高品質インターネット　6－2
(2003. 2. 6)

インターネットアクセスにおけるHTTP通信リクエスト発生及びドキュメントサイズ分布について

上村　正行†　細谷　謙介†　篠　秀明††　八名　和夫†††

† 法政大学工学部　〒184-8584 東京都小金井市梶野町 3-7-2
†† 三洋電機株式会社技術開発本部　〒503-0195 岐阜県安八郡安八町大森 180
††† 法政大学アメリカ研究所　米国カリフォルニア州バーリンゲーム市 504
E-mail: †{kamimura,hosoya}@bme.ei.hosei.ac.jp, ††shino@k.hosei.ac.jp, †††yana@huric.org

あらまし　高品質でコストパフォーマンスに優れたネットワークを構築するためには，トラヒックを構成する要素の統計的性質を把握することが必要不可欠である．そこで本論文では，現在多くの情報ネットワークにおいてトラヒックの主流を占めるHTTPトラヒックに着目し，その構成要素であるHTTP通信リクエストとドキュメントサイズ分布に関する統計的解析法について述べる．さらに，この提案手法に基づいて法政大学小金井キャンパスネットワークにおけるHTTPトラヒックの構成要素に対する統計解析結果を示す．本論文で検証した通信リクエスト発生とドキュメントサイズ分布に関する統計的性質は，ネットワークトラヒックの構成的モデリングを詳細に行ううえで重要であり，トラヒックシミュレーションへの応用の他，推定される統計量からネットワークを構成する要素パラメタを推定する逆問題に対する解法を与えることも期待される．

キーワード　インターネットアクセス，HTTP通信リクエスト，時間尺度変換，ポアソン性検定，ドキュメントサイズ，混合確率分布関数

HTTP Communication Request Occurrences and Document Size Distribution of the Internet Access Network Traffic

Masayuki KAMIMURA†, Kensuke HOSOYA†, Hideaki SHINO††, and Kazuo YANA†††

† Faculty of Engineering, Hosei University　Kajino 3-7-2, Koganei-shi, Tokyo, 105-0123 Japan
†† Technical Development Headquarters, Sanyo Corporation　Omori 180, Ampachi-cho, Ampachio-gun, 565-0456 Japan
††† Hosei University Research Institute　California 800 Airport Blvd. Suite 504 Burlingame, CA 94010
E-mail: †{kamimura,hosoya}@bme.ei.hosei.ac.jp, ††shino@k.hosei.ac.jp, †††yana@huric.org

Abstract　To build an excellent network by the high quality in the cost performance, it is indispensable to grasp the statistical character of the element which constitutes the traffic. Then, this paper describes the statistical analyzing method about the HTTP communication request and document size distribution which are the composition element, paying attention to the HTTP traffic which occupies the mainstream of traffic in many present information network. Furthermore, based on this proposal technique, this paper shows the statistics analysis result of the composition element of the HTTP traffic in the Hosei University Koganei campus network. The statistical character about communication request occurrences and the document size distribution which were verified in this paper is important when performing the composition-modeling of network traffic in detail.

Key words　Internet Access, HTTP Communication Request, Time Scale Conversion, Poisson Hypothesis, Document Size, Mixture of Probability Distribution

1. まえがき

近年, 国内のインターネットユーザ数の急増に伴い，インターネットが本格的な通信インフラとして普及するようになった．特に大学等教育機関においてトラヒックのほとんどがインターネットからLAN内に向けたWWW閲覧のためのHTTPトラヒックとなっている状況下で，LAN設計を有効的，効率的に行うためには，WWWアクセスによるHTTPトラヒックの統計的性質を理解した上での統計的なモデル化が重要である．

LANトラヒックの相関係数やスペクトル構造に関する研究において，LANトラヒックが自己相似性を有する長時間相関特性を持つことが明らかにされ[1]～[6]，自己相似性をもったトラヒック時系列の網設計の検討もなされている[7], [8]．これら相関構造に対するモデル化は，唯一の情報源としての観測トラヒックデータの統計解析により，トラヒック時系列としての統計的性質を明らかにするもので，現象的モデルと位置付けることができる．

