

長時間トラフィックに含まれるトレンドの効果

五十嵐 隆治*, 入江 若菜**, 高橋 秋典*, 正木 忠良***, 佐々木 芳宏***
上田 浩†, 奈須野 裕††, 岩谷 幸雄†††, 木下 哲男††††

*秋田大学工学資源学部情報工学科, **秋田大学大学院工学資源学研究所, ***秋田大学総合情報処理センター
†群馬大学総合情報メディアセンター, ††苫小牧工業高等専門学校電気電子工学科
†††東北大学電気通信研究所, ††††東北大学大学院情報科学研究科

本研究では 1 週間以上の長時間に渡り観測した LAN トラフィック時系列から Low-Pass フィルタリングにより Diurnal Variation (日間変化) に基づくトレンドを抽出し, このトレンドを除去した後のランダム成分に関し, 自己相似性の観点から検討してみた. この結果, トレンド除去後の時系列にも自己相似性が認められるという結果を得た.

キーワード: ネットワークトラフィック, トレンド, ハーストパラメータ, R/S 解析

Effect of Trend of Long-Time Network Traffic

Ryuji Igarashi*, Wakana I-rie**, Akinori Takahashi*, Tadayoshi Masaki***, Yoshihiro Sasaki***
Hiroshi Ueda†, Yutaka Nasuno††, Yukio Iwaya†††, Tetsuo Kinoshita††††

*Department of Computer Science and Engineering, Akita University,
**Graduate School of Engineering and resource science, Akita University
***General Information Processing Center, Akita University, †Library and Information Technology Center, Gunma University
††Department of Electrical and Electronic Engineering, Tomakomai National College of Technology,
†††Research Institute of Electrical Communication (RIEC), Tohoku University,
††††Graduate School of Information Sciences, Tohoku University

Traffic time series observed in longer period than some tens of hours has such trend as diurnal variation or weekly variation. In case stochastic processes are investigated, a separation between random component and deterministic trend or periodicity should be practiced. In this paper a stochastic property of the trend-removed traffic is discussed in terms of self-similar nature. Our results in small scaled LANs tell that the time series after removing the trend still hold a self-similar nature.

Keywords: Network traffic, Trend, Hurst parameter, R/S analysis

1. はじめに

パケットトラフィック時系列は従来の通信路モデルで採用されていたポアソン過程での記述は困難で, 自己相似過程を考慮したモデル化が適切であると指摘されている¹⁾. しかしパケットトラフィックが自己相似過程に従うと指摘された当初から, その特性は実は自己相似性ではなく, 非定常ランダム過程が見かけ上自己相似過程として観測される擬似自己相似過程ではないか, ということも議論されている²⁾. 実際トラフィックには顕著な長期的変化として, Diurnal Variation (日間変化) や週に渡る変化があり, User Pattern (使用傾向) も特異なトラフィック疎通傾向を呈することがある. このような状況はトポロジー形態とも相まって, 複雑なトラフィック特性を誘起すると考えられる. さらに近年大きな不都合を惹起しているのがインターネット上での DoS・DDoS 攻撃や不正使用である. これら以外にも特定のストリーミングの増加や特別な興味を喚起するホームページへの閲覧の特異な増加などは, 広い意味で異常トラフィックと認識される³⁾. DDoS のような Flood 攻撃はトラフィックの

Diurnal Variation から見ると短時間異常と見なせるが, アプリケーションに基づく広い意味での異常は Diurnal Variation の時間スケールでも観測される異常である⁴⁾. このような異常の抽出のためにはトラフィック時系列の信号処理的な取り扱いや⁵⁾⁻⁸⁾ 主成分分析に基づく長時間, 中間時間, 短時間トラフィックの解析⁹⁾ が有効であると考えられる. これらはトラフィック時系列を複数の成分に分解したのち, その特性を解析するという提案である. トラフィックには Diurnal Variation があるため一般的には非定常過程であるとみなされている. ゆえにこのような場合の信号処理的手法としては, 通常, Wavelet 解析が用いられるが⁹⁾⁻¹¹⁾, Fourier 解析も依然有用なツールである⁸⁾.

