

インターネット需要の長時間相関：電子メールの場合

松原義継, 日永田泰啓, 只木進一

佐賀大学 総合情報基盤センター

概要

インターネットのデータ流量にはベキ則揺らぎがあることが知られている．その起源のひとつの候補であるインターネット需要の長時間相関を考える．例として組織内電子メールの送信需要の長時間相関を観測した．送信リクエスト回数及び送信データ量について，一日周期及び一週間周期を除いたデータに対して DFA (Detrended Fluctuation Analysis) を実行し，ベキ則揺らぎを見つけた．

Long-Term Correlation in Internet Demand: Case of E-mail Sending.

Yoshitsugu Matsubara, Yasuhiro Hieida, Shin-i-chi Tadaki

Computer and Network Center Saga University

Abstract

We consider about the long-term correlation of the Internet demand which is one candidate of power-law fluctuation in data flow. We observed the long-term correlation of the demand for e-mail sending in our organization as an example. By applying DFA (Detrended Fluctuation Analysis) to the number of e-mail sending requests and the amount of sent data, in which we extract the average daily and weekly trends, the power-law fluctuations are found.

1 はじめに

インターネットを始めとしたネットワークの構造とその上の動的現象について関心が寄せられている．特に，スケールフリー性等の構造の性質，ベキ則揺らぎ等の動的性質が共通に現れることが興味深い．

インターネットは，その構造が AS (Autonomous Systems) レベル等でスケールフリーであることが指摘されている [1]. また，RTT (Round Trip Time) に現れるベキ則揺らぎ [2] や流量に現れるベキ則揺らぎ [3]，さらに負荷に現れるベキ則揺らぎ [4, 5] 等が論じられている．しかし，スケールフリーである構造と動的性質に現れるベキ則との関係は明確ではない．たとえば，渋滞転移への臨界現象としてベキ則揺らぎを説明する試みもある [6, 7, 8]．

インターネットを構成するハードウェア及びソフ

トウェアには，様々な待ち行列機構や排他制御機構があり，それらがベキ則揺らぎを生成する可能性はある．しかし，いずれも特徴的時間が短く，数ヶ月に及ぶ相関を生成すると考えることには困難がある．インターネットのデータ流量に現れるベキ則揺らぎの他の起源のひとつとして，インターネット需要のベキ則揺らぎの可能性もある．

インターネットのデータ流量は様々なサービスに対応した要素で構成されている．全てのサービスの需要を調査することは困難であることから，その中から計測しやすいサービスの一つである電子メールの送信需要を取り上げる．電子メールは，インターネット上における主要なコミュニケーション手段の一つである．その量は統計的に十分であり，メールサーバのログファイルから容易に計測できる．なお

かつ、ネットワークの構造の影響の無いものとして、組織内のメールサービスを対象とする。組織内の電子メールサービスでは、通常、組織の主たるメールサーバが設置され、ほとんどのユーザはそのサーバへSMTP (Simple Mail Transfer Protocol) を用いて電子メールを送信する。電子メールの送信元端末およびメールサーバの両方が組織内にあれば、電子メールは組織外のネットワークを介して送信されることはない。組織内ネットワークの構造は、組織外ネットワークも含めたネットワーク全体と比較して極めて単純であり、送信時の遅延もほとんどない。つまり、今回の対象における時系列は組織外ネットワークの構造の影響をほとんど受けていない。

本研究では、佐賀大学の全教職員向けの送信用メールサーバのデータ (送信リクエスト回数および送信データ量) の時系列を用いる。これら2つの時系列を用いて、どちらにベキ則揺らぎがあるのかを明らかにする。その結果、両時系列には、長時間にわたるベキ則揺らぎが存在することを示す。

2 対象データ

佐賀大学は、5学部からなり、約1,700名の教職員を有する総合大学である。これらの教職員のほとんどが利用する送信用メールサーバの2008年5月9日から9月30日でのデータを用いることにする。データは1時間単位で集計し、その全時間は3,480時間である。

この間に学内からの送信リクエストは505,217回であった。この件数の中には、メールサイズの制限(10MB)を超えたために送信できなかったものも含まれている。これは、送信需要の性質を調査することが目的であるためである。また、毎週水曜日午後2,3時間行われる定期保守や2008年8月22日に約4時間行われた計画停電のために送信用メールサーバを停止している期間も含んでいる。

