

## 確率過程による LAN トラヒックのモデル化における一考察

上田 浩\* 那須野 裕† 岩谷 幸雄\* 五十嵐 隆治‡ 木下 哲男§

\* 東北大学 電気通信研究所 † 日本テレコム/東北大学東北大学大学院工学研究科電気・通信工学専攻  
‡ 秋田大学 工学資源学部 § 東北大学 情報シナジーセンター

### 要旨

東北大学電気通信研究所の基幹ネットワークスイッチの実測トラヒックを確率過程として捉え、ネットワーク状態と自己相似性の強弱を示す Hurst パラメータの関連について、ネットワークアプリケーションごとに検討を試みた。その結果、定常状態においては Hurst パラメータは長期間一定の値に収まっており、ネットワーク状態の変化に追隨して Hurst パラメータが変化する場合があることが分かった。

## Note on Modeling for LAN Traffic by Stochastic Process

Hiroshi UEDA\* Yutaka NASUNO† Yukio IWAYA\* Ryuji IGARASHI‡ Tetsuo KINOSHITA§

\* Research Institute of Electrical Communication, TOHOKU University

† Japan Telecom Corporation / Department of Electrical and Communication Engineering, Graduate School of Engineering, Tohoku University

‡ Faculty of Engineering and Resource Science, AKITA University

§ Information Synergy Center, TOHOKU University

### Abstract

We consider about Hurst parameter and network status by network application in R.I.E.C, Tohoku Univ. As a result, we observed that the case Hurst parameter varies with network status.

## 1 はじめに

これまで用いられてきたネットワークトラヒックに関するモデル化の例として、電話網における呼の到着のポアソン過程としてのモデル化を挙げる事ができる。ポアソン過程は電話の呼の到着間隔  $\tau$  が指数分布  $\tau = e^{-\mu\tau}$  に従う確率過程<sup>1</sup>であり、単位時間あたりの呼の到着数  $\mu$  のみで記述できるため、電話網の設計とそのトラヒック予測において実用上大きな利点がある。

一方、IP 網のトラヒックに関しては、1994 年以来、LAN トラヒックに自己相似性があるとの様々な報告がなされている [1, 2, 3, 4]。自己相似性とは、“

<sup>1</sup> ネットワーク上の様々なイベント (呼の発生時期、通信時間、必要帯域など) はいずれも不確定でランダムであるが、これらの事象がランダムに、ある確率分布に従って生起していると考えられ、確率過程を用いたモデル化によって、ネットワークの挙動をある程度理解できる。

対象の持つ特徴が長さや時間などのスケール変換に対して不変である性質”と定義できる [5]。LAN トラヒックにおいては、注目する単位時間のスケールに関わらず同様の変動が観測されることである [6]。

ポアソン過程に従うトラヒックは、注目する単位時間のスケールが大きくなると、ポアソン過程の平均値に漸近し平滑化されるのに対し、実測 LAN トラヒックは、注目する時間スケールが大きくなっても (すなわち、拡大しても) パースト性が失われない自己相似的挙動を示す場合があるため、ポアソン過程でモデル化するのは困難である場合があると言われている。

一方、近年の研究結果から、独立な多数のユーザがアクセスしているインターネットのバックボーントラヒックにはある種のランダム性があり、ポアソン過程に非常に近い性質を示すとの報告 [7] や、LAN、WAN など測定対象のネットワークの種類に

よって、また、観測の時間スケールに依存して、トラヒックの特性が自己相似であったり、ポアソンのであったりすることが指摘されている [8, 9] ため、トラヒックの性質はネットワークの利用状況に依存すると予想される。

ネットワークの利用状況を詳細に調べるための一つの手段として、http, ftp などのネットワークアプリケーション別のトラヒック解析を挙げるができる<sup>2</sup>。トラヒックはネットワーク上の(複数の)サービスのパケットを重畳したものであるから、ネットワークアプリケーション別のトラヒックの自己相似性を検証することに興味を持たれる。

本報告は、東北大学電気通信研究所の基幹ネットワークスイッチで観測したトラヒックを確率過程として捉え、その自己相似性を全トラヒックのみならず、ネットワークアプリケーション別に解析し、自己相似性の顕現に関し詳細に考察するものである。

