

適正時間スケールを利用した 長期依存性トラヒックの近似解析

計 宇生 藤野貴之 阿部俊二

趙偉平 松方 純 浅野正一郎

国立情報学研究所

〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: netgrp@netsp.r.nii.ac.jp

あらまし インターネットトラヒックには長期依存性ないし自己相似性が広く観察されている。本研究では、実際に観測されたネットワークトラヒックのデータを利用して、その長期依存性がトラヒックのキューイング性能に与える影響について検討する。FBM モデルを使ったキューイング性能の近似解析は、厳密な自己相似性を持つトラヒックのために提案されているが、ここでは、適正時間スケールの概念を利用することによって、より一般的な長期依存特性をもつトラヒックの近似解析方法を提案する。また、実際のトラヒックデータを用いたシミュレーションの結果を使って、近似解析の結果との比較、検証を行う。

キーワード 長期依存性, 自己相似性, 適正時間スケール, FBM, インターネットトラヒック

Performance Approximation of Long-Range Dependent Traffic Using Relevant Time Scales

Yusheng JI Takayuki FUJINO Shunji ABE Weiping ZHAO

Jun MATSUKATA and Shoichiro ASANO

National Institute of Informatics

2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: netgrp@netsp.r.nii.ac.jp

Abstract It is well known that long-range dependence or so-called self-similarity is widely observed in internet traffic. In this study, we discuss on the influence of long-range dependence in queueing performance by using our observed traffic data from the Science Information Network (SINET). Although Fractional Brownian Motion (FBM) model was proposed in the literature to deal with such fractal traffic characteristics, it is only efficient for approximating the performance of strict second order self-similar traffic. Here we propose a method for approximation of the performance of traffic with general long-range dependence by applying the notation of relevant time scales. Simulation results from real traffic data are used for the verification of our method.

Keyword Long-Range Dependence, Self-Similarity, Relevant Time Scale, FBM, Internet Traffic

1. はじめ

インターネットトラフィックには長期依存性ないし自己相似性が広く観察されている。そのような長期依存性や自己相似性は LAN や、広域ネットワーク、または動画像のトラフィックなどにみられ、多重化されたネットワークトラフィックに広く存在していることが報告されている^{[1]-[3]}。本研究では、実際に観測されたネットワークトラフィックを利用して、その長期依存性がキューイングの性能に与える影響について検討する。ここでは、厳密な自己相似トラフィックにのみ適用できる FBM による近似解析を、適正時間スケール概念を利用することによって、より一般的な長期依存特性をもつトラフィックへの適用方法を提案する。また、実際に観測されたトラフィックデータを用いて、比較、検証を行う。

2. SINET のトラフィックと長期依存性

筆者らは国立情報学研究所で運用されている学術情報ネットワーク (SINET)^[4] の IP トラフィックに対して観測を行った。それにより、観測期間内にネットワーク回線を通過する個々の IP パケットの到着時刻とパケット長のデータが得られた。それらデータに対する統計解析の結果、一般的に知られているようなネットワークトラフィックにおける長期依存性ないし自己相似性がみとめられた^[5]。本節では、実測されたネットワークトラフィックの統計特性について述べる。

一般に、定常増分を有する確率過程 $X(t)$ の自己相関関数が $r(k)$ であり、次のような条件を満足する場合、 $X(t)$ が 2 次の自己相似性を有するといひ、 H が自己相関関数の時間的減衰の速さ (遅さ) を示すパラメータで、ハーストパラメータという。

$$r(k) = \frac{1}{2} \{ (k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \} \quad (1)$$

また、 $X(t)$ の連続した m 個 (または長さ m) の平均

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(t), \quad k=1,2,\dots$$

を考えると、2 次の自己相似性は次式のようにも表される。

$$\text{Var}[X^{(m)}(k)] = \sigma^2 m^{-2\beta} \quad (2)$$

σ^2 は $X(t)$ の分散であり、 $\beta=2(1-H)$ である。また、 $X^{(m)}(k)$ の自己相関関数 $r^{(m)}(k)$ が、 m の増大に対し

