

ハーストパラメータとネットワークトラフィック量変化との相関

五十嵐 隆治*, 井上 博勝**, 宮林 尚英**, 田村 雄介***, 増田 有悟***, 岩谷 幸雄****, 坂田 真人*****, 横山 博之*, 藤原 克哉*, 高橋 秋典*, 玉本 英夫*, 行松 健一*

*秋田大学工学資源学部情報工学科, **秋田大学大学院鉱山学研究科, ***秋田大学大学院工学資源学研究科
****東北大学電気通信研究所やわらかい情報システム研究センター, *****秋田大学工学資源学部電気電子工学科

ネットワークトラフィックに自己相似性があることが指摘されて以来、モデリングを含めた種々のトラフィック解析がなされている。当該時系列に厳密な意味での自己相似性が存在する場合、自己相似性はハーストパラメータ $H(0.5 < H < 1)$ で表されるが、実トラフィックはネットワークのアーキテクチャ、プロトコル、ないしは稼働状況により、常に理論的な意味での自己相似過程に従っているとは考え難い。本研究では、1年以上の長期間に渡り連続して測定したトラフィックを R/S 解析により調査してみた。その結果、 H の値はトラフィック量変化にも依存して変わることが推測された。

キーワード: ネットワークトラフィック, ハーストパラメータ, R/S 解析, コレログラム

Relationship between Hurst Parameters and Amount Variations of Network Traffic

Ryuji Igarashi*, Hiro-katsu Inoue**, Nao-hide Miya-bayashi**, Yuusuke Tamura***, Yuugo Masuda***, Yukio Iwaya*****, Masato Sakata*****, Hiroshi Yokoyama*, Katsuya Fujiwara*, Akinori Takahashi*, Hideo Tamamoto*, Ken-ichi Yukimatsu*

*Department of Computer Science and Engineering, Akita University, ** Graduate School of Mining and Engineering, Akita University, *** Graduate School of Engineering & Resource Science, Akita University, **** Flexible Information System Research Center, RIEC(Research Institute of Electrical Communication), Tohoku University, *****Department of Electric and Electronic Engineering, Akita University

The Analysis of the network traffic, including the model construction, has been practiced since a self-similarity was detected in the traffic flow. In case a stationary process obeys a 2nd order self-similar process in a strict sense, the process is characterized by a Hurst parameter $H(0.5 < H < 1)$. A real traffic which depends on the network architecture, protocol, or operation condition is hard to be imagined to be quite consistent with the theory. According to the authors' observation for more than one year, H seems to have a relationship with the temporal variation of the traffic amount.

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

1. はじめに

W. E. Leland et al.¹⁾ によりネットワークトラフィックには自己相似性が認められることが指摘されて以来、いろいろな研究が進められて来た²⁾⁻⁶⁾。自己相似性を有している過程には長期依存性があることも知られていて^{1), 2)}、自己相似性を特徴付けるパラメータはハーストパラメータ H と呼ばれている。これはナイル川の流量調査で初めてこの現象を統計的に解析した H. E. Hurst⁷⁾ にちなんで命名されたものである。

このときハーストが用いた方法が R/S 解析ないしは RAR(Rescaled Adjusted Range)統計⁸⁾と呼ばれているもので、B. Mandelbrot⁹⁾ 等により数学的な基礎が与えられた。 H を求める方法としては R/S 解析、分散-時間法、ペリオドグラム法等、種々の方法が提案されている^{1), 5)}。ただし、 H 自体が統計量であり、R/S 解析、分散-時間法、ペリオドグラム法等のグラフ的な方法^{1), 2)}では、求められた値の精度に限界があることが指摘されている。しかし著者らは、視覚的に評価しつつ H を求め得る R/S 解析は、実際のトラフィック変動の検討には依然有効な手法であると考えている¹⁰⁾。また、実際の現象は決して理論に都合良くは適合しないことを考えると、むしろ R/S 解析により得られるパラメータは実際の変動過程の何らかの性質を表す統計量であるとも解釈でき、実トラフィックの特性の検討に積極的に用い得るパラメータであるとも解釈できる。著者らはこのような考えの下で実測トラフィックに R/S 解