## 2 自己相似性を持つトラヒックについて

本報告では時間区間  $(t, t+1]$  における到着パケット数を定常確率過程  $X_t$  として扱う。いま、 $X_t$  を  $m$  個ずつのブロックにまとめた過程を考える。すべての  $m \geq 1$  に対して、 $X_t$  と

$$X_t^{(m)}(H) = \frac{1}{m^H} \sum_{k=0}^{m-1} X_{mt+k} \quad (1)$$

が同じ確率法則に従うとき、 $X_t$  は  $H$  を自己相似パラメータを持つ自己相似過程であるという。 $H$  は Hurst パラメータと呼ばれる。 $m$  の変化は単位時間スケールの変化に相当し、 $m$  にかかわらず、もとの確率過程の時間スケールを  $m$  倍し、大きさのスケールを  $1/m^H$  倍した確率過程はもとの確率過程と同じ確率法則に従うことを意味している。

特に、 $X_t^{(m)}$  の自己相関関数

$$r_t^{(m)} \rightarrow r(t) \sim t^{-(2-2H)} \quad (2)$$

のとき、確率過程  $X_t$  は漸近的な 2 次の自己相似性を持つという。これは、十分大きな  $m$  に対しても自己相関関数の減衰が非常に緩やかであり、その積分値が発散することを意味している。

式 (1) より、 $H$  が 1 に近いほど自己相似性が強いといえる。このことは、確率過程の  $H$  によるモデル化の可能性を示唆している。

<sup>2</sup>たとえば [10] では、WWW サーバのファイル分布が heavy-tail であることがトラヒックの自己相似性の原因となっていることを主張している。

$H$  の推定方法には幾つかの手法があるが、本報告では R/S 解析と呼ばれる手法を用いる。R/S 解析とは、R/S 統計量と呼ばれる量を両対数グラフにプロットし、その回帰直線の傾きを求めることで  $H$  を推定する方法である。

確率過程  $X_t$  を  $m$  個ずつに分けたとき、 $l = 1, 2, \dots$  番目のブロックについて、ブロック内での平均値を  $X_l^{(m)}$  とするとき、

$$Z(l, m, i) = \sum_{j=(l-1)m+1}^{(l-1)m+i} X(j) - iX_l^{(m)} \quad (3)$$

$$(4)$$

で示される、 $i$  番目の時点での平均値からのずれを計算し、以下の R/S 統計量

$$\frac{R(l, m)}{S(l, m)} = \frac{\max Z(l, m, i) - \min Z(l, m, i)}{\sqrt{\frac{1}{m} \sum_{i=(l-1)m+1}^{lm} \{X_i - X_l^{(m)}\}^2}} \quad (5)$$

を求める。Hurst は、ナイル川の流量解析から経験的に  $m \rightarrow \infty$  のとき、 $1/2$  より大きい  $H$  に対して R/S 統計量が  $m^H$  に比例することを発見した [11]。すなわち、

$$\log E \left[ \frac{R}{S} \right] \approx a + H \log m \quad (6)$$

と近似できる。RS 統計量  $-m$  の両対数プロットについて、回帰直線を求めたときの傾きが Hurst パラメータの推定値となる。

## 3 解析結果と考察

**トラヒック観測手法** 東北大学電気通信研究所(以下通研)から外へのトラヒックは、4 台の基幹ルータと 1 台の対外接続装置を通して行くため、これらのトラヒックを収集することを考える。詳細には以下の通りである。

1. 対外接続装置、基幹ルータそれぞれの上位リンクをポート単位でミラーリング
2. ミラーリングしたトラフィックを、Gigabit Ethernet Controller を装備した FreeBSD OS のサーバ (Pentium4 / 2.80EGHz) 5 台で、tcpdump コマンド [12] を用いてファイルにパケットダンプ
3. 5 分ごと (HH:03, HH:08, HH:13 ...) にパケットダンプしたファイル (以下ログファイル) に日付と時間で名前を付けて保存
4. 5 分ごと (HH:00, HH:05, HH:10 ...) にログファイルを再び tcpdump で解析し、ネットワークアプリケーション (以下アプリケーション) ごとのパケット数をカウントしファイルに保存