本研究では小規模な教育研究機関におけるトラフィックデータを扱うが, このデータには定常的な Diurnal Variation が認められる. Diurnal Variation が定常と見なせるような長時間トラフィックにおいては, Fourier 解析によるフィルタリングも妥当であると考えられる. 我々は Fourier 解析を用いた周波数領域での理想フィルタリング, すなわち sinc 関数による Low-Pass フィルタリ

グにより長時間トラフィックの定常成分抽出を試みた。本研究ではトラフィック時系列を定常成分とランダム成分とに分離し、分離した後のランダム時系列の特性を検討することにより、長時間トラフィック特性の基礎的な知見を得ることを目的とする。

2. 測定

2.1 条件

本研究では、ノード数が 200 程度の教育研究機関 A、およびその 10 数倍程度となる教育研究機関 B の 2 種の機関において得られたトラフィックデータを取り扱う。フロー量は単位時間 Δt 当たりのパケット数 pps で表し、 $\Delta t=1$ s, 10 s, 600 s の 3 領域での取り扱いとした。

2.2 定常成分とランダム成分

大規模 ISP におけるバックボーントラフィック時系列 x_t は、統計学において通例用いられる線形なモデル¹⁰⁾

$$x_t = y_t + T_t + S_t, \\ y_t: \text{Random Component}, T_t: \text{Trend}, \\ S_t: \text{Seasonality} \text{-----} \quad (1)$$

ではなく、

$$x_t = \sqrt{am_t} y_t + m_t, \\ y_t: \text{Random Component}, m_t = T_t S_t, \\ T_t: \text{Trend}, S_t: \text{Seasonality}, \\ a: \text{Peakedness} \text{-----} \quad (2)$$

のようなモデル⁴⁾が妥当なかもしれないが、本研究では容易に基礎的な知見を得ることを目的とし、(1)式のモデルにより解析を進めた。

2.3 トラフィック時系列の例

Fig.1 に機関 A において単位時間 $\Delta t=600$ s としたときの測定結果の例を示す。このタイムスケールでは Diurnal Variation が明確に認められる。なお以下においてトラフィックのモデルは(1)式によるが、記法としては

$$x_t = (Y_t + A_t) + X_{t,T} = y_t + X_{t,T},$$

$$Y_t: \text{Random Component}, A_t: \text{Anomaly}, \\ X_{t,T}: \text{Stationarily Deterministic Component} \\ = \text{Trend} + \text{Seasonality} \text{-----} \quad (3)$$

とし、 y_t は広義には異常成分がある場合も含むものとする。 $X_{t,T}$ は広義のトレンドとし、統計学でいうところのトレンドと

Seasonality (Cycle) を含む成分とする。なお Fourier 解析による Low-Pass フィルタリングを適用するに当たり

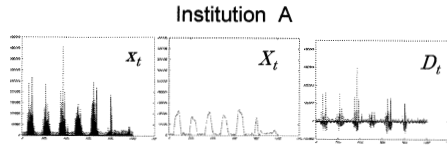


Fig.1 Exemplification of the original traffic time series x_t , its filtered result X_t , and the subtracted time series D_t as $D_t = x_t - X_t$ obtained from either an educational and research institution A or B under $\Delta t=600$ s.

トラフィック時系列: x_t

Low-Pass フィルタの Cut-Off 周波数: f_L
(FFT を適用するデータ点数 N は $N=8,192$ とし、このときの f_L は $f_L=32$)

差分時系列: $D_t = x_t - X_t$, $X_t = \text{IFFT}[\text{LPF}[X_{t,L}]]$,

$$X_{t,L} = \text{LPF}[X_t], X_t = \text{FFT}[x_t]$$

とした。 X_t は原時系列 x_t のフィルタリング結果で、(3)式の $X_{t,T}$ である。機関 A での結果は 2005/6/13 からの Mon~Sun で、Mean Packet Rate $f_p = 51.7$ pps, Mean Interval $\tau_M = 19,300$ μ s, 機関 B での結果は 2007/11/6 からの Mon~Sun で、Mean Packet Rate $f_p = 2,695$ pps, Mean Interval $\tau_M = 623$ μ s である。

3. 結果

データ処理は $\Delta t=600$ s, 10 s, 1 s に関し実施し、原時系列 x_t ならび差分時系列 D_t に対し ACF (自己相関関数)、および R/S 解析によるハーストパ