* 秋田大学工学資源学部情報工学科
** 秋田大学大学院鉱山学研究科
*** 秋田大学大学院工学資源学研究科
**** 東北大学電気通信研究所やわらかい情報システム研究センター
***** 秋田大学工学資源学部情報電気電子工学科
* Department of Computer Science and Engineering, Akita University
** Graduate School of Mining and Engineering, Akita University
*** Graduate School of Engineering and resource science, Akita University
**** Flexible Information System Research Center, RIEC, Tohoku University
***** Department of Electric and Electronic Engineering, Akita University

析を適用し、これにより得られるハーストパラメータを、トラフィック量の時系列的变化と対応付けて調べてみた。

実トラフィックより得られた R/S Pox Plots¹⁾は、FGN(Fractional Gaussian Noise)より生成されたデータ系列による R/S Pox Plots とは相違した形を示す場合がある。FGN は理論的な自己相似過程のモデルであり、Mandelbrot 等が積極的に用いたものである。また、実トラフィックから得た R/S Pox Plots より H を求めようとする場合、Pox Plots のどの範囲で H を求めるのかにより、得られる H の値には相違が生じる。従って実トラフィックより求められる H は、厳密には、ハーストパラメータとは呼べない。むしろ R/S 解析により求められる、トラフィックを特徴付けるひとつのパラメータである、と解釈するのが妥当と考えられる。このときトラフィック量と H はどのような関係を示すのかに関し、以下に述べて行く。

2. トラフィックの測定

2.1 測定

本研究で用いたトラフィックデータは、著者らの機関に設置してあるサーバー上で

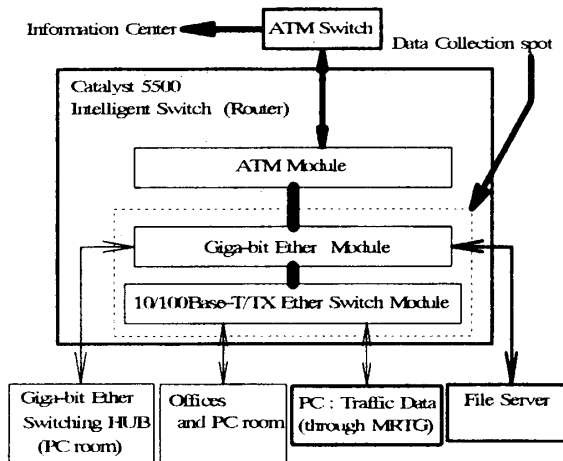


Fig.1 Schematic description of the data collection environment.

MRTG により取得したもので、データ取得点周囲のネットワークは図 1 のような構成になっている。本研究では 24 ポートから得たトラフィックの合計トラフィックを用いた。MRTG においては 5 分間の平均トラフ

ックを時系列上の 1 離散時間 t における値 X_t とし、データ取得後この X_t を順次データファイルにセーブした。トラフィック測定中であってもこのデータ X_t は随時取り出し、解析に供することができる。

現在も測定は進行中であり、測定開始時よりほぼ 1 年半が経過している。ただし上述のように離散時間 t は、その単位が 5 分であるため、 X_t は 1 日経過後で 288 点、1 ヶ月経過後で 8,640 点、1 年の経過に対して高々 105,120 点と、従来の研究で試みられている測定¹⁾⁻⁶⁾よりは時間的には粗い測定である。また図 1 からわかるように、ここで対象としているネットワークは小規模なもので、ここで測定されるトラフィックは本ネットワークに特有なものであると考えられる。