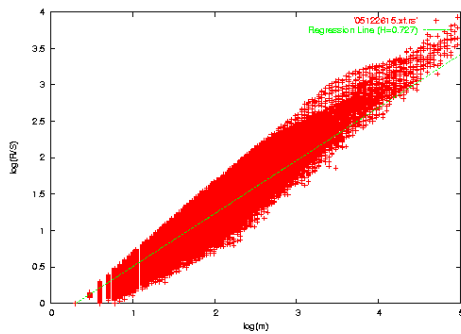


図 1: Pox plot diagram on RIEC traffic, Sep. 12/26 15:00-16:00

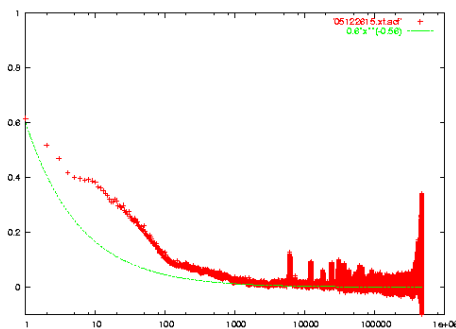


図 2: Autocorrelation function on RIEC traffic, Sep. 12/26 15:00-16:00

本手法にて取得したトラフィックデータのうち、対外接続装置の上位リンクに流れるトラフィックを解析した。

**アプリケーション別の Hurst パラメータ推定** 2005 年 12 月 26 日 15 時から 16 時のパケットトラフィックをサンプル時間 0.01 s を用いサンプリングし、R/S 解析を行い Hurst パラメータを推定した結果を図 1 に示す (この図のことを Pox plot diagram という)。H=0.72 であり、このとき、自己相関関数は式 (2) より、 $t^{-(2-2H)} = t^{-0.56}$  に比例する減衰特性を持つと推定される。図 2 より、計算した自己相関関数と、 $0.6 * t^{-0.56}$  が一致していることが分かる。以上より、通研のトラフィックには自己相似的な特性があり、さらに長期間のデータの傾向並びにアプリケーション別の自己相似性に興味を持たれる。

**長期間のデータからの Hurst パラメータ推定** 2005 年 9 月 1 日～11 月 30 日の 3 ヶ月間、5 分ごとにパケット数をカウントしたデータを用い、R/S 解析によりネットワークアプリケーション別の Hurst パラメータを推定した。3 ヶ月の Hurst パラメータ推定

表 1: Hurst parameters and rates classified by applications on Sep-Nov. 2005

Applications	H Parameter	Ave. # of Packet	Rate(%)
ALL	0.810657	225258.1	-
HTTP	0.815704	85867.5	38.12
Shell	0.767001	13904.14	6.173
FTP	0.659990	552.6	0.245
MAIL	0.698158	5124.83	2.275
DNS	0.735163	5483.48	2.434
ARP	0.498984	0.2	0.000
NetBIOS	0.767349	1196.0	0.531
ICMP	0.735883	1308.82	0.581
RTSP	0.761934	4565.14	2.027

表 2: Hurst parameters and rates classified by applications on Sep. 2005

Applications	H Parameter	Ave. # of Packet	Rate(%)
ALL	0.806629	210515.9	-
HTTP	0.810084	83598.2	39.71
Shell	0.761240	9067.9	4.308
FTP	0.656358	61.5	0.0292
MAIL	0.698465	5526.1	2.625
DNS	0.716187	5631.6	2.675
ARP	0.493772	0.2	0.0001
NetBIOS	0.757383	1128.8	0.5362
ICMP	0.750535	960.1	0.4561
RTSP	0.755208	3117.0	1.481

値を表 1 に、9 月、10 月、11 月それぞれの推定値をそれぞれ表 2, 3, 4 に示す。アプリケーションの内訳については、HTTP には HTTPS を、Mail には SMTP/POP3/IMAP をそれぞれ含むものとする。

アプリケーション別の Hurst パラメータの推定値から以下の知見を得ることができる。

- HTTP, RTSP(ストリーミング), ICMP など、データの送受信にバースト性が高いアプリケーションの自己相似性が強い。
- FTP など、一定時間一様なトラフィックを発生させるアプリケーションの自己相似性は弱い。

さらに、表 1 と表 2～4 の比較から、アプリケーション別の Hurst パラメータは 3 ヶ月間ほぼ一定であることが分かる。このことから、準リアルタイムで Hurst パラメータを推定することによるネットワーク状態検出への応用の可能性が示唆される。

**ネットワーク利用状況の変化と Hurst パラメータ** ネットワークの利用状況の変化と Hurst パラメータの関係について考察するため、2005 年 9 月 28 日に行なわれた、通研のシステム移行時の Hurst パラメータを 5 分ごとに推定した。結果を図 3 に示す。

表 3: Hurst parameters and rates classified by applications on Oct. 2005

Applications	H Parameter	Ave. # of Packet	Rate(%)
ALL	0.812845	228157.6	-
HTTP	0.813749	86663.3	37.98
Shell	0.780706	19745.6	8.650
FTP	0.652916	129.51	0.060
MAIL	0.694190	5097.6	2.230
DNS	0.723321	5307.9	2.330
ARP	0.505288	0.2	0.000
NetBIOS	0.764893	1266.4	0.560
ICMP	0.690234	1049.5	0.460
RTSP	0.759443	4551.4	1.990