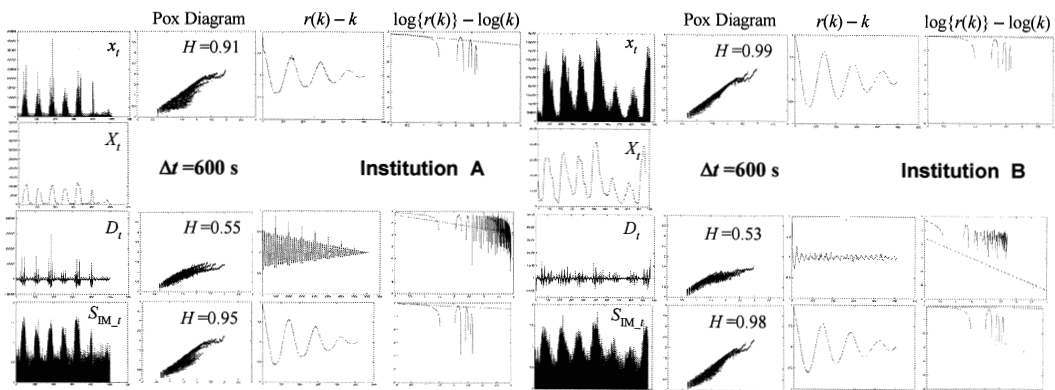


Fig.2 Exemplification of the original traffic time series x_t , its filtered result X_t , and the subtracted time series $D_t = x_t - X_t$ obtained from either an educational and research institution A or B under $\Delta t=600$ s. Also illustrated is a simulated series $S_{IM,t} = x_{t,FGN,0.5} + X_t$ superposing an uncorrelated FGN series $x_{t,FGN,0.5}$ having $H=0.5$ to the filtered series X_t .

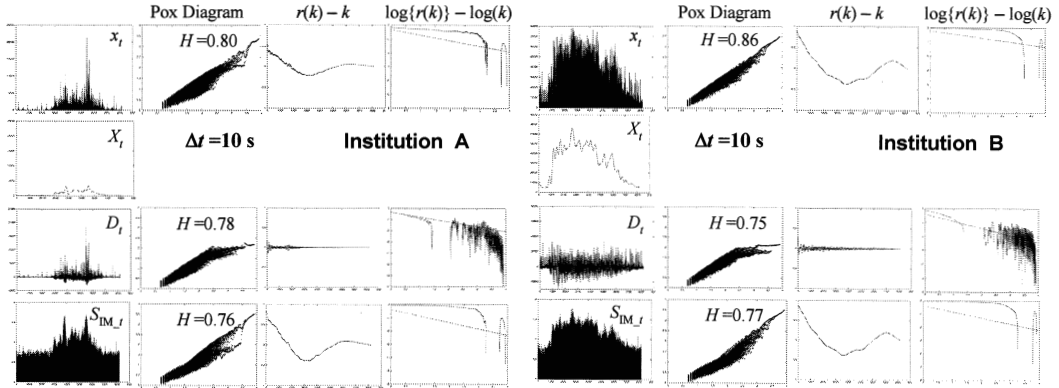


Fig.3 Exemplification of the original traffic time series x_t , its filtered result X_t , and the subtracted time series $D_t = x_t - X_t$ obtained from either an educational and research institution A or B under $\Delta t = 10$ s. Also illustrated is a simulated series $S_{IM,t} = x_{t,FGN,0.5} + X_t$ superposing an uncorrelated FGN series $x_{t,FGN,0.5}$ having $H=0.5$ to the filtered series X_t .

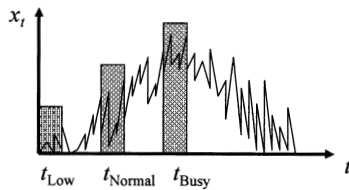


Fig.4 Explanation of the selection of Low, Normal, and Busy hour in a day which correspond to small, medium, and frequent traffic flow in a LAN.

ラメータ H と Pox Diagram を導出した。またシミュレーション系列として、抽出したトレンド成分 X_t と、無相関過程となる $H=0.5$ のときの FGN (Fractional Gaussian Noise)¹¹⁾ $x_{t,FGN,0.5}$ との重畳過程 $S_{IM,t} = x_{t,FGN,0.5} + X_t$ も生成してみた。長時間トラフィックのバケットフロー量 f_p は、機関 A で $f_p=50$ pps 程度、機関 B で $f_p=2700$ pps 程度であった。また 2 次の自己相関過程の ACF は

$$r(k) \sim H(2H-1)k^{2H-2} \quad (4)$$

となり¹²⁾ $r(k)$ の \log - \log プロットは直線となるので、結果の整理にはこの関係も利用した。

3.1 タイムスケールが大である場合の結果

Fig.2 に機関 A, B における $\Delta t=600$ s のときの結果を示す。フィルタリングされた時系列 X_t および normal-normal プロットの ACF から明らかであるが、明確な Diurnal Variation が認められる。両機関とも原時系列 x_t の H の値は $H>0.5$ 、差分時系列 D_t の H の値は $H \approx 0.5$ であるが、ACF の \log - \log プロット形には両機関において相違が認められる。