て、漸近的に式 (1) の右側になる場合、あるいは、式 (2) が m の増大に対して漸的に成り立つ場合、漸近的な 2 次の自己相似性を有するという。

一方、自己相関関数の k に対する減衰が緩やかであることを表す性質として、長期依存性がある。確率過程 $X(t)$ が長期依存性を有する場合、次の式が成り立つ。

$$\lim_{k \rightarrow \infty} r(k) = k^{-\beta} L(k) \quad (3)$$

ただし、 $L(t)$ は任意の $x > 0$ に対して、

$\lim_{t \rightarrow \infty} L(tx)/L(t) = 1$ を満足するような、 t の十分大

きな領域では緩やかに変化する関数である。長期依存性は漸近的な 2 次の自己相似性よりも少し広い概念で、式 (2) の $L(k)$ が定数である場合は漸近的な 2 次の自己相似性となり、 $\beta=2(1-H)$ である。2 次の自己相似性は式 (2) の性質から、異なる時間スケール上における $X(t)$ の m 個の平均 $X^{(m)}(k)$ の分散特性の m による変化などで見ることができる。

次に SINET で観測されたトラフィックの特性について述べる。図 1 は 2001 年の 2 月から 5 月にかけて、SINET バックボーンで観測されたトラフィックで、1 時間分のデータを単位時間で区切られた場合の単位時間内の到着ビット数の分散特性 (Variance-Time Plot) を示したものである。ここでは十分な観測精度が確保できる上、時間スケールの充分小さな領域の特性も表現できるように、単位時間を 10^{-3} 秒とした。

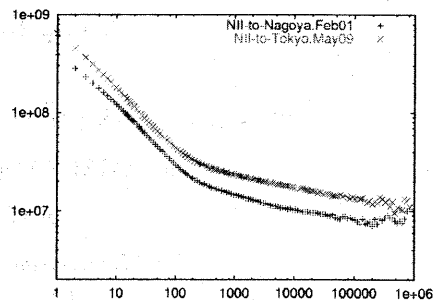


図 1 SINET トラフィックの Variance-Time Plot

図 1 に 2 系列のトラフィックデータの分散特性を示したが、2 つの系列のトラフィックではレートが異なるため、異なる分散値を示しているが、分散特性の時間スケール上での減衰の様子がよく似ていることがわかる。さらに、それぞれの系列のデータをみると、時間スケール m の増大による

$X^{(m)}(k)$ の分散の減衰の速さは m の領域によって異なっている。 m の比較的小さい (0.2 秒以下の) 領域においては減衰が比較的速く、ポアソン特性に近い (ポアソン分布のハーストパラメータが 0.5 であるのに対し、ここでは 0.6 以上に相当する)。それに対して、比較的長い時間スケール (0.2 秒以上) においては、分散の減衰がより緩やかになり、ハーストパラメータ 0.9 以上に相当する。図 1 からわかるように、SINET トラヒックも長期依存性を有している。

文献[6]などでは、Fractional Brownian Motion (FBM) を用いて自己相似性トラヒックを近似することが提案されている。到着率が λ 、分散係数が a の確率過程が FBM であるとすれば、時刻 t までに到着するトラヒックの量が

$$A(t) = \lambda t + \sqrt{\lambda a} Z_H(t) \quad (4)$$

で表される。 $Z_H(t)$ は平均 0、分散 $\text{Var}[Z_H(t)] = |t|^{2H}$ 、ハーストパラメータ $H \in [0.5, 1)$ を持つ FBM 確率変数である。この場合、 $A(t)$ の増分過程 $Y(k) = A(k+1) - A(k)$ ($k > 0$) は Fractional Gaussian Noise (FGN) と呼ばれ、式(1)で表す厳密な 2 次の自己相似性を満足する。

FBM は正規性を有しているため、このような過程は、 $\{\lambda, a, H\}$ の 3 つのパラメータで完全に決まる。中央極限定理によれば、たくさんの 2 次の自己相似性を有するような、同一の分布をもつ独立な確率過程 (i.i.d. 過程) の集約は FBM になる。そのため、FBM は集約トラヒックをモデル化するために比較的有効であることが多く報告されている^{[7][8][9]}。