2.1 測定結果

2.1.1 トータルトラフィック

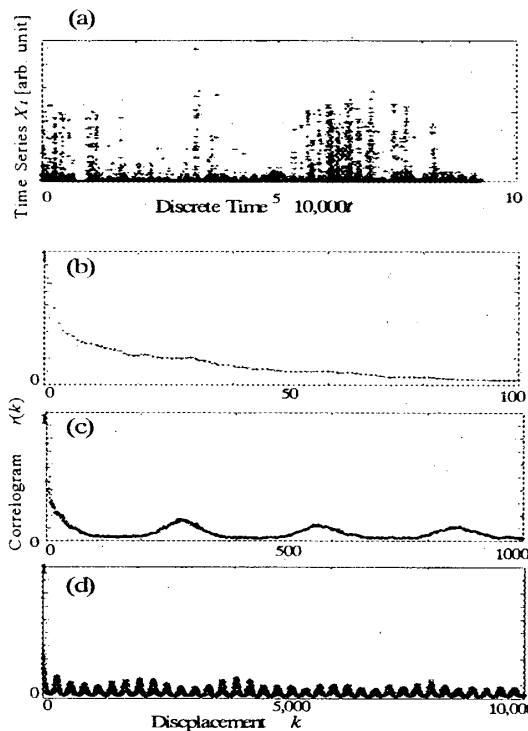


Fig.2 Obtained traffic X_t and its correlograms for various displacement ranges.

上記条件下で得られたトラフィックを時系列データ X_t として表したのが図 2-(a)で、このときの総データ数、すなわち図 4 に示した、離散時間 t の最終値 $t=N$ は 90,000 である。この X_t にはバースト性が認められる

ことがわかるが、図 2-(a)のみからでは自己相似性の明確さは未だ明言し難い。

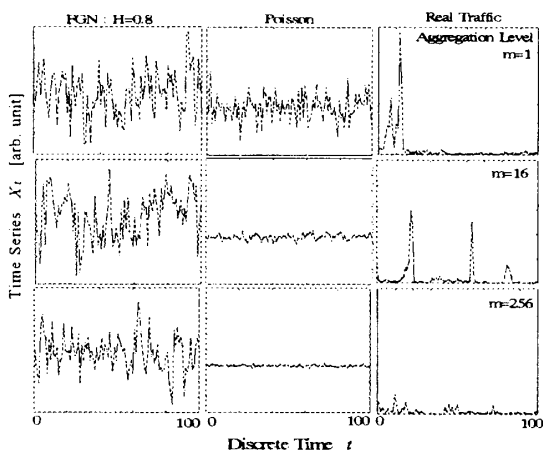


Fig.3 Investigation example of the self-similar nature between FGN, Poisson, and the observed traffic in terms of their time series X_t .

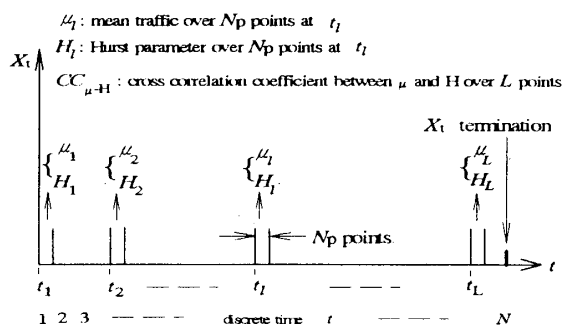


Fig.4 Illustration how to obtain H and the mean traffic μ at each spot along a time series X_t , the temporal variation of the traffic. After the derivation of μ_1 and H_1 , crosscorrelation coefficient $CC_{\mu-H}$ is obtained.

自己相似性とはスケール不変な性質，すなわちフラクタル性であるとも解釈されていて^{1), 2), 4)}，その特徴は，時間領域でのフラクタル性の表現である aggregated process^{1), 4)}を調べることで確認できる．いま aggregation level m を $m=1 \sim m=256$ としたときの $X_t^{(m)}$ を各々，FGN, Poisson, 実測トラフィックに対し描いた様子を図 3 に示す．FGN は $H=0.8$ の場合の様子である．図よりこの m の範囲では FGN, 実測トラフィックともバースト性が保存されていて，自己相似性が確認される．Poisson に関しては理論に従い，大なる m に対しバースト性

が消失していることがわかる．ただし図 3 における確認は定性的な検討であり，これより直ちに実測トラフィックが，すべての範囲において理論的な意味での自己相似性を有しているとは言えない．