3.2 中程度タイムスケールときの結果

Fig.3 に $\Delta t=10$ s のときの結果を示す。このスケールでは 1 日の変化をカバーしていて、緩慢なトレンドの効果が現れている。両機関とも H は 0.8

前後となっているが、Pox Diagram の形状は 2 次の自己相関過程のそれとは少し異なっている¹³⁾。ここで得られた ACF の \log - \log プロットは(4)式とほぼ一致している。

3.3 タイムスケール小のときの結果

ここでは Leland et al.¹⁾ にならい、Fig.4 に示した 1 日における Low, Normal, Busy Hour というフロー量範囲での結果を Fig.5 に示す。原時系列 x_t では機関 A で $0.7 < H < 0.8$ 、機関 B で $H \approx 0.9$ 、差分時系列 D_t では機関 A で $H \approx 0.7$ 、機関 B で $H \approx 0.8$ であった。

4. 検討

TCP のストップ・ウェイト・フィードバックの機能を考慮した場合、トラフィックにおける支配的なランダム性は自己相関性であると考えられる¹⁴⁾。以下、原時系列 x_t 、差分時系列 D_t 、および無相関過程 $x_{t,FGN,0.5}$ とフィルタリングされた過程 X_t との重畳過程 $S_{IM,t} = x_{t,FGN,0.5} + X_t$ に関し、R/S 解析の結果ならびに ACF の観点からの検討を試みる。

4.1 自己相関性

4.1.1 自己相関過程と自己相関関数

コンピュータにより求められた FGN 系列¹¹⁾ と(3)式より得られる ACF の \log - \log プロットを Fig.6 に示す。Fig.6 より、実測トラフィックの H を(3)式に代入した結果と実測トラフィックの ACF の \log - \log プロットの比較は自己相関性の検討には妥当であることがわかる。

4.1.2 原時系列および差分時系列

SNMP による被観測トラフィックでは、トレンドを除去した後の過程にも自己相関性が認められるという指摘もある¹⁵⁾。これよりは小さなタイムスケールである $\Delta t = 10$ s と $\Delta t = 1$ s の場合の Fig.3 および Fig.5 において、トレンド除去後の差