一方、文献[10]などで指摘されているように、TCP の Self-Clocking 効果などにより、トラヒックの集約は短い時間スケールにおいてポアソンの性質を示すことがある。図 1 の SINET トラヒックの分散特性を見ても、時間スケールの短い領域では、ポアソンの性質に近い特性が現れている。したがって、観測された SINET のトラヒックは厳密的な自己相似性ではなく、漸近的な自己相似性、またはより一般的な長期依存性をもっていると言える。

3. トラヒックの性能と適正時間スケール

一般的に、トラヒックの性能は、問題視されている性能に対応した時間スケールにおけるトラヒックの特性に依存する。長期依存性のトラヒックは、キューイング性能に直接影響を与えるトラヒックの変動具合が長い時間スケールにおいて

保たれるため、比較的長い待ち行列の生成につながる事が知られている。しかし、待ち行列の長さを有限な領域で注目している限り、それに対応した時間スケールでのトラヒックの特性が問題になる^{[11][12]}。そのような待ち行列長に対応した時間スケールを適正時間スケール (Relevant Time Scale) と呼ばれ、文献[12]では長期依存性トラヒックの適正時間スケールを導出している。

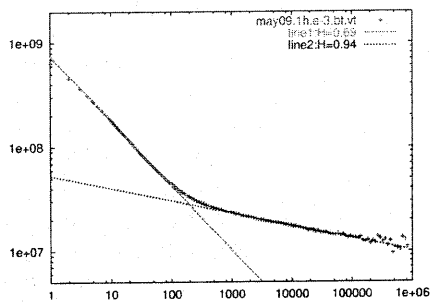
バッファ長を B とした場合、到着率 λ 、ハーストパラメータ H の正規性確率過程がサービス率 C のサーバにおけるキューイング性能の適正時間スケール T は次式で表される。

$$T = \frac{B}{C - \lambda} \frac{H}{1 - H} \quad (5)$$

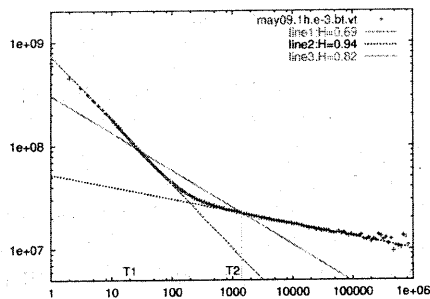
FBM のような正規性を持つ自己相似性確率過程は、 λ, a および H によって決まることが前節で述べた通りであるが、その性能は対応する時間スケールにもよる。

本節では、式 (5) で示した適正時間スケールを一般的な長期依存性をもつトラヒックの解析に利用することを考える。観測された SINET トラヒックのように、異なる時間スケールにおいて、トラヒックの分散特性の減衰の速さが異なっているため、まずは図 2 (a) のように、分散特性 (Variance-Time 特性) を複数 (ここでは 2 本) の直線で近似的に表す。トラヒックの分散特性が直線上にのっている場合は、厳密的な 2 次の自己相似性を有することになる。さらに正規性を仮定すれば、それぞれの直線に対応したものは、一つの FBM 過程として考えることができる。

次に、このようにできた、パラメータの異なる複数の FBM 過程の性能を比較した場合、適正時間スケールを利用すれば、それぞれの直線に対応した FBM 過程がどの領域で性能に最も大きな影響を与えるかがわかる。正規性を持つ確率過程のキューイング性能は、到着量の変動、すなわち 2 次モーメント (分散) によって直接影響される。図 2 (a) で示したような 2 本の直線の場合、短い時間スケールでは、 $H_1=0.69$ に対応したもの (これを直線 1 と呼ぶことにする)、長い時間スケールでは、 $H_2=0.94$ に対応したもの (これを直線 2 と呼ぶ) の $X^{(m)}(k)$ の分散の値が大きく、二つの直線が交差するところ ($m=0.2$ 秒) で $X^{(m)}(k)$ の分散の値が同じになる。したがって、短い時間スケールでは直線 1、長い時間スケールでは直線 2 のほうがキューイング性能により悪い影響を与えることが直感でもわかる。