2.1.2 H 導出時の R/S Pox Plots 上での計算範囲への依存

図 4 に R/S Pox Plots より H を求めるときの計算範囲の設定法を示す．なおここで示

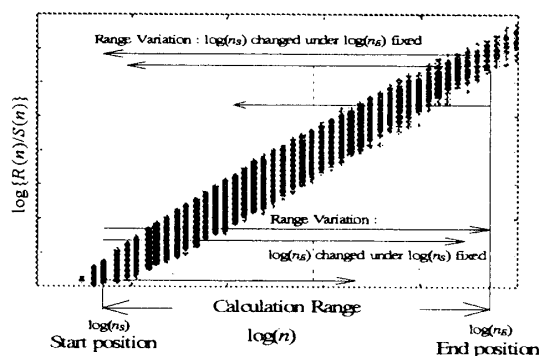


Fig.5 Illustrative explanation with regard to the calculation range in an R/S pox diagram.

した $R(n)$, $S(n)$ は各々累積範囲と標本標準偏差で，元々は Hurst により導入された統計量⁷⁾であるが後に Mandelbrot 等⁹⁾により洗練された取り扱いとなり，さらに Leland 等¹¹⁾によりネットワークトラフィックの解析に鋭意用いられた量である．Leland 等に従うと標本平均 $X(n)$ により

$$R(n)/S(n) = 1/S(n) [\max(0, W_1, W_2, \dots, W_n) - \min(0, W_1, W_2, \dots, W_n)];$$

$$W_k = (X_1 + X_2 + \dots + X_k) - kX(n),$$

$$k = 1, 2, \dots, n$$

ただし n は図 4 の記号では， $1 \leq n \leq N_p$ となる．また Hurst によると，変動を有する過程は統計量 $R(n)/S(n)$ の期待値 $E\{R(n)/S(n)\}$ が Hurst パラメータ H と n に対し $E\{R(n)/S(n)\} \cong c n^H$ となり， $1/2 < H < 1$ のとき当該過程は 2 次の自己相似過程であると言われる^{1), 2)}．

R/S Pox Plots を用いて H を導出する場合，diagram の横軸 (n または $\log(n)$ 軸) の極端に大きな部分および小さな部分は避けるべきであることが，トラフィック研究の初期においてすでに指摘されていた²⁾．しかし

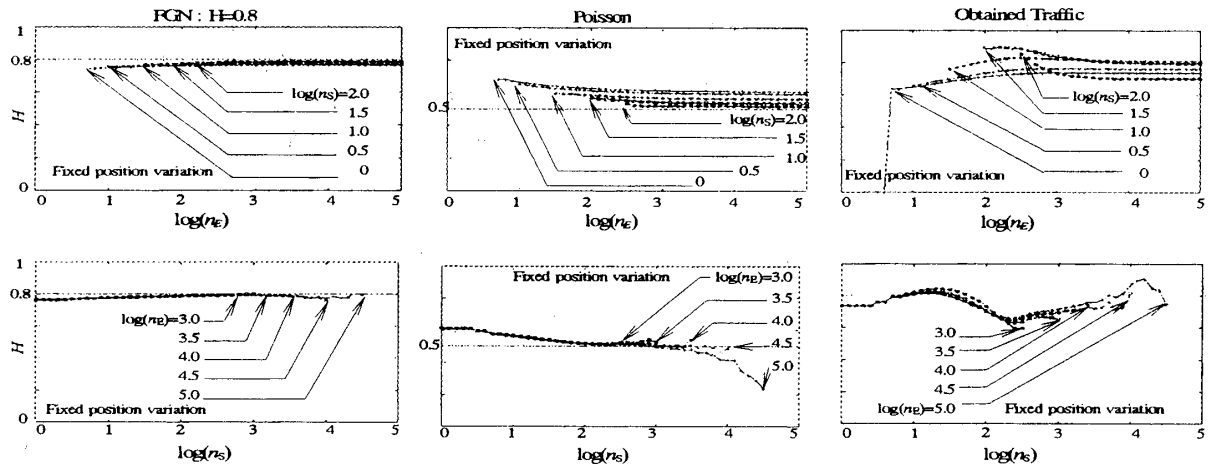


Fig.6 Discrepancies of H between diverse parameters. Upper: in case end points are changed, with start points fixed for various n_S . Lower: in case start points are changed, with end points fixed for various n_E .

