪循環に陥っている場合も多い。

問題の解決のためには、

- (1) インターネットのトラヒック特性の把握
- (2) それに基づいたシステム性能評価モデルの構築

(3) 待ち行列理論/シュミレーションの適用による必要資源(処理能力、回線容量)量の把握が必要となる。文献3)では、インターネット上のWWW(World Wide Web)⁴⁾トラヒックに焦点を絞り、大阪大学情報処理教育センターにおけるプロキシサーバのログ情報を分析し、その特性を明らかにした。以下にその概要を述べる。

2.1 ログ情報の収集

まず、2ヶ月間のログ情報から得られたメディア別の構成を表1に示す。これを見ると、画像がリクエスト回数で60.2%、伝送バイト数で

表1 WWWドキュメントのメディア別構成

メディア	リクエスト (回数)	伝送バイト (Mbyte)	平均サイズ (Kbyte)
テキスト	321,563	1,750	5.4
画像	729,127	6,692	9.2
音声	2,014	324	161.0
映像	838	719	858.1
C G I	73,545	426	5.8
圧縮ファイル	15,081	793	52.6
その他	69,591	1,286	18.5

55.8%を占め、ついでテキストが多い。一方、音声や映像はあわせて全体の0.3%にすぎないが、伝送バイト数ではテキストファイルの約3分の2の8.4%と大きな割合を占めている。また、平均サイズは5.4~858.1KbyteでWWWのドキュメントサイズが非常に広い範囲を取ることがわかる。

2.2 トラヒック特性

次に、トラヒックモデルを構築するために、対象とするトラヒック特性を確率分布関数によって近似する。具体的には、候補とする確率分布関数のパラメータをサンプルデータから推定してモデルを作成し、最適なモデルを判定するというものである。対象とするトラヒック特性は、ドキュメントサイズ、リクエスト間隔、ドキュメントごとのアクセス頻度である。これらは、ネットワーク設計を行う際に重要になる。候補とする確率分布関数は、正規分布、指数分布などよく知られた分布とともに、表1に示したドキュメントサイズのように広い範囲をとる分布を表現する必要がある。そこで、二重指数分布、対数正規分布、対数二重指数分布、パレード分布も候補とした。例えば、対数正規分布は、

$$F(x) = \int_0^x \frac{1}{\sqrt{2\pi}\sigma y} \exp\left[-\frac{(\log y - \zeta)^2}{2\sigma^2}\right] dy$$

であらわされるものである。

2.3 分析結果

詳細は省略するが、ドキュメントサイズに最も適合したのは対数正規分布であった。図1にログ情報から得られたドキュメントサイズの分布と対数正規分布、指数分布によるモデルをあ

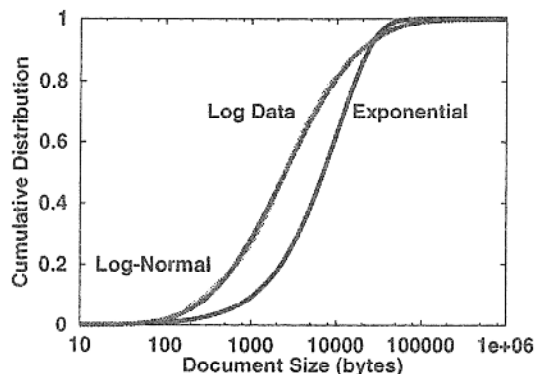


図1 ドキュメントサイズの分布

わせて示す。図1より、対数正規分布によるモデルが、ドキュメントサイズの分布と非常によく適合していることがわかる。一方、数学的な取り扱いが容易でネットワーク設計によく用いられてきた指数分布⁵⁾は、実際のデータとの差が大きい。

リクエストの発生間隔についても同様の統計的分析により、対数正規分布がもっとも適合すること、ただし、ネットワーク設計を行う際に重要となる最繁忙時間の分布は指数分布に従うことを確認した。

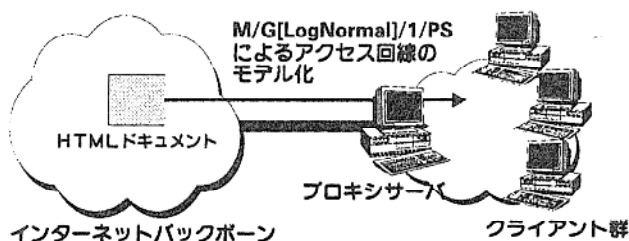


図2 アクセス回線モデル

2.4 待ち行列システムによるモデル化

次に、前章の分析結果に基づき、アクセス回線を待ち行列システムでモデル化する。対象とするアクセス回線モデルを図2に示す。本モデルでは、ユーザはLANを介してプロキシサーバに接続されており、プロキシサーバは中継線を通じてバックボーンネットワークに接続される。従って、ドキュメントは外部のサーバからアクセス回線を通してユーザへ伝送される。