(a)



(b)

図2 分散特性の直線近似

ここでは、2つのFBM過程のハーストパラメータが $H_1 < H_2$ であるため、式(5)における他のパラメータが同じである場合、適正時間スケールは $T_1 < T_2$ となる。そこで、2つのFBM過程の他のパラメータのとり値が揃っている、2本の直線の交点の0.2秒周辺について考える。 $T_2 = 0.2$ 秒の時の T_1 および $T_1 = 0.2$ 秒の時の T_2 を計算すると、それぞれ $T_1 = 0.028$ 秒、 $T_2 = 1.4$ 秒になる。したがって、0.028秒より短い時間スケールでは直線1に対応したもの、1.4秒より長い時間スケールでは直線2に対応したもののほうがより性能に悪影響を与えることと言える。0.028秒と1.4秒の間の領域では、この方法では比較することができないため、ここでは、それらよりもさらに性能に悪影響を与えるような、図2(b)で示すような3本目の直線、直線3 ($H=0.82$) を考える。 $T_1 = 0.028$ 秒と $T_2 = 1.4$ 秒の間の領域では、直線3の分散の値が最も大きいため、直線1と直線2よりも性能への影響が大きい。また、実際のトラヒックの分散特性も、直線1と直線2が交差する辺りでは、二つの直線よりも分散の値が大きくなっているため、それを直線で近似するためには直線3が最も適当

と考えられるからである。

上述の方法を一般化すれば、分散特性が図2のような折れ線に近い形をしている場合、まず、 n ($n \geq 1$) 個の点で交差する $n+1$ 本の直線を書き、さらに、 n 本の直線を書き添えることによって、計 $2n+1$ 本の直線で、 $2n+1$ 個のFBM過程に対応させることができる。

FBMを利用した待ち行列長の近似解析の結果[6][13]によれば、待ち行列長の補分布 $\Pr(Q > x)$ は漸近的に次のような式になる。

$$\Pr(Q > x) \sim e^{-x^\gamma} \quad \text{as } x \rightarrow \infty \quad (6)$$

ただし、ここでは、

$$\gamma = \frac{1}{2a\lambda(1-H)^2} \left(\frac{(C-\lambda)(1-H)}{H} \right)^{2n}$$

である。

式(6)を上述のようにできた3つの直線のそれぞれに対応したFBM過程に適用すれば、図3のような、3つの性能曲線ができる。3つの曲線の交差する部分をよりわかりやすく示すために、図3では、log-logスケールを用いた。また、待ち行列に加わる負荷が図3の(a)と(b)では、それぞれ66.7%と80%であり、負荷の違いによって3つの曲線の交わる位置が異なることがわかる。

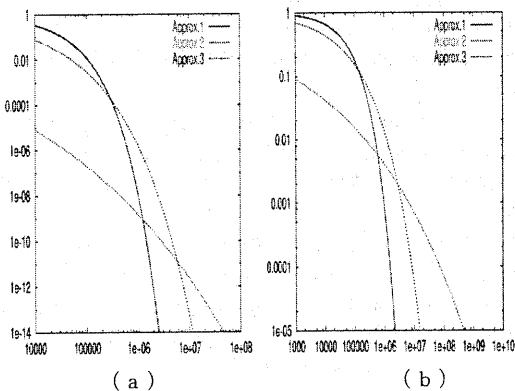


図3 複数のFBMモデルによるキュー長の近似

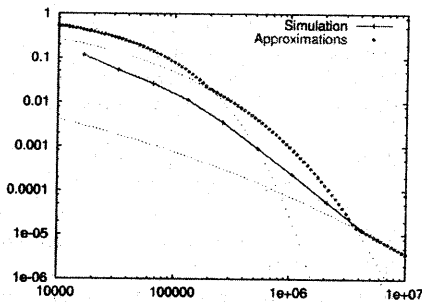
図2(b)の分散特性を有する実測データのキューイング性能は、3つの曲線の中の最も性能の悪い部分を通る曲線で近似することが考えられる。ここでは、図3の3つの曲線のもっとも性能の悪い部分をつなげると、曲線と曲線の繋ぎ目のところ(特に直線3に対応する性能曲線と直線2