どの範囲の n が妥当であるかの定量的な指摘はなく、ここに H 導出における不確定的な要素があるように思われる。

FGN($H=0.8$), ならびに Poisson 過程($H=0.5$) に対する H の変化の様子も示した。ここで用いた FGN は Software In The Internet Traffic Archive より取得したもので、FFT アルゴリズムに基づくものである¹¹⁾。Poisson ほどではないが FGN においても計算範囲に対する依存性、特に n_S に対する依存性が認められる。実測トラフィックに対してはこの傾向がさらに顕著である。

2.1.3 時系列 X_t の複数スポットより求めた H とトラフィック量 μ との相関

上述の結果を受け、図 4 に示した t 上の

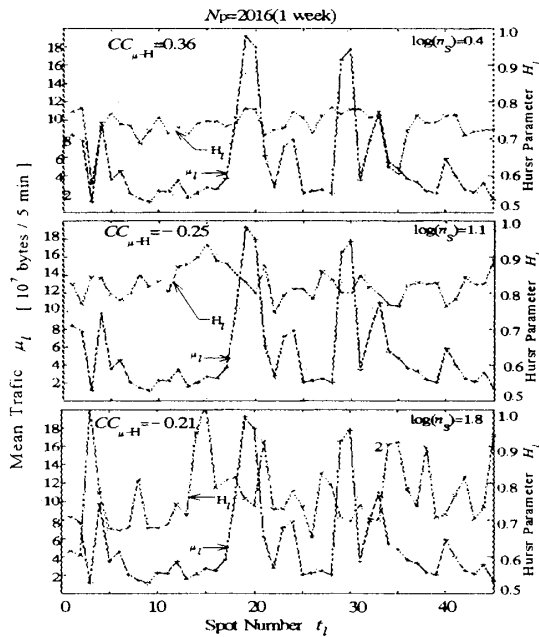


Fig.7 Correlation between μ_t and H_t along a time series X_t over $N_p=2016$ pints. $CC_{\mu-H}$ is obtained for diverse $\log(n_S)$.

R/S 解析による H の導出に際し著者等はこの不確かさを可能な限り除去する目的で、図 4 に示したような方法で H の計算範囲依存性を調べてみた。図 5 に示した計算開始点 n_S および終了点 n_E を変えて導出した H の変化の様子を図 6 に示す。ここでは比較のために、シミュレーションにより得た

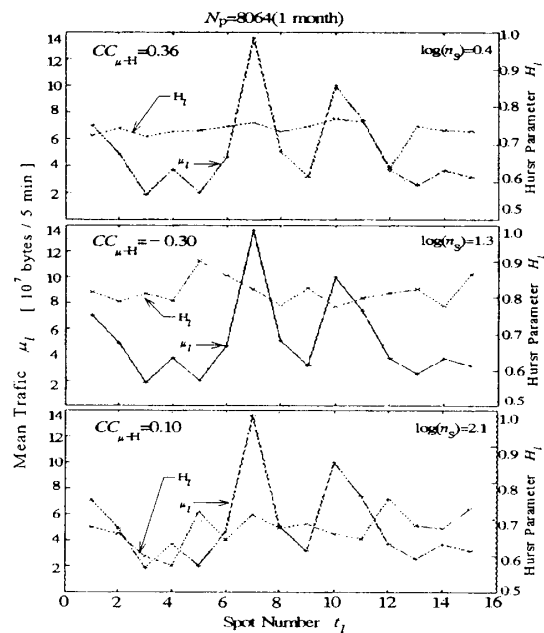


Fig.8 Correlation between μ_t and H_t along a time series X_t over $N_p=8064$ pints. $CC_{\mu-H}$ is obtained for diverse $\log(n_S)$.