一方，あるクライアントから発生するユーザリクエスト及びそれに付随して発生する2次リクエスト，各リクエストによりダウンロードされるドキュメントが多数重畳し総合的なトラヒックを形成していると考えるならば，構成的な立場でのモデル化が可能である．この場合，トラヒックを構成する要素であるリクエストと通信対象となるドキュメントサイズの統計的性質を明らかにすることが重要である．このようなトラヒックの構成要素を考慮したモデル化は，トラヒック変動の振る舞いをより深く理解するうえで重要であり，最近，宮原ら[9]によって端緒が開かれ，篠ら[10]によって構成的モデリングの第一歩が踏み出されたといえる．

そこで本稿では，構成的モデリングを念頭に置き，トラヒックを構成する要素のうち特に重要であるユーザによるHTTPリクエスト発生とドキュメントサイズの統計的性質の解析法とその結果について論ずる．まず第2章では，HTTP通信リクエスト発生に関して，短区間のリクエスト発生がポアソン過程に従うと仮定し，リクエスト発生の非定常性を考慮した上でポアソン性検定を行うことで，その仮定を確かめている．第3章では，ドキュメントサイズの分布について，複数の確率分布関数を複合的に用いることで，ドキュメントサイズ分布を詳細に表せることを示している．

2. HTTP通信リクエストに関する統計解析

一般にスパースな点過程の独立な重ねはポアソン過程に収束し[11]，多くのユーザリクエストが重畳して得られるLAN上の通信トラヒックは1次近似的にポアソン過程に従うと考えられる．[10]ではリクエスト発生の非定常生を考慮し，リクエスト生起密度関数にパラメトリックなモデルとしてフーリエ級数を用いて長時間(1日)のリクエストについてポアソン性検定を行い，ユーザリクエストが定常ポアソン過程ではないことを否定できないという結論を得ている．しかし，リクエスト発生の極めて低い時間帯に関してはポアソン性が棄却されてしまい，1日という長時間リクエストに関するフーリエ級数を用いた生起密度関数のモデル化に課題を残した．

そこで本章では，短時間(1時間)のリクエストに関して生起密度関数に指数分布を用いたモデルを採用し，非定常性を考慮した上でポアソン性検定を行い，短区間リクエストの発生がポアソン性を否定できないとの仮説を検証するための解析手法を述べ，次いでその結果を示す．なお，本稿で用いるユーザリクエストデータ及びドキュメントサイズデータは，法政大学小金井キャンパスに設置されているProxyサーバ上で稼働するsquidより得られるアクセスログから，各々必要部分を抽出して作成したものである．実測リクエストデータの加工法は[10]を参照されたい．

2.1 解析手法

2.1.1 時間尺度変換(Time Scale Conversion:TSC)

事象生起密度関数を$\lambda(t)$とする非定常ポアソン過程に従う事象生起時刻列を$\{t_1, ..., t_N\}$とすれば，一般に式(1)に示す時間尺度変換により事象生起密度1の定常ポアソン過程に従う事象生起時刻列$\{u_1, ..., u_N\}$へ変換できる[11]．

$$u_k \equiv \Lambda(t_k) \equiv \int_0^{t_k} \lambda(\sigma)d\sigma \quad (k = 1, ..., N) \tag{1}$$

よって帰無仮説H_0を「通信リクエスト生起時刻列は事象生起密度関数を$\lambda(t)$とするポアソン過程に従う」としたときの統計的検定は，事象生起時刻列の時間尺度変換後における定常ポアソン性の検定に帰着される．つまり$\lambda(t)$の適切なパラメトリックなモデル$\lambda(t;\theta)$が想定可能であれば，観測データに基づく最ゆうパラメタ$\hat{\theta}$をモデルに代入した$\lambda(t;\hat{\theta})$による時間尺度変換により現実的な検定を構成することができる[12]．

2.1.2 指数分布のパラメタ推定法

さて，定常ポアソン過程における生起事象間隔は指数分布に従うことから，パラメトリックなモデルとして式(2)に示す指数分布を考える．

$$\lambda(t;\theta) = \theta_1 \cdot e^{\theta_2 t} \quad (\theta_1 > 0) \tag{2}$$

事象生起密度関数$\lambda(t)$を式(2)とするポアソン過程の対数ゆう度関数$\mathcal{L}(\theta)$は，

$$\mathcal{L}(\theta) = \frac{\theta_1}{\theta_2}(1 - e^{\theta_2 T}) + N \ln \theta_1 + \theta_2 \sum_{i=1}^{N} t_i \tag{3}$$