に対応する性能曲線の間)では滑らかさが欠けてしまっているが、それは図2(b)で近似に用いた複数の直線の勾配の違いによるものである。したがって、勾配の差が少ない直線をより多数本用いれば、より滑らかな近似性能曲線を得ることができる。

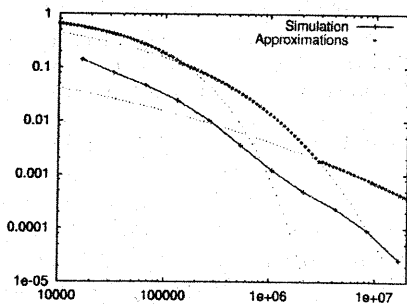
4. 観測トラヒックの性能との比較

前節で述べた近似方法を検証するために、観測された SINET のトラヒックデータを使って、シミュレーションによってそのキューイング性能を調べた。

まず、有限バッファの場合のバッファ溢れ率についてシミュレーションを行った。図4にバッファ溢れ率とバッファ長の関係を log-log スケールで示した。さらに、前節で述べたような複数の FBM モデルによる異なる時間スケールでの近似値も合わせて図4に示した。ここでは、3つの性能曲線の最も性能の悪い部分を太い点線で強調し、他の部分は細い点線で示してある。図4の(a)と(b)はそれぞれ負荷が75%と80%の時である。



(a) 負荷 75%の時

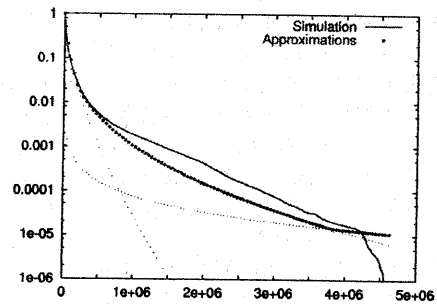


(b) 負荷 80%の時

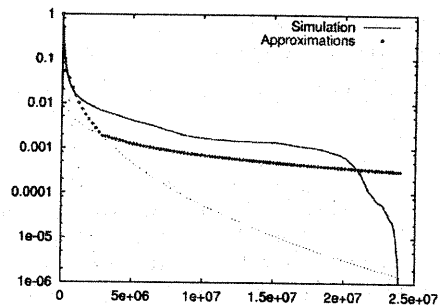
図4 バッファ溢れ率の評価

図4では、待ち行列長の補分布の近似解析値がシミュレーション値のバッファ溢れ率よりも大きくなっており、特にバッファ長が短い時や、負荷の大きいときにはその差が顕著である。それはバッファ溢れ率を考えた場合、バッファから溢れたものがサーバのサービスの対象から除かれるため、待ち行列長の補分布に比べ、待ち行列が見かけ上短くなるためであると考えられる。バッファ溢れ率が大きいときには、特にその差が大きいので、負荷の大きいときや、バッファ長が短いときには大きな差を生じているのである。したがって、バッファ溢れ率だけでは待ち行列長の補分布の近似方法の有効性を検証できないが、解析の結果はシミュレーションの結果と矛盾を生じていないことだけは言える。

次に、観測されたトラヒックデータを使ってシミュレーションで実際の待ち行列長分布を調べた。図5ではシミュレーションによる結果と近似解析の結果を比較している。図4と同じように図5の(a)と(b)はそれぞれ負荷が75%と80%のときである。



(a) 負荷 75%の時



(b) 負荷 80%の時

図5 待ち行列長の補分布の評価

シミュレーションで扱えるデータが有限であるため、待ち行列長の長い区間で得られた値には誤差が生じているが、そのような部分を除けば、シミュレーションによって得られた結果は近似解析の結果よりもやや悪い性能を示すものの、一般的には同じような傾向を示していることが言える。