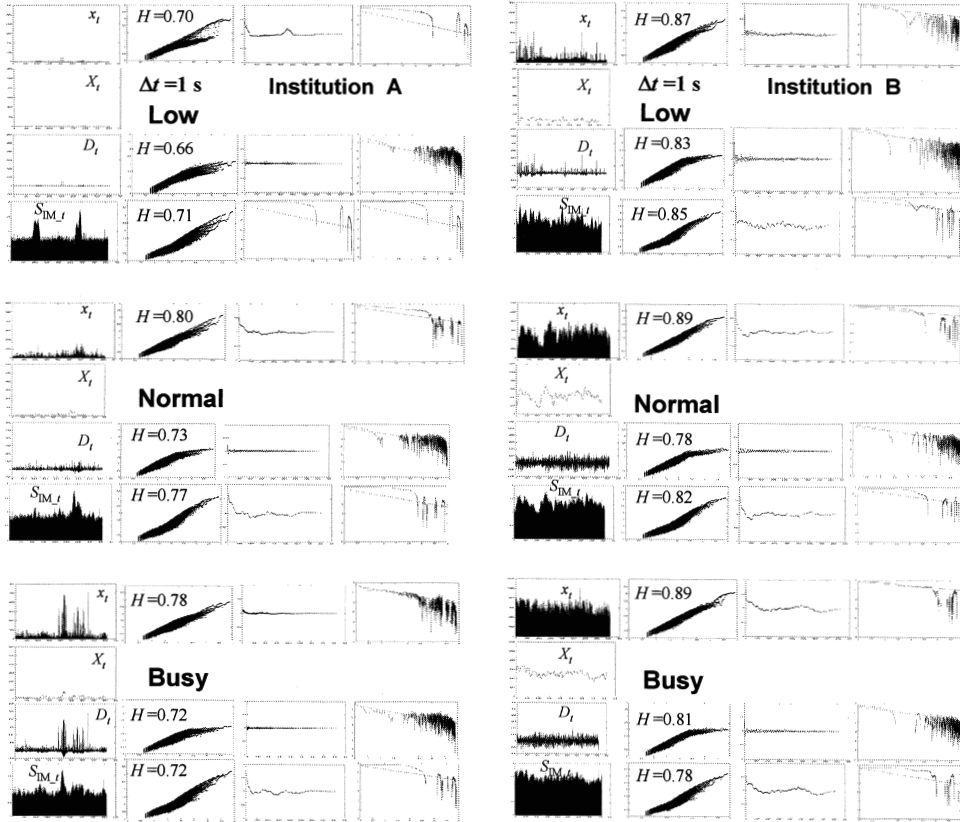


Fig.5 Inter-institutional comparison among Low, Normal, and Busy Hour of the traffic time series x_t , its filtered result X_t , the subtracted time series $D_t = x_t - X_t$, and a simulated series $S_{IM,t} = x_{t,FGN,0.5} + X_t$ under $\Delta t = 1$ s.

分時系列 D_t の H は $0.6 < H < 0.8$ 程度であり、無相関過程ではないことがわかる。

4.1.3 シミュレーション系列

シミュレーション系列 $S_{IM,t}$ は 3 節に述べたような重畳時系列である。Fig.2, Fig.3, Fig.5 に示したように、タイムスケール $\Delta t = 600, 10, 1$ s のいずれのいずれにおいても両機関における $S_{IM,t}$ は

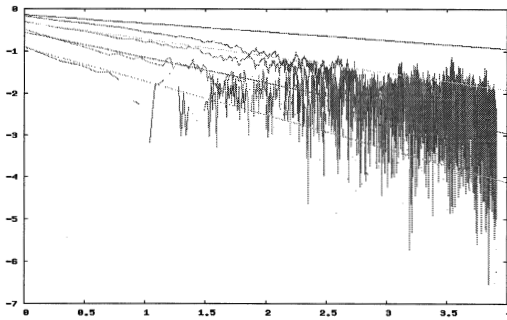


Fig.6 Drawing of the ACF of 2nd-order self-similar processes in log-log plot with respect to FGN series and theoretical expression for $H=0.6, 0.7, 0.8, 0.9$.

原時系列 x_t と相似していて、実トラフィックの自己相似性は擬似的であること暗示している。

4.2 パケット到着時間隔分布からの検討

トラフィックに Diurnal Variation がある場合にはランダム過程の非正常性の検討も必要である。パケットの到着時間間隔を τ , 当該時間帯でのパケットの平均到着率を f_p [pps], 当該時間を T_p とすると、ポアソン過程における時間間隔 $d\tau$ でのパケット到着の頻度、すなわち IATD (Interarrival Time Distribution) $dN_p(\tau)$ は

$$dN_p(\tau) = f_p^2 T_p \exp\{-f_p \tau\} d\tau \quad \text{----- (5)}$$

と表せ¹⁶⁾、頻度と時間間隔 τ の log-normal プロットは直線となる。機関 A, B の Low, Normal, Busy Hour での IATD の結果を Fig.7 に示す。Fig.7 の log-normal プロットは直線とはならず、パケット間隔は定常なポアソン過程のそれではないことがわかる。詳細なモデリングは実施していないが、ここで扱っているトラフィックは擬似自己相似性の観点からの検討も必要であると考えられる²⁾。

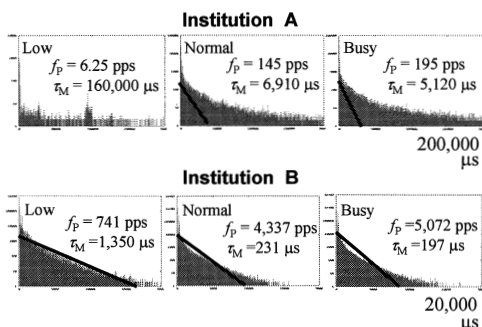


Fig.7 Interarrival Time Distribution (IATD) for either institution A, or B. Definite result is that none mode of them is identical to Poisson processes. A kind of tail distribution is imagined.