であり，$\mathcal{L}(\theta)$を最大とするパラメタが最適なモデルを与えることになる．パラメタθ_1, θ_2の最ゆう推定値$\hat{\theta}_1, \hat{\theta}_2$は，式(3)の対数ゆう度方程式に関して，$\frac{\partial L(\theta)}{\partial \theta_1} = 0$, $\frac{\partial L(\theta)}{\partial \theta_2} = 0$と置き，次式，

$$\sum_{i=1}^{N} t_i + N\left(\frac{1}{\theta_2} - \frac{T e^{\theta_2 T}}{e^{\theta_2 T} - 1}\right) = 0 \tag{4}$$

$$\theta_1 = N \frac{\theta_2}{e^{\theta_2 T} - 1} \tag{5}$$

を連立させて解くことで得られる[13], [14]．

このようにして，実測データから式(2)で提案した指数分布モデルに従う事象生起密度関数の最ゆうパラメタ推定値を求め，時間尺度変換を施し，定常ポアソン性検定を行うことにする．

表1 時間尺度変換前におけるポアソン性検定結果
Table 1 Results of test for stationaly Poisson hypothesis before TSC.

インターバル χ^2 適合度	9.068×10^1
5% 有意水準	4.380×10^1
dispersion 検定 d	4.316×10^2
95% 信頼区間	$(7.336 \times 10^1, 1.284 \times 10^2)$
Sherman 統計量 S	5.050×10^{-1}
95% 信頼区間	$(3.444 \times 10^{-1}, 3.911 \times 10^{-1})$

*d 及び S の定義は [15] 参照

表2 時間尺度変換後におけるポアソン性検定結果
Table 2 Results of test for stationaly Poisson hypothesis after TSC.

インターバル χ^2 適合度	4.357×10^1
5% 有意水準	4.380×10^1
dispersion 検定 d	1.273×10^2
95% 信頼区間	$(7.336 \times 10^1, 1.284 \times 10^2)$
Sherman 統計量 S	3.791×10^{-1}
95% 信頼区間	$(3.456 \times 10^{-1}, 3.898 \times 10^{-1})$

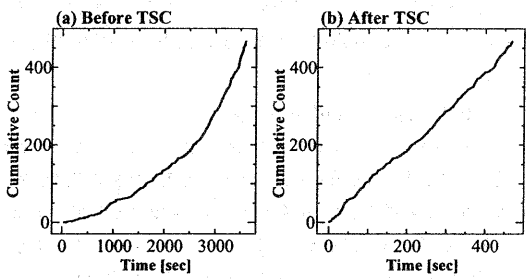

図1 リクエストの累積カウント数 (a) 時間尺度変換前 (b) 変感後
Fig. 1 Comulative request count, (a)Before TSC, (b)After TSC.

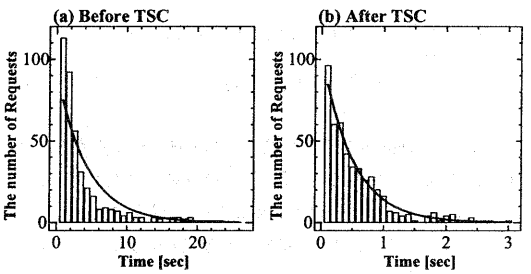

図2 インターバル分布 (a) 時間尺度変換前 (b) 変感後
Fig. 2 Interval histogram of artificially generated HTTP request occurrences, (a)Before TSC, (b)After TSC.

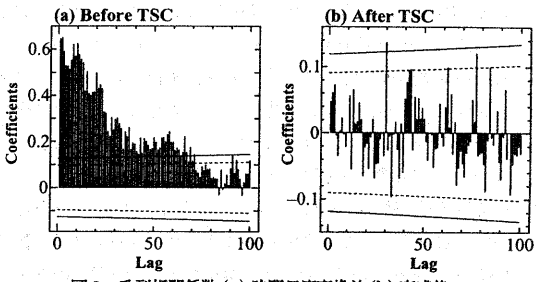

図3 系列相関係数 (a) 時間尺度変換前 (b) 変感後
Fig. 3 Serial correlation coefficients, (a)Before TSC, (b)After TSC.