表 4: Hurst parameters and rates classified by applications on Nov. 2005

Applications	H Parameter	Ave. # of Packet	Rate(%)
ALL	0.811274	252965.8	-
HTTP	0.821725	91738.4	36.27
Shell	0.760737	15732.82	6.220
FTP	0.667794	1743.3	0.690
MAIL	0.700622	4618.0	1.830
DNS	0.769630	5471.7	2.160
ARP	0.514227	0.190	0.000
NetBIOS	0.748986	1278.6	0.510
ICMP	0.748571	2280.7	0.900
RTSP	0.773874	7093.88	2.800

システム移行時刻である、12:00 を含む時間帯では、いずれも Hurst パラメータが減少しているため、この時間帯においては、システムの切り替えに伴いトラヒックの特性が変化していることが窺える。

## 4 おわりに

本報告では、IP 網のパケットトラヒックを確率過程として捉え、その自己相似性の検証に加え、アプ

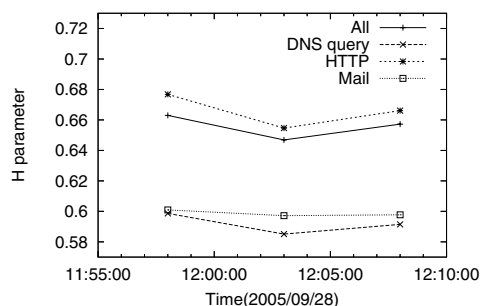


図 3: Variation in Hurst parameter at system switch

リケーションでスライスして自己相似性パラメータを推定することで、ネットワークの状態検出への応用の可能性を検討した。長期間観測したトラヒックの Hurst パラメータはほぼ一定の値に収まっており、これをネットワークの定常状態と定義して、Hurst パラメータの変化とネットワーク状態の変化の関連に明確な関連があるかどうか、システム移行時のトラヒックを例に挙げて検討した。

今後は、観測をさらに長期間続け、ネットワークの何らかの状態変化と自己相似性の関連をさらに明確に示すことができるように解析を続けて行く必要がある。

## 参考文献

- [1] Leland, W. E., Taqqu, M. S., Willinger, W. and Wilson, D. V.: On the Self-Similar Nature of Ethernet Traffic (Extended Version), *IEEE/ACM Transactions on networking*, Vol. 2, No. 1, pp. 1–15 (1994).
- [2] Beran, J., Sherman, R., Taqqu, M. S. and Willinger, W.: Long-Range Dependence in Variable-Bit Rate Video Traffic, *IEE Trans. Commun.*, Vol. 43, No. 2/3/4, pp. 1566–1579 (1995).
- [3] Taqqu, M. S., Teverovsky, V. and Willinger, W.: Estimators for long-range dependence: an empirical study, *Fractals*, Vol. 3, No. 4, pp. 785–798 (1995).
- [4] Willinger, W., Taqqu, M. S., Leland, W. E. and Wilson, D. V.: Self-Similarity in High-Speed Packet Traffic; Analysis and Modeling of Ethernet Traffic Measurements, *Statistical Science*, Vol. 10, No. 1, pp. 67–85 (1995).
- [5] 小沢利久: 長期依存性/自己相似性を持つトラヒックのモデル, システム/制御/情報, Vol. 43, No. 3, pp. 117–122 (1999).
- [6] Stallings, W.: *High-Speed Networks: TCP/IP and ATM Design Principles*, Prentice Hall, chapter 8, pp. 181–207 (1998).
- [7] Zhang, Z., Ribeiro, V., Moon, S. and Diot, C.: Small-time scaling behaviors of Internet backbone traffic: An empirical study, *Proc. INFOCOM 2003*, pp. 1826–1836 (2003).
- [8] 佐藤昌平, 吉田万貴子: 次世代インターネットとトラヒック工学, 電子情報通信学会論文誌, Vol. J85-B, No. 6, pp. 875–889 (2002).
- [9] Karagiannis, T., Molle, M., Faloutsos, M. and Broido, A.: A Nonstationary Poisson View of Internet Traffic, *INFOCOM 2004*, Vol. 23, No. 1, IEEE, pp. 1559–1570 (2004).
- [10] Park, K., Kim, G. and Crovelli, M.: On the relationship between file sizes, transport protocols, and self-similar network traffic, *Proc. IEEE International Conference on Network Protocol*, pp. 171–180 (1996).
- [11] Hurst, H. E.: Long-Term Storage Capacity of Reservoirs, *Trans. Amer. Soc. Civil Eng.*, Vol. 116, pp. 770–799 (1951).
- [12] LBNL's Network Research Group: tcpdump. <ftp://ftp.ee.lbl.gov/tcpdump.tar.Z>.