

LANにおけるアプリケーション別トラフィックモデルの検討

NTTマルチメディアネットワーク研究所 *石橋圭介, ISHIBASHI Keisuke
 01009830 NTTマルチメディアネットワーク研究所 小沢利久, OZAWA Toshihisa

1. はじめに

インターネットを初めとするコンピュータ通信ネットワークではアプリケーションの増加に伴い、トラフィックの特性も多様化してきている。よって、これら多様なトラフィックに対し、システムの性能評価や網の設計・制御を適切に行うためには各種トラフィックの特性を把握し、標準的なモデルを提示しておく必要がある。本報告ではLAN上で様々なインターネットアプリケーションのトラフィック測定を行い、それらアプリケーションの情報発生間隔とサイズ分布をモデル化する。

2. モデル化対象

ここではプロトコル階層として図1を想定し、アプリケーションレイヤで発生するトラフィックをモデル化の対象とする。具体的にはWWWにおけるコンテンツ、ftpにおけるファイル、telnet

アプリケーション
TCP/UDP
IP
Ether

図1: プロトコル階層

におけるキーボード入力、TV会議における画像フレーム等を転送単位と呼び、この転送単位の発生間隔分布及び、サイズ分布をモデル化する。ところで、LANで測定されるIPパケットトラフィックはTCPのフロー制御や再送の影響を受けており、また、最大パケットサイズを超えるデータは分割されて送出されるため、アプリケーションのトラフィック発生過程を直接表現しているとは言えない。そこで本検討ではパケットヘッダの情報より個々の端末、アプリケーション間のトラフィックを抽出し、目的とするモデル化を行った。

3. LANトラフィック測定とモデル化

分析、モデル化に用いたデータは筆者が使用するLANから測定したアプリケーション別のトラフィックである。また、他研究機関の公開トラフィック測定データ [1] も参考のために同様に解析した。以下、整理をやすくするために、ftpのようなトラフィックをバルク転送型、telnetのようなトラフィックをインタラクティブ転送型、TV会議のようなトラフィックをストリーム転送型と呼ぶ。まず、バルク転送型、インタラクティブ転送型トラフィックに関して述べる。

図2, 3にWWWにおけるコンテンツサイズ、コンテンツ発生間隔を指数分布でモデル化したものを示す。指数分布の平均は実測データの算術平均により決定した。指数分布では分布

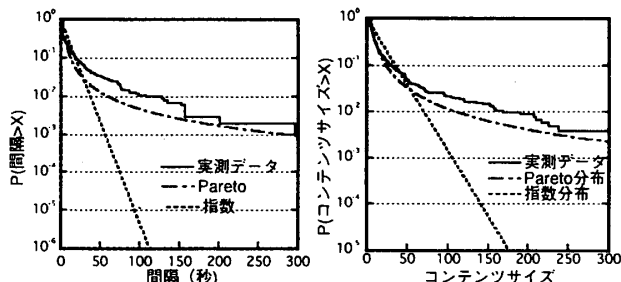


図2: 間隔の分布

図3: コンテンツサイズ分布

の裾の領域で実測データの分布からの乖離が見られる。これより、実測トラフィックを指数分布でモデル化すると、大サイズ発生確率の過小評価、発生間隔の分散の過小評価（バースト性の過小評価）等性能評価上好ましくない結果を導くと考えられる。ところで図の実測データのように長く裾を引く分布に対しては、”heavy-tailed” と呼ばれるクラスの分布が合うことが報告されている [2]。heavy-tailed 分布とは次式 (1) で表される性質を持つ分布であり、Pareto 分布、Weibull 分布などがその例である。

ここでは累積分布関数が次式 (2) の形となる

$$P(X > x) \approx x^{-\beta} \quad (x \rightarrow \infty) \quad (1)$$

Pareto 分布によりモデル化した。ここで β は形状パラメータと呼ばれるものであり、 a は最小値を決めるパラメータである。 b は log-log プロットの傾きから最小 2 乗法を用いて求め、 a は実測データの平均から求めた。

図2, 3から Pareto 分布と実測データの分布の

$$P(X \leq x) = 1 - (a/x)^{\beta} \quad (x > a) \quad (2)$$

裾はよくフィットしていることが分かる。しかし、Pareto 分布には最小値 a が存在するため、発生間隔の分布として用いた場合、一定時間トラフィックが発生しないモデルとなる等、0に近い領域で分布がフィットしないことが知られている [3]。ところで、これらトラフィックの発生の仕組みを考えた場合、例えばWWWトラフィックの発生間隔ではユーザのクリックする間隔と同一ページに含まれる複数データ転送の間隔など、時間スケール毎に異なる挙動が混在していると考えられる。

そこで、確率変数が0付近で指数分布、裾の部分でPareto分布に従う、累積分布関数が次式 (3) で表されるモデルを提案する。ここではこの分布をHybrid分布と呼ぶ。パラメータ b の推定はPareto分布の場合と同様に決定する。平均

は実測データの平均より決定した。

$$P(X \leq x) = \begin{cases} 1 - \exp(-x/\mu) & \text{if } x \leq \mu\beta \\ 1 - (\mu\beta/e x)^\beta & \text{if } x > \mu\beta \end{cases} \quad (3)$$

(e :自然対数の底, μ :平均)

次に指数, Pareto, Hybrid 分布の適合度を検定するため χ^2 検定を行った。結果は, 測定データの分散が非常に大きいのでどれも 95% 棄却閾外であったが, ほぼ全てのアプリケーションの分布で Hybrid 分布が最も良い適合度を示した。