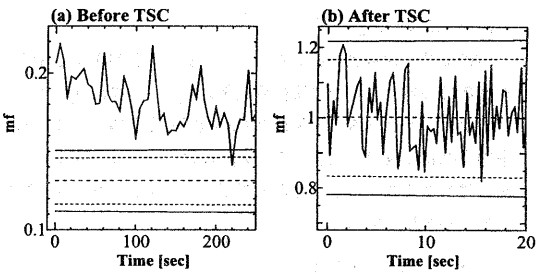

図4 強度関数 (a) 時間尺度変換前 (b) 変感後
Fig. 4 Intensity function, (a)Before TSC, (b)After TSC.

2.1.3 定常ポアソン性検定

本稿では，ポアソン過程を再生過程としてとらえることで，ポアソン性検定としてインターバル分布の指数分布に対する χ^2 適合度検定，系列相関関数の無相関性，強度関数による検定を採用した．さらに，ネットワークトラヒック解析において重要なバーストの有無等，インターバルの変動性に着目した指標として知られる dispersion 検定及び Sherman 統計量を加えて総合的なポアソン性検定を構成した．dispersion 検定に用いる d 統計量及び Sherman 統計量はいずれもインターバルが規則的に，つまり一定間隔になるほど小さな値をとり，バーストを含み不規則さが増すほど大きな値をとる [15].

2.2 ポアソン性の検定結果

以上述べてきた提案手法をもとにして，2002 年 5 月 14 日 (火) の午前 8 時から 9 時までの 1 時間分のユーザリクエストにおけるポアソン性検定を，時間尺度変換前の結果を表 1に，変換後の結果を表 2に示す．また，図 1にユーザリクエストの累積カウント数を，図 2にインターバル分布と指数分布の適合性を視察するためのヒストグラムを，図 3に系列相関係数を，図 4に強度関数を示す．図はそれぞれ (a) に時間尺度変換前，(b) に時間尺度変換後の結果を示している．表 1，2においてdispersion 検定では自由度を 100 とした．また，図 2におけるヒストグラムのビンの数は 30 であり，図中の実線は指数分布を表し，図 3における系列相関関数ではラグを 100 とし，図 4と併せて内側の破線は 5%，外側の実線は 1% の有意水準を表す．

時間尺度変換前のすべての検定においては定常ポアソン性が棄却された．これは非定常性をもつデータに対して定常ポアソン性検定を行った結果であり，非定常ポアソン過程に対して定常ポアソン性検定を適用すると，当然のことながら仮説は棄却されてしまう．しかし，この段階では過程のポアソン性が棄却されたのか，定常性が棄却されたのかを知ることはできない．

そこで，時間尺度変換を施し定常ポアソン過程点生起時刻列へ変換の後，定常ポアソン性検定を行った．その結果，時間尺度変換後では定常ポアソン性検定により仮説が棄却されないことから，対象となる短区間の実測リクエストデータが非定常ポアソン過程であることが正しく示唆され，さらに非定常性を考慮するために時間尺度変換を施すことの有用性が示される．

3. ドキュメントサイズ分布に関する統計解析

本章ではトラヒックを構成する重要な要素の一つであるリクエストから得られるドキュメントサイズの統計解析手法を提案し，このドキュメントサイズ分布の統計的性質を明らかにする．

3.1 統計解析用データ

本章の統計解析用データは，アクセスログからユーザリクエストに直接関係するTCP部分のドキュメントサイズを抜き出し，ドキュメントサイズのヒストグラムを作成しこれを利用した．図5に2002年10月7日から14日までの1週間分のリクエスト全体のドキュメントサイズに対して，ビン幅に含まれる頻度が全体の何％にあたるかを表す相対頻度によるヒストグラムで示した．これ以降のヒストグラムのビン幅は全て50byteである．この1週間で約220万(2,190,310)のアクセス(リクエスト数)が観測され，総伝送バイト数は約2Tbyte，平均サイズは約9Kbyte，95％サイズは27.8Kbyteである．95％サイズとは，27.8Kbyte以下のドキュメントサイズ数が全体の95％を占めていることを意味する．WWWのドキュメントサイズが非常に広い範囲をとることがわかる．なお，詳細については文献[9]に詳しい．

3.2 解析手法

本節では，対象となるドキュメントサイズ分布をある確率分布関数で表現するための解析手法を述べる．ドキュメントサイズのサンプルデータから候補とする確率分布関数のパラメタを最ゆう法にて推定しモデルを作成，最適なモデルを判定するために適合度を測り，どの確率分布がドキュメントサイズを詳細に表すかを検証する．