シミュレーションによるキューイングの性能が悪くなる原因として考えられるのは、近似解析ではトラヒック分布の正規性を仮定しているが、実測データでは必ずしも正規性が成り立つとは言えないことである。実際に使ったデータの歪度と尖度を計算してみると、サンプル数が360万に対して、それぞれ-0.14と-0.52である。このように、トラヒックデータの歪度と尖度がトラヒックのキューイング性能に悪影響を及ぼすとの報告もある^[14]。

5. おわりに

インターネットトラヒックの多くは、厳密的な自己相似性を有するのではなく、より一般的な長期依存性をもつことが多い。これまでに提案されているFBMは厳密的な自己相似性をもつトラヒックの性能の近似解析には有効であるが、本研究では、より一般的な長期依存性トラヒックに対しての近似方法を検討した。トラヒックの2次モーメントの時間スケール上の特性を考慮して、適正時間スケールと呼ばれるものを利用することによって、より一般的な長期依存性トラヒックの性能を近似する方法について提案した。また、実測トラヒックを用いたシミュレーションの結果と比較を行った。適正時間スケールを考慮することによって、現実のトラヒックによりよい近似が得られることを示した。

今後、その他のトラヒックデータを使って本近似手法の検証を進めるとともに、トラヒックの特性が正規性から外れたときには性能にどのような影響を及ぼすかについてもより詳しく調べていく予定である。

文 献

- [1] W. Leland, M. Taqu, W. Willinger and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol.2, no.1, pp.1-15, 1994.
- [2] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Trans. Networking*, vol.3, no.3, pp.226-244, 1995.
- [3] T. Kato, Y. Ji and S. Asano, "A Study on the Self-Similarity of MPEG2 Video Traffic and the Effect for Transmission Quality," *Proc. 10th International Packet Video Workshop*, Paper No.10, 2000.
- [4] 浅野, "学術情報ネットワーク," *電子情報通信学会誌*, vol.81, no.4, pp.402-406, 1998.
- [5] 藤野, 阿部, 計, 趙, 松方, 浅野, "実トラヒック測定データを用いたシミュレーションによるインターネット品質保証制御方式の一検討," *電子情報通信学会技術研究報告*, SSE2000-306, pp.505-510, 2001.
- [6] I. Norros, "On the Use of Fractional Brownian Motion in the Theory of Connectionless Network," *IEEE JSAC*, vol.13, no.6, pp.953-962, 1995.
- [7] 阿部, 藤野, 計, 松方, 浅野, "学術情報ネットワークにおける自己相関トラヒックの解析," *インターネットコンファレンス 2001 論文集*, pp.151-158, 2001.
- [8] J. Beran, R. Sherman, M. Taqu and W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic," *IEEE Trans. Commun.*, vol.43, no.2/3/4, pp.1566-1579, 1995.
- [9] A. Erramilli, O. Narayan and W. Willinger, "Experimental Queuing Analysis with Long-Range Dependent Packet Traffic," *IEEE/ACM Trans. Networking*, vol.4, no.2, 1996.
- [10] 古屋, 中村, 野本, "IPバックボーントラヒックの局所ポアソン性に関する時間スケールとネットワーク設計への応用," *電子情報通信学会技術研究報告*, IN2001-108, pp.9-16, 2001.
- [11] M. Montgomery and G. De Veciana, "On the Relevance of Time Scales in Performance Oriented Traffic Characterizations," *Proc. IEEE INFOCOM'96*, pp.513-520, 1996.
- [12] A. Neidhardt and J. Wang, "The Concept of Relevant Time Scales and Its Application to Queuing Analysis of Self-Similar Traffic (or Is Hurst Naughty or Nice?)" *Proc. ACM SIGMETRICS'98*, pp.222-232, 1998.
- [13] 小沢, 町原, 石橋, "マルチメディアトラヒック理論の最新動向," *電子情報通信学会誌*, vol.81, no.5, pp.506-515, 1998.
- [14] T. Mori and R. Kawahara, "A Study on the Difference in Marginal Distribution of Network Traffic," *電子情報通信学会技術研究報告*, IN2001-107, pp.1-7, 2001.