しかし, 性能評価という視点からはもっと別の方法で分布推定のよしあしを判定しておく必要がある。そこで, ここでは WWW の実測データによる分布とそれを上記の各分布によってモデル化したものを入力トラフィックとして用いた場合の, 単一待ち行列における待ち行列長分布, 及び, TCP フロー制御を介した場合の平均転送時間をシミュレーションにより比較した。結果を図 4, 5 に示す。これより性能評価上, Hybrid 分布が良い近似を与えていることがわかる。ただし, シミュレーションではモデル化したサイズ, 発生間隔分布からランダムに生成した系列を入力トラフィックとして性能評価を行った。

次にストリーム転送型トラフィックに関して述べる。可変ビットレートの動画フレームサイズについては log-normal 分布, Γ 分布によるモデル化がよくフィットすることが報告されている [4] [5]。ここでは H261 符号化方式によるテレビ会議のフレームサイズに対し両分布を用いてフィッティングをおこなった。両分布ともパラメータとして測定データの平均, 分散を用いた。結果を図 6 に示す。明らかに両分布とも良い適合を示している。また平均対数尤度による比較においても有意な差は認められなかったので, 両分布をモデル化の候補とできる。ただし動画トラフィックでは連続するフレームサイズの間に強い相関があることが知られており, さ

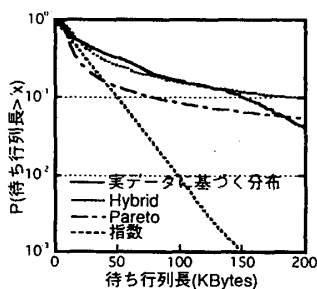


図 4: 待ち行列長分布

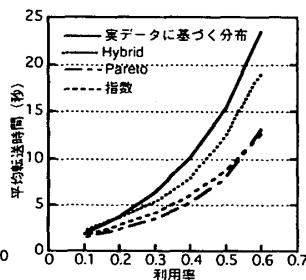


図 5: 平均転送時間

らにそのモデル化が必要である。

次に, インターネットホーンについてであるが, 通常は固定サイズの packets を固定間隔で送出している。しかし, アプリケーションによっては有音区間を

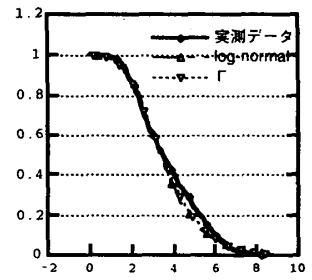


図 6: フレームサイズ分布

検知し, 有音区間のみ packets を発生するものもある。この場合有音区間のみ固定間隔で固定サイズの packets を発生するモデルが適当である。今回の測定データでは, 有音区間長は平均 1.29sec, 無音区間長は平均 1.91sec の指数分布になった。

表 1 に今回の結果をまとめる。

表 1 (上段: サイズ, 下段: 発生間隔)

フロータイプ	AP	分布	パラメータ値
バルク転送型	FTP	Hybrid	平均 83.9 ~ 172KB, $\beta=1.14 \sim 1.45$
		Hybrid	平均 23.0 ~ 129sec, $\beta=0.64 \sim 2.26$
	WWW	Hybrid	平均 5.31 ~ 15.3KB, $\beta=1.25 \sim 1.71$
		Hybrid	平均 1.81 ~ 29.7sec, $\beta=1.10 \sim 1.92$
	E-mail	Hybrid	平均 3.35 ~ 3.63KB, $\beta=1.34 \sim 1.54$
		指数	平均 9.51 ~ 1.16sec
インタラクティブ型	telnet	指数	平均 17.3 ~ 39.5bytes
		Hybrid	平均 0.47 ~ 1.16sec, $\beta=1.14 \sim 1.17$
ストリーム型	TV-会議	Γ , log-normal	平均 3.26 ~ 3.74KB, 分散 = $1.76e+6 \sim 2.54e+6$
		一定	1/8sec ~ 1/30sec (APIによる)
	MPEG	log-normal	Iフレーム: 平均 60.4KB, 分散 = $3.92e+8$ Pフレーム: 平均 23.1KB, 分散 = $2.15e+8$ Bフレーム: 平均 7.28KB, 分散 = $2.31e+8$
		一定	1/8sec ~ 1/30sec (APIによる)
		一定	108 ~ 188bytes (APIによる)
	Internet phone	一定	0.125sec ~ 0.198sec (APIによる)

AP: Application

4. まとめ

LAN トラフィック実測データを基にアプリケーション毎のトラフィックモデルを提示した。今後の課題としてはサイズや間隔の分布だけでなく相関も考慮したトラフィックモデルの提示が挙げられる。

参考文献

- [1] The Internet Traffic Archive <<http://town.hall.org/Archives/pub/ITA/>>.
- [2] V.Paxon et al., "Wide Area Traffic: The Failure of Poisson Modeling," IEEE/ACM Trans. Net., Jun, 1995, pp.226-244.
- [3] Bruce A.Mah, "A Empirical Model of HTTP Network Traffic," Proc of IEEE INFOCOM'97, 1997
- [4] D.Heyman et al., "Modeling Teleconference Traffic from VBR Video Coders," Proc. of IEEE ICC'94, 1994, pp 1744-1748.
- [5] M.Krunz et al., "Statistical Characteristics and Multiplexing of MPEG Streams," Proc. of IEEE INFOCOM'95, 1995, pp455-462.