送信データ量 u_Q の時系列データの一部を図1に示す。この図における横軸は2008年6月1日00:00から2週間分である。この図からは、1日周期でその量が周期的に変化していることを読み取ることができる。なおかつ、減衰傾向も読み取れるが、これは今回の時系列データ全体の傾向ではないことに注意しておく。送信リクエスト回数 u_N の時系列にも同様の周期的挙動が現れる。

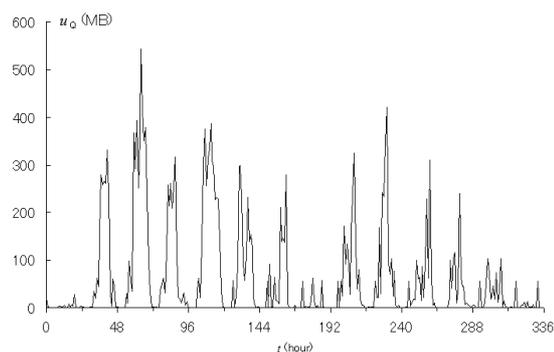


図1: 送信データ量 u_Q のサンプル

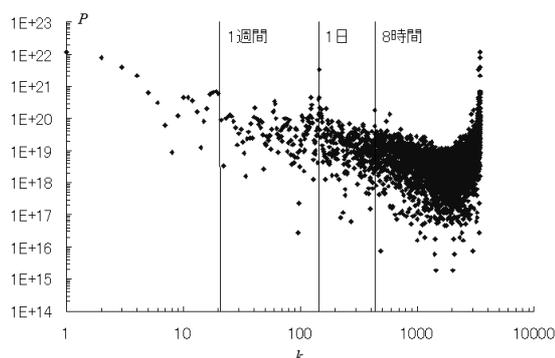


図2: 送信データ量 u_Q のパワースペクトル。3本の縦線は、それぞれ周期1週間、1日、8時間に対応する。

3 パワースペクトルによる分析

一般的な時系列解析の手法であるパワースペクトルによる分析結果を示す。全時間 T 内の時間 t ($0 \leq t < T$) における時系列データ $u(t)$ に対し、そのパワースペクトルを以下のように定義する。

$$P(k) = \left| \sum_{t=0}^{T-1} u(t) \exp\left(\frac{-2\pi k t}{T} i\right) \right|^2 \quad (1)$$

図2に送信データ量の、図3に送信リクエスト回数のパワースペクトルを示す。1日と1週間のピークを見ることができる。全体としてベキ則のような傾向があるが、指数を特定することは困難である。

4 DFA による分析

パワースペクトルではベキ指数の特定が困難な理由は、非定常データを定常波の集合で捉えることが難しいためである。そこで、非定常な時系列データの長期相関分析に用いられるDFA(Detrended Fluctuation

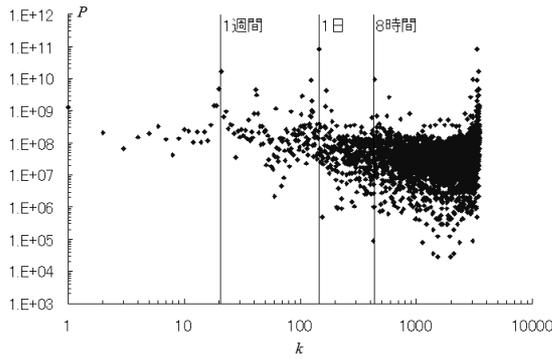


図 3: 送信リクエスト回数 u_N のパワースペクトル．3本の縦線は、それぞれ周期1週間、1日、8時間に対応する．

Analysis) により分析する [9, 10] ．

DFA による分析方法は、始めに元の時系列データ $u(t)$ に対して、その全体の平均値 $\langle u \rangle = T^{-1} \sum_{t=0}^{T-1} u(t)$ からの偏差和であるプロファイル $y(t)$ を作る．