5. むすび

規模の異なる 2 箇所の教育研究機関において得られた長時間トラフィック時系列を、周波数領域における理想フィルタリングによりトレンド成分とランダム成分とに分離し、トレンド除去後のランダム成分に関し検討してみた。その結果、明確な Diurnal Variation が認められるようなタイムスケール (Aggregation Level = 600 s) でも、また緩やかなトレンドが認められるようなタイムスケール (Aggregation Level = 1, 10 s) においてもランダム成分は無相関な系列とはならないことがわかった。今後はこの知見を基にしてリアルタイムな異常検知法の提案などを行っていく予定である。

文 献

- 1) W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic", *Computer Communications Review*, Vol.23, No.4, pp.183-193, (1993).
- 2) Thomas Karagiannis, Mart Molle, Michalis Faloutsos, Andre Broido: "A nonstationary Poisson view of internet traffic", IEEE INFOCOM 2004 - The Conference on Computer Communications, Vol. 23, No. 1, pp. 1559-1570, (March 2004).
- 3) P. Barford, J. Kline, D. Plonka, and A. Ron: "A signal analysis of network traffic anomalies", Proc. of the ACM Internet Measurement Workshop (IMW) 2002, pp.71-82, (Marseille, France), (November 2002).
- 4) M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates and Y. Zhang: "Experience in Measuring Backbone Traffic Variability: Models, Metrics, Measurements and Meaning," Proceedings of the 18th International Teletraffic Congress (ITC '2003), Berlin, Germany, (September 2003).
- 5) P. Barford, J. Kline, D. Plonka, and A. Ron: "A signal analysis of network traffic anomalies", Proc. of the ACM Internet Measurement Workshop (IMW) 2002, pp.71-82, (Marseille, France), (November 2002).
- 6) Augustin Soule, Kavé Salamatian, Nina Taft: "Combining Filtering and Statistical Methods for Anomaly Detection", Proceedings of the Internet Measurement Conference on Internet Measurement Conference (IMC '05), Berkeley, CA, pp.331-344, (Oct. 2005).
- 7) Stilian Stoeva, Murad S. Taqqa, Cheolwoo Parkb and J.S. Marron: "On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of Internet traffic", *Computer Networks*, Vol.48, No. 3, pp.423-445, (June 2005).
- 8) William Aiello, Anna Gilbert, Brian Rexroad, Vyas Sekar : "Sparse Approximations for High Fidelity Compression of Network Traffic Data", Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement, Berkeley, California, USA, pp.253-266, (October 2005).
- 9) A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. Kolaczyk, and N. Taft: "Structural Analysis of Network Traffic Flows", *ACM SIGMETRICS Performance Evaluation Review*, Vol.32, No.1, pp.61-72, (June 2004).
- 10) P. J. Brockwell and R. A. Davis, *Time Series-Theory and Methods*, 2nd Ed., (Springer, New York, 1991).
- 11) Software In The Internet Traffic Archive: Available at <http://ita.ee.lbl.gov/html/contrib/fft-fgn.html>
- 12) J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, Long-Range Dependence in Variable-Bit Rate Video Traffic, *IEEE Transactions on Communications*, Vol.43, No.2/3/4, pp.1566-1579, (February/March/April 1995).
- 13) 入江若菜, 五十嵐隆治, 高橋秋典, 岩谷幸雄, 木下哲男, 上田 浩, 非定常性を付加したシミュレーショントラフィックにおけるハーストパラメータの変化, 平成 19 年度電気関係学会東北支部連合大会講演論文集, 2C-16, p113, 2007.
- 14) Sukeke Fukudaa, Hideki Takayasub, Misako Takayasu: "Self-similar Traffic Originating in the Transport Layer", *Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS'01)*, pp.55-60, SCS, Phoenix, USA, (Jan. 2001).
- 15) Matthew Roughan and Joel Gottlieb: "Large-scale Measurement and Modeling of Backbone Internet Traffic", *Proc. SPIE*, Vol.4865, pp.190-201, Boston, MA, USA, (July 2002).
- 16) 小野仁子, 田村雄介, 増田有悟, 五十嵐隆治, 高橋秋典, 岩谷幸雄, 坂田真人: プロトコルに依存したネットワークトラフィックの解析, 平成 18 年度電気関係学会東北支部連合大会講演論文集, 1G-10, p249, (2006).