すなわち、本モデルは、アクセス回線をサーバ、ドキュメントを客とする単一サーバ待ち行列システムとしてモデル化できる。ただし、実

際にはドキュメントはTCP/IP上のHTTPプロトコル⁴⁾によって伝送されるため、パケット単位に分割されて処理が行われる。そこで、処理規律は中継線のパケット伝送能力を転送中ドキュメント数で均等に分割されるPS(Processor Sharing)を用いる。待ち行列システムにおける到着間隔の分布は、最繁忙時間のリクエスト間隔に適合した指数分布とする。また、サービス時間分布は、ドキュメントサイズ分布に対応し、回線速度をCとすれば、次式で与えられる。

$$G(t) = \int_0^t \frac{1}{\sqrt{2\pi\sigma Ct}} \exp\left[-\frac{(\log Ct - \zeta)^2}{2\sigma^2}\right] dt$$

2.5 WWWトラフィックの遅延時間分布

本節では、アクセス回線上でのドキュメント転送に要する遅延時間分布を数値例によって示す。システムモデルの妥当性はコンピュータ上のシュミレーションによって確認する。一例として、アクセス回線を512Kbpsにした場合の遅延時間の確立密度を図3に示す。また、分布

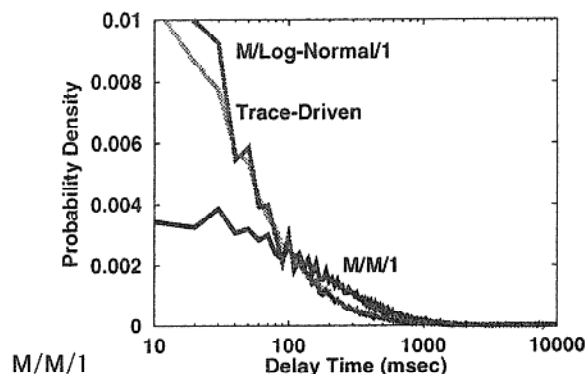


図3 遅延時間確率密度(中継線: 512Kbps)

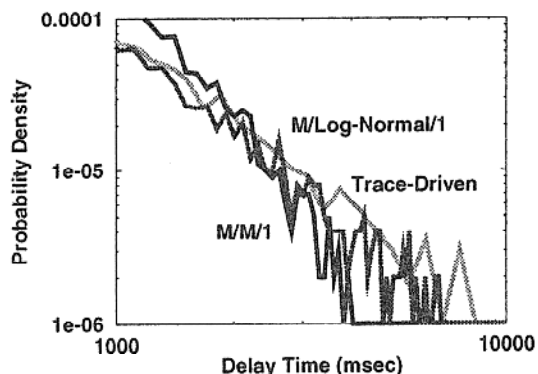


図4 遅延時間確率密度の裾野部分(中継線: 512Kbps)

の裾野部分を拡大したものを図4に示す。なお、図中には、実際の利用記録を入力とするトレースデータ駆動型シュミレーションとM/M1/PSによる遅延時間分布を同時に示している。図より、本モデルの妥当性がわかる。本モデルを利用することにより、WWWトラフィックの基本的特性が把握できるようになり、ユーザに対して提供すべき遅延時間を満足するアクセス回線容量が決定できるようになる。

3. 今後の課題

本稿で紹介した結果は、あるサブネット(情報処理教育センター)で発生するWWWトラフィック特性を明らかにし、それをアクセス回線の容量設計に適用したものであるが、実際には、プロキシサーバにおけるキャッシング機能や、バックボーンネットワークの性能特性、あるいは、WWWサーバのドキュメント転送能力など、容量設計に影響を与える要因はまだあり、研究の緒についたばかりである。現在、これらの問題に取り組んでいるところである。

参考文献

- 1) 村田正幸, “マルチメディアコンピュータネットワークの通信品質保証” 電子情報通信学会誌(掲載予定), vol.81, 1998.
- 2) 村田正幸, 宮原秀夫, “通信トラフィック理論とその応用 [I] - [VII]”, 電子情報通信学会誌(講座連載), vol.77-78, 1994-1995.
- 3) M. Nabe, M. Murata and H. Miyahara, “Analysis and modeling of World Wide Web traffic for capacity dimensioning of the Internet access lines”, to appear in Performance Evaluation, 1998.
- 4) N.J. Yeager and R.E. McGrath, “Web server technology”, Morgan Kaufmann Publishers Inc., 1996.
- 5) 滝根哲哉, 村田正幸 “通信網における待ち行列—理論の応用と課題—”, オペレーションズ・リサーチ学会誌 vol.43, PP.264-271, May 1998.