複数スポット $t_l (l=1, 2, \dots, N)$ において H と μ を求めた結果を図 7 および図 8 に示す. 各 t_l において R/S 解析に必要な $R(n)/S(n)$ を求めるときの n は図 4 に示した N_p に対し, 図 7 では $0.4 \leq \log(n) \leq \log(N_p) = \log(2016) = 3.3$ とした. この $N_p=2016$ は離散時間 t の 1 点が 5 分であることより, 2016 点 = 7 days = 1 week に対応している.

2.1.2 で述べたように, R/S 解析により導出される H は計算開始点 n_s に依存するので, 図 7 においては $\log(n_s) = 0.4, 1.1, 1.8$ の 3 モードに関し H を求めた. 図 7 においては t 上での μ と H の相関関係の有無は明確には判別できない. 故に図 7 には, 同一の N_p に対し求めた μ と H の相互相関係数 $CC_{\mu-H}$ も併記した. これより $CC_{\mu-H}$ は $\log(n_s)$ の値に依存していて, 本測定に関してはその絶対値は $\log(n_s)$ の逆順に一致し, $\{\log(n_s): 0.4 < 1.1 < 1.8\}$ に対し $\{|CC_{\mu-H}|: 0.36 > 0.25 > 0.21\}$ であった.

図 8 も図 7 と同様の処理で, ここでは $0.4 \leq \log(n) \leq \log(N_p) = \log(8064) = 3.9$ とし, $\log(n_s) = 0.4, 1.3, 2.1$ の 3 モードに関し H を求めてみた. この $N_p=8064$ は上と同様に, 8064 点 = 30 days = 1 month に対応している. ここでも $CC_{\mu-H}$ を求めてみたが, その絶対値は $\log(n_s)$ の逆順に一致し, $\{\log(n_s): 0.4 < 1.3 < 2.1\}$ に対し $\{|CC_{\mu-H}|: 0.36 > 0.30 > 0.10\}$ であった.

3. 結果の検討

3.1 H 導出時のデータ幅 N_p との関係

理論的な自己相似過程は定常確率過程ゆえ, R/S 解析により H を求めようとする場合, それに供するデータの点の数, すなわちデータ幅 N_p は大であるほど望ましい. しかしネットワークのすべての稼動期間において理論的な自己相似過程に従うという保障のない実トラフィックにおいては³⁾, N_p をどのような値にするのかはネットワーク特性を把握する上で興味深い視点である. ただしトラフィックに長期依存性が存在することを考慮すると, N_p を全く任意に設定するのは不適當である. N_p の設定の目安として著者等は, X_t の長期依存性が X_t の自己相関関数 (実験値においてはコレログラム)

に現れることに注目した. 実際に $0 \leq r(k) \leq 1$ で規格化したコレログラム $r(k)$ は, X_t の期待値を $\mu_X = E\{X_t\}$ と表すと

$$r(k) = E\{(X_t - \mu_X)(X_{t+k} - \mu_X)\} / E\{(X_t - \mu_X)^2\}$$

と定義できる. これにより求めた実トラフィック X_t の $r(k)$ を図 2-(b), (c), (d) に示す. 長期依存性を有し且つ自己相似となる理論過程である FGN のシミュレーションにより得た X_t の, $H = 0.6, 0.7, 0.8, 0.9$ に対するコレログラムを図 9 に示したが, $H = 0.8$ ないしは $H = 0.9$ に対する $r(k)$ は, その漸近性^{1), 2)}より, $k = 1000$ でも $r(k) \neq 0$ であることが明白である. ただし, 多少恣意的ではあるが, $H < 0.8$ に対し $k > 1000$ では $r(k) \cong 0$ と見なしてみると, $k > 1000$ であるような k の値を N_p とすることは, 粗い検討では妥当と考えられる. ここで改めて図 2-(c) を見てみると, 本測定で得られた X_t に対し $N_p = 2016$, または $N_p = 8064$ と設定したことは妥当であると判断した.