3.2.1 パラメタ推定法

最ゆう法とは任意に取り出した無作為標本が，ある確率分布関数に従うとき，関数に内包される不定パラメタの推定量を見つけ出す一方法である．不定パラメタ数をm，ゆう度関数を$\mathcal{L}(\theta_j), j=1,...,m$としたとき，$\mathcal{L}(\theta_j)$の値を最大せしめる最ゆう推定値$\hat{\theta_j}$を求めればよい．一般にゆう尤度関数は積の形で与えられることが多いことから，対数の形で表し，

$$\ln \mathcal{L}(\theta_j) = \ln \prod_{i=1}^{n} f(x_i;\theta_j) = \sum_{i=1}^{n} \ln f(x_i;\theta_j) \qquad (6)$$

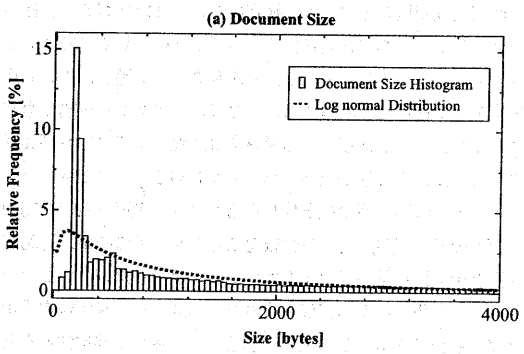

図5 リクエスト全体のドキュメントサイズ分布

Fig. 5 The document size distribution of the whole request.

としてゆう度関数の対数の最大値を考える．つまり，

$$\frac{\partial \ln \mathcal{L}(\theta_j)}{\partial \theta_j} = 0 \quad (j=1,2,...,m) \qquad (7)$$

を解いて$\mathcal{L}(\theta)$を最大ならしめる$\hat{\theta}$を求めることにする．

本稿ではモーメント法とNewton-Raphson法(以降NR法と呼ぶ)を段階的に利用した2段階最ゆう推定を行うことにした．これは，NR法による数値計算を行う場合，最ゆう推定値から大きくはずれた値を初期値として与えると根が収束しない可能性があり，これを回避するためである．つまりモーメント法によりパラメタの近似解を探り，そこで得られたパラメタの近似解をNR法の初期値として与え，最ゆう推定値を得る手法をとることにする．

3.2.2 適用する確率分布関数

図(5)のドキュメントサイズヒストグラムにおいて，2段階最ゆう推定法により推定したパラメタを用いて対数正規分布に当てはめた結果を図(5)に破線で示す．裾の部分に関してはよく適合していると言えるが，全リクエストの約50％以上を占める前半の立ち上り部分はよく表しているとは言いがたい．そこで確率分布関数を複数混合した複合確率分布関数を考えることにする．式(8)に複合確率分布関数の一般式を定義する．

$$\begin{aligned} F_{mix}(x;\Theta) &= \omega_1 F_1(x;\Theta_1) + \omega_2 F_2(x;\Theta_2) + \cdots \\ &= \sum_{i=1}^{n} \omega_i F_i(x;\Theta_i) \end{aligned} \qquad (8)$$

ここで，nは確率分布関数の数，ω_iは各確率分布関数$F(x;\Theta_i)$の重みである．本稿では2種類の分布関数を複合させることにし，以下の2タイプの複合分布を考えることにする．

- モデルタイプ1：正規分布＋対数正規分布

$$F_1(x;\Theta) = \omega_1 F_n(x;\Theta_n) + \omega_2 F_l(x;\Theta_l) \qquad (9)$$

- モデルタイプ2：指数分布＋対数正規分布

$$F_2(x;\Theta) = \omega_1 F_e(x;\Theta_e) + \omega_2 F_l(x;\Theta_l) \qquad (10)$$

ここで，$F_n(x;\Theta_n)$，$F_l(x;\Theta_l)$，$F_e(x;\Theta_e)$は各々，正規分布，対数正規分布，指数分布を表しており，各分布関数のパラメタを，$\Theta_n=\{\mu_n,\sigma_n\}, \Theta_l=\{\mu_l,\sigma_l\}, \Theta_e=\{a,b\}$とすることで，次式で与えられる．

$$F_n(x;\Theta_n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \int_0^x exp\left[\frac{-(y-\mu_n)^2}{2\sigma_n^2}\right] dy \qquad (11)$$

$$F_l(x;\Theta_l) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \int_0^x exp\left[\frac{-(\log y-\mu_l)^2}{2\sigma_l^2}\right] dy \qquad (12)$$

$$F_e(x;\Theta_e) = 1 - exp\left[\frac{-(x-a)}{b}\right] \quad (b>0) \qquad (13)$$