$$y(t) = \sum_{i=0}^t [u(i) - \langle u \rangle] \quad (2)$$

そのプロファイル $y(t)$ を長さ l の区間に分割し、 n 番目の区間におけるトレンド $\tilde{y}_n(t)$ を最小二乗法により求める．プロファイル $y(t)$ のトレンドからのズレ

$$y_i(t) = y(t) - \tilde{y}_n(t), \quad nl \leq t < (n+1)l \quad (3)$$

から標準偏差 $F(l)$ を求める．

$$F(l) = \sqrt{\frac{1}{T} \sum_{t=0}^{T-1} y_i^2(t)} \quad (4)$$

$F(l)$ が区間長 l に対して

$$F(l) \sim l^\alpha \quad (5)$$

と振舞う場合は、パワースペクトルにベキ則揺らぎがあることに対応している．

$$P(k) \sim k^{-\beta}, \quad \beta = 2\alpha - 1 \quad (6)$$

前述のパワースペクトルにおいて、1日周期の成分が強く長期相関の分析に影響を及ぼすことから [11]、データ $u(t)$ から一日平均 $\tilde{u}_{\text{daily}}(\tau_d)$ を差し引いた $u'(t)$ を基に DFA 分析を行う．

$$\tilde{u}_{\text{daily}}(\tau_d) = \frac{1}{D} \sum_{d=0}^{D-1} u(d \times T_{\text{day}} + \tau_d) \quad (7)$$

$$u'(t) = u(t) - \tilde{u}_{\text{daily}}(t \bmod T_{\text{day}}) \quad (8)$$

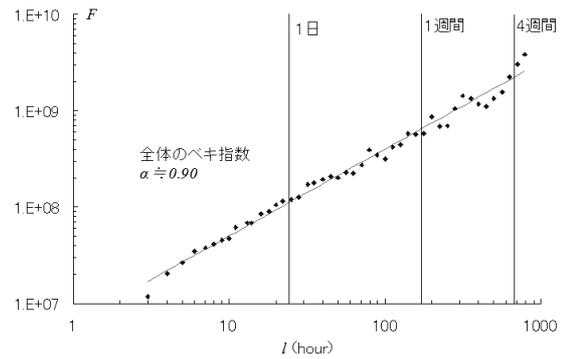


図 4: u_Q から1日平均を差し引いた送信データ量 u'_Q の DFA 分析結果

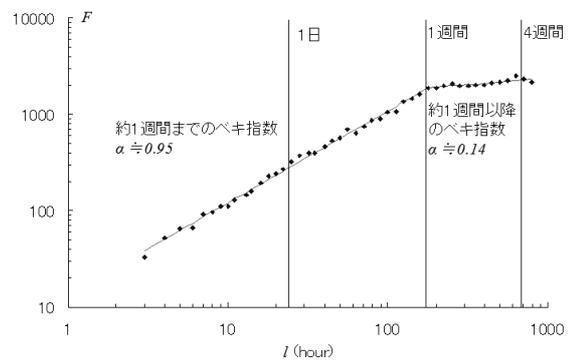


図 5: u_N から1日平均を差し引いた送信リクエスト回数 u'_N の DFA 分析結果

ここで、 $0 \leq \tau_d < T_{\text{day}} = 24$ 時間、 $D (= T/24)$ は観測日数である．

一日平均を差し引いた送信データ量に対する DFA の結果を図 4 に示す．数時間から数週にかけて、ほぼ直線となっており、この範囲におけるベキ指数は $\alpha \approx 0.90$ である．対応するパワースペクトルのベキ指数は約 0.89 となる．つまり、メールの送信データ量には、長時間にわたってベキ則揺らぎが存在することがわかる．組織内の電子メール送信データ量であるため、ネットワークの構造がベキ則揺らぎの起源と考えることは困難であり、利用者の行動にベキ則揺らぎがあると考えるのが自然である．

一日平均を差し引いた送信リクエスト回数に対する DFA の結果を図 5 に示す．数時間から一週間までは直線となっており、それより長い時間領域ではほぼ平らとなっている．ほぼ直線の部分での傾きは $\alpha \approx 0.95$ であり、送信データ量と大きな違いはない．