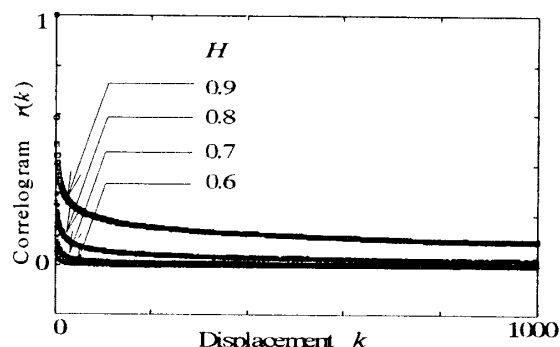


Fig.9 Correlogram example derived from FGN with regard to various values of H

なお図 2-(c) ないしは (d) に示されているような, $r(k)$ が周期性を示す長期依存性に関しては McLeod 等が Gota 川の流量変動の解析に際し議論した例がある⁸⁾

3.2 時系列上での H とトラフィック量 μ との相関

本測定トラフィックに関しては H とトラフィック量 μ との相関係数は, 十分大きいとは言えない. H と μ との相関が大きい, または小さいときの H はトラフィックの性質に関しどのような意味を有しているのだろうか.

$1/2 < H < 1$ は長期依存性の存在を示唆し

ているが、我々が得たトラフィックに対する H は $H \approx 0.8$ であり、 H の値のみからは Hurst が指摘したような自然現象の変動に相似しているようである。現象として長期依存性があるということはアフターエフェクトが存在しているということであり、図 4 の N_p 毎に区切ったときの H の変化はその影響を表していると考えられる。この意味からは H と μ との間に大きな相関が存在してもいいように思われるが、今回の測定値に関しては顕著な相関があるとの明言は困難である。ただし 2.1.2 に述べたように R/S pox plots において H 導出のための計算範囲を変えると H と μ との相互相関係数 $CC_{\mu-H}$ は変化した。この事実は、 H をトラフィックの評価に用い得る可能性を示唆していると考えられる。

4. まとめ

現在、今回の結果のみより H によるトラフィック特性評価の妥当性を結論付けることは多少困難であり、今後、より詳細な検討が必要である。ただしネットワークのアーキテクチャないしはプロトコルとトラフィックの混雑具合に関連するアフターエフェクトが、時系列としての H の値に顕現するのであれば、ハーストパラメータ H はネットワーク特性の有用な評価指標になり得ると考えられる。

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) J. Beran, R. Sherman, M. S. Taqqu, 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).
- 3) M. Takayasu, H. Takayasu, and K. Fukuda: Dynamic Phase Transition Observed in Internet Traffic Flow, *Physica A*, vol.277, pp.248-255, (2000).
- 4) D. Chakraborty, A. Ashir, T. Suganuma, G. Mansfield Keeni, T. K. Roy, and N. Shiratori, Self-similar and Fractal Nature of Internet Traffic, *International Journal of Network Management*, Vol.14, pp.119-129, (2004).
- 5) M. S. Taqqu, V. Teverovsky, and W. Willinger; "Estimators for long-range dependence: an empirical study", *Fractals*, vol.3, No.4, pp.785-798,(1995).
- 6) W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson; Self-Similarity in High-Speed Packet Traffic; Analysis and Modeling of Ethernet Traffic Measurements, *Statistical Science*, Vol.10, No.1, pp.67-85 (1995).
- 7) H.E. Hurst: Long-Term Storage Capacity of Reservoirs, *Trans. Amer. Soc. Civil Eng.*, Vol.116, pp.770-799, (1951).
- 8) A. I. McLeod and K. W. Hipel: Preservation of the Rescaled Adjusted Range; 1. A Reassessment of the Hurst Phenomenon, *Water Resour. Res.*, vol.14, No.3, pp.491-508, (1978).
- 9) B. B. Mandelbrot, and J. R. Wallis: Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence, *Water Resour. Res.*, vol.5, pp.967-988, (1969).
- 10) 田村雄介, 宮林尚英, 五十嵐隆治, 高橋秋典, 横山博之, 玉本英夫, 岩谷幸雄: ハーストパラメータによるネットワークトラフィックの評価, 電気学会電子・情報・システム部門大会講演論文集, GS9-7, pp.867-873, (2003).
- 11) Software In The Internet Traffic Archive: <http://ita.ec.lbl.gov/html/contrib/fit-fgn.html>