3.2.3 モデルの検定法

モデルの適合度を検定する際，χ^2適合度検定によりモデルの適合性を検定する方法が一般に知られているが，ヒストグラムのビン(クラス)幅が大きくなると検定が棄却されやすくなってしまい，ドキュメントサイズのヒストグラムのようにビン幅が非常に多い場合には適さないと考えられる．

そこで，2つの確率分布間の近さを評価する尺度として，情報幾何学の分野で導入された Kullback-Leibler 情報量 (以降 KL 情報量と呼ぶ：相対エントロピーとも呼ばれる) を利用してモデル間距離を測ることでモデルの適合度を検討することにする．2つの確率分布関数 $F(x), G(x)$ の密度関数を $f(x), g(x)$ とするとき，KL 情報量は以下の式で与えられる．

$$I(G|F) = \int g(x) \log \frac{g(x)}{f(x)} dx \qquad (14)$$

ここで $f(x) \propto constant$ とすると，$-I(G|F)$ はエントロピーを表す．式 (14) の $I(G|F)$ は，確率変数 X の値を観測することが，分布 $G(x)$ に比べ $F(x)$ に対して提供する情報量を意味する．なお一般には KL 情報量は $I(G|F) = I(F|G)$ なる関係は成り立たない．これは上述したように，確率変数 X が，分布 $G(x)$ に比べ $F(x)$ に対して提供する情報量を表しており，分布 $F(x)$ と $G(x)$ を対等に扱っていないからである．しかし，ここで，

$$\begin{aligned} J(G|F) &= I(G|F) + I(F|G) \\ &= \int (g(x) - f(x)) \log \frac{g(x)}{f(x)} dx \end{aligned} \qquad (15)$$

なる量を考えると，$J(G|F) = J(F|G)$ が成り立つ．これを分離情報量 (ダイバージェンス:Divergence) と呼ぶ．式 (15) のダイバージェンスは，分布関数 $F(x), G(x)$ に対する重みを対等とみなした場合の情報量を示しており，このダイバージェンスが大きいほど，分布 $F(x)$ と $G(x)$ は弁別されやすくなる．すなわち，ダイバージェンスは2つの分布間の距離を表す一つの尺度として捉えることができる [16].

また離散型分布に対する KL 情報量およびダイバージェンスはそれぞれ次のように表される．

$$I(G|F) = \sum_{i=0}^{\infty} \log \frac{p_{gi}}{p_{fi}} dx \qquad (16)$$

$$J(G|F) = \sum_{i=0}^{\infty} (p_{gi} - p_{fi}) \log \frac{p_{gi}}{p_{fi}} dx \qquad (17)$$

ただし，p_{fi}, p_{gi} は $X = x_i (i = 0, 1, 2, ..., \infty)$ となる確率を表す．本稿では，与えられた各確率分布関数より密度関数を求め，式 (17) におけるダイバージェンスを用いてモデル間の距離を測ることでモデルの当てはまりのよさを検討する．

3.3 解析結果

前節までの解析手法に従って，図5に示したドキュメントサイズのヒストグラムに対して複合確率分布関数を適合させた結果を示す．図6にモデルタイプ1の正規分布＋対数正規分布，図7にモデルタイプ2の指数分布＋対数正規分布を当てはめた結果を，表3に各モデルタイプ毎のダイバージェンスを示す．

図5の単一の確率分布関数 (対数正規分布) を適合させた結果と見比べても，前半部分の立ち上がりの部分，そして裾の部分と明らかに複合確率分布関数を当てはめた方が，よりドキュメントサイズの分布を詳細に表していることがわかる．さらに，複合確率分布関数のうち，モデルタイプ1の正規分布＋対数正規分布よりもモデルタイプ2の指数分布＋対数正規分布をあてはめた方がダイバージェンスの値が小さくなっていることから，よりドキュメントサイズ分布を詳細に表していることが分かる．

しかし，本稿ではダイバージェンスの値のみを用いてモデルの適合度を比較している．つまりダイバージェンスの値の小さいモデルの方が，実測されたドキュメントサイズ分布により近いというだけであり，実際には，「得られたドキュメントサイズ分布はある複合確率分布関数に従う」との仮説が正しいかどうか，統計的な適合度検定を行わなければならないが，これは今後の課題としたい．以上のように，提案した手法を基に複合確率分布関数を当てはめることでドキュメントサイズをより詳細に表現可能であることが示された．

表3 ドキュメントサイズと適合させたモデルとのダイバージェンス
Table 3 Divergence between document size and the fitted mixture of probability distribution model.