図 1 に示したように、送信データ量には一日周期

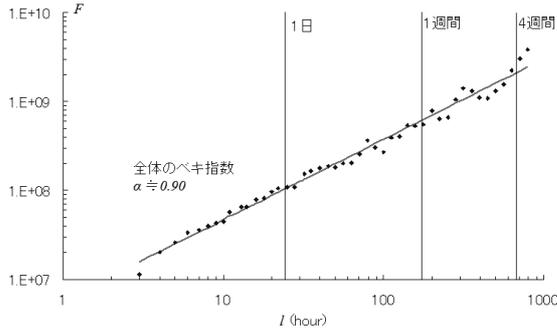


図 6: u'_Q から 1 週間平均を差し引いた送信データ量 u''_Q の DFA 分析結果

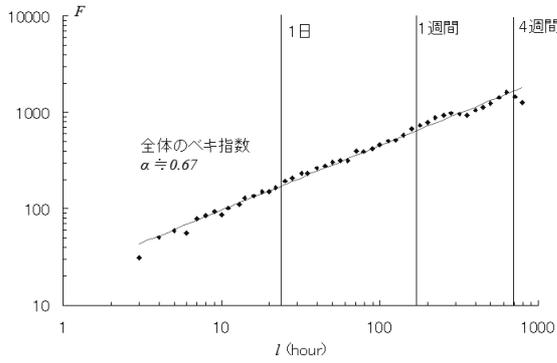


図 7: u'_N から 1 週間平均を差し引いた送信リクエスト回数 u''_N の DFA 分析結果

他に一週間の明らかな周期性がある．送信リクエスト回数にも同様の周期性が現れる．一日平均と同様に一週間平均を取り除き，DFA 分析を行うこととする．

$$\tilde{u}_{\text{weekly}}(\tau_w) = \frac{1}{W} \sum_{w=0}^{W-1} u'(w \times T_{\text{week}} + \tau_w) \quad (9)$$

$$u''(t) = u'(t) - \tilde{u}_{\text{weekly}}(t \bmod T_{\text{week}}) \quad (10)$$

ここで， $0 \leq \tau_w < T_{\text{week}} = 7 \times 24$ 時間， W は観測週数である．

送信データ量について見ると，一日平均に加えて一週間平均を差し引いても，大きな変化は見られない(図 6)．一方，送信リクエスト回数では，一週間平均を引くことで直線部分がほぼ全時間領域に広がったように見える(図 7)．しかし，同時に傾きも小さくなっており，ランダム ($\alpha = 0.5$) に近づいている．

5 まとめ

電子メール送信における送信リクエスト回数および送信データ量において，ベキ則に従うゆらぎがあることが DFA 分析により見つかった．組織内の電子メール送信の分析であるため，ネットワークの構造が原因と考えるよりも，利用者の行動にベキ則があると考えられる．

現在使用しているデータは 4 週間である．それ以上長い領域での挙動の解析は今後の課題である．

参考文献

- [1] M. Faloutsos, M. P. Faloutsos, and C. Faloutsos: ACM SIGCOMM Comp. Commun. Rev. **29** (1999) 251.
- [2] M. Takayasu, H. Takayasu, and T. Sato: Physica A **233** (1996) 824.
- [3] S. Tadaki: J. Phys. Soc. Jpn. **76** (2007) 044001.
- [4] M. Takayasu, A. Tretyakov, K. Fukuda, and H. Takayasu: in *Traffic and Granular Flow '97*, ed. D. E. Wolf (Springer, Berlin, 1998) P. 57.
- [5] K. Fukuda, H. Takayasu, and M. Takayasu: Physica A **287** (2000) 289.
- [6] M. Takayasu and H. Takayasu: Fractals **1** (1993) 916.
- [7] S. Yukawa and M. Kikuchi: J. Phys. Soc. Jpn. **65** (1996) 916.
- [8] S. Tadaki, M. Kikuchi, Y. Sugiyama, and S. Yukawa: J. Phys. Soc. Jpn. **68** (1999) 3110.
- [9] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger: Phys. Rev. E **49** (1994) 1685.
- [10] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger: Chaos. **49** (1995) 82.
- [11] S. Tadaki, M. Kikuchi, A. Nakayama, K. Nishinari, A. Shibata, Y. Sugiyama, and S. Yukawa: J. Phys. Soc. Jpn. **75** (2006) 034002.