ダイバージェンス (正規分布＋対数正規分布)	1.67×10^1
ダイバージェンス (指数分布＋対数正規分布)	7.99×10^0
単一分布関数 (対数正規分布) を用いた場合のダイバージェンス (図5)	6.09×10^1

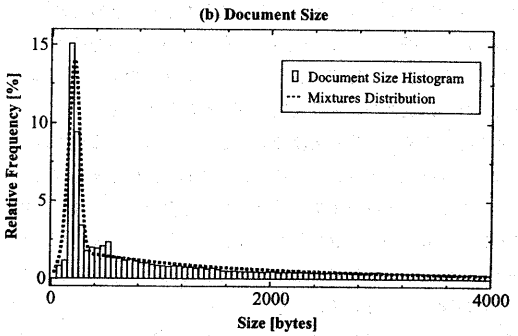

図6 混合確率分布関数 (正規＋対数正規) を当てはめたドキュメントサイズ分布
Fig. 6 The document size distribution fitting the mixtured distribution. (Normal Dist. & Log-Normal Dist.)

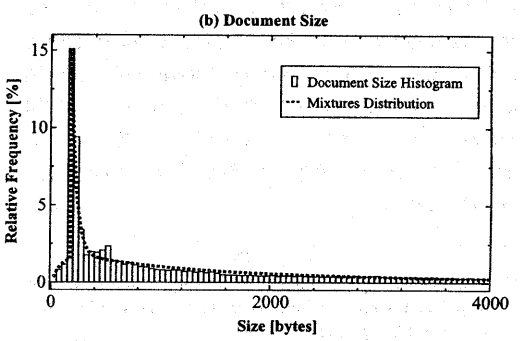

図7 混合確率分布関数 (指数＋対数正規) を当てはめたドキュメントサイズ分布
Fig. 7 The document size distribution fitting the mixtured distribution. (Exponential Dist. & Log-Normal Dist.)

4. むすび

本稿では，ユーザが直接発生させるHTTPリクエストとそのリクエストから得られるドキュメントサイズに注目して統計解析手法を提案し，解析を行った．短区間のユーザリクエストに対して，非定常性を考慮するために時間尺度変換を行いポアソン性検定を行った結果，対象となる短区間の実測リクエストデータが非定常ポアソン過程であることが正しく示唆された．本稿では1区間のみの結果表示に留まったが，1日分のユーザリクエストに関して短区間ごとに全ての区間を検証する必要がある．また，確率分布関数を複合させることでドキュメントサイズの詳細なモデル化が可能であることを示した．本稿ではリクエスト全体のドキュメントに対しての統計結果であるが，実際のドキュメントはテキストファイルや画像ファイルなど，拡張子ごとに様々なタイプのドキュメントが存在し，タイプごとにそのドキュメントサイズの分布が異なることが予想されるので，その検討も行う必要がある．

実際にはユーザが発生させるリクエストを起点として2次的にバースト的なリクエストが発生する．構成的なWWWトラヒックのモデル化を考えた場合，このバーストの性質を明らかにすることもリクエストの統計的性質を考える上で重要である．ユーザリクエスト，2次的リクエスト(バースト)，ドキュメントサイズ分布等トラヒックを構成する要素の統計的性質を明らかにすることで，WWWトラヒックの構成的モデル化が可能となる．このようなトラヒックの構成的モデルに基づいて現実的な疑似トラヒックの作成，シミュレーションを行うことで，ネットワークトラヒックの制御や予測等，幅広い応用が期待できると考えられる．

文　献

[1] W.E. Lerand, W.Willinger, M.S. Taqqu, and D.V. Wilson, "On the self-similar nature of Ethernet traffic," SIGCOMM '93, pp. 183-193, 1993.

[2] W. Willinger, M.S. Taqqu, W.E. Lerand, 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.

[3] 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.

[4] H. Michiel, and K. Laevens,"Teletraffic engineering in a broad-band era," Proc. IEEE, vol. 85, no. 12, pp. 2007-2033, 1997.

[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，" 会誌「システム / 制御 / 情報」, vol. 43, no. 3, pp. 117-122, 1999.

[6] 大久保 智史, 篠 秀明, 八名 和夫, "ネットワークトラヒック時系列のフラクタル性と信号モデル，" 第16回 ゆらぎ現象研究会抄録集, pp. 19-20, Nov, 2001.

[7] K.R. Krishnan, "A new class of performance results for a fractional Brownian traffic model," Queueing Systems, vol. 22, pp. 277-285, 1996.

[8] 小沢利久, 町原文明, 石橋佳介, " マルチメディアトラヒック理論の最新動向，" 信学誌, vol. 81, No. 5, pp. 506-515, 1998.

[9] 名部正彦, 馬場健一, 村田正幸, 宮原秀夫, " インターネット・アクセスネットワーク設計のためのWWWトラヒックの分析とモデル化，" 信学論(B-I), vol. J80-B-I, no.6, pp. 428-437, 1997.

[10] 篠秀明, 北澤慶一, 八名和夫, " インターネットアクセスネットワークにおけるHTTP通信リクエスト発生の非定常解析法" 信学論(B), vol. J84-B, no. 8, pp. 1494-1504, 2001.

[11] E. Cinlar, "Superposition of point processes," in Stochastic Point Processes Statistical Analysis, Theory, and Applications, P.A.W. Lewis Ed., Wiley, New York, 1972.

[12] 八名和夫, " 非定常点過程に対するポアソン性検定，" 信学論(A), Vol. J67-A, No. 5, pp. 431-438, 1984.

[13] I. Bar-David, "Communication under the Poisson regime," IEEE Trans. Inf. Theory, IT-15, 1, pp. 31-37, 1996.

[14] K. Yana, N. Takeuchi, Y. Takikawa, M. Shimomura, "A Method for Testing an Extended Poisson Hypothesis of Spontaneous Quantal Transmitter Release at Neuromuscular Junctions," Biophysical Journal, vol. 46, pp. 313-330, 1984.

[15] D.R. Cox, and P.A.W. Lewis, "The statistical analysis of series of events," Methuen & Co. , London, 1966.

[16] 澤田 清, 三道 弘明, 藤井 進 "Kullback-Leiblerの情報量に基づくソフトウェアの信頼性実証試験に関する離散型モデル，" 信学論(A),vol.J83-A, no. 3, pp. 830-833, 1996

付　録

1. dispersion検定

時間軸を等間隔に分割した各区間における事象数を $n_i (i = 1, ..., K)$, n_i の標本平均を $\bar{n}$ としたとき，n_i の $\bar{n}$ からの偏差の指標は，

$$d = \sum_{i=1}^{K} \frac{(n_i - \bar{n})^2}{\bar{n}} \qquad \text{(A·1)}$$

と定義され，式(A·1)を用いたポアソン性の検定をdispersion検定という．ポアソン過程において d は自由度 $K-1$ の χ^2 分布に従う．d 統計量は間隔が規則的になり一定値に近付けば小さい値をとり，事象生起がクラスタ化されバースト状である場合大きな値をとる．

2. Sherman統計量

観測時間 T で正規化した生起時刻を $v_i = t_i/T$ とすると，

$$E[v_i - v_{i-1}] = \frac{1}{n+1} \qquad i = 1, \ldots, n+1 \qquad \text{(A·2)}$$

となる．ただし $t_0 = 0, t_{n+1} = T$, n は時間 t 内の事象生起数とする．ここで，Sherman統計量は $v_i - v_{i-1}$ の期待値からの偏差を累計し，

$$S_n = \frac{1}{2} \sum_{i=1}^{n+1} |v_i - v_{i-1} - \frac{1}{n+1}| \qquad \text{(A·3)}$$

と定義される．Sherman統計量はdispersion検定同様インターバルが規則的になる程小さな値をとり，不規則さが増すほど大きな値をとる．ポアソン過程のもとでは，

$$E[S_n] = \frac{1}{e} \qquad \text{(A·4)}$$

$$var[S_n] = \frac{0.05908}{n} - \frac{0.07145}{n^2} + o(\frac{1}{n^3}) \qquad \text{(A·5)}$$

となる．Sherman統計量は漸近的に正規分布となることが知られており式(A·4)，式(A·5)を用いてポアソン性の検定が行うことができる．