早稲田大学 産業経営研究所『産業経営』
第43号 2008年11月 pp. 3-13

次世代ネットワーク（NGN）時代における ウェブサーバシステム応答遅延評価法

～拡散過程による確率モデル化と解析～

髙 橋 敬 隆

要 旨

リアルタイム性を必要とするインターネットコマースにおいては，ウェブサーバへのアクセス応答遅延時間を評価・予測することがオペレーションズ・リサーチ分野で重要かつ喫緊の研究課題となって来ている。アクセス応答遅延時間を短縮するために，通常複数の地理的に異なるサイトにあるプロキシを介することによりウェブサーバシステムの性能改善が図られている。しかしながら，先行研究にはプロキシを考慮したウェブサーバの応答遅延解析はあまり見当たらない。本稿では，上述した課題を解決するため，複数プロキシ・一般到着モデルを対象に，拡散過程により定式化する。複数プロキシサーバからの出力過程を特徴付ける統計量（平方変動係数）を計算することによりウェブサーバへのアクセスメカニズムを解明している。ウェブサーバシステムにおける残余仕事量を拡散過程と見做したモデル化を提案し，複数プロキシ・一般到着ウェブサーバシステムにおける平均応答遅延時間の公式を導出している。本モデルの特殊な場合（単一プロキシ・ポアソン到着モデル）に対し，ここで得られた結果は既存公式に一致していることが示される。

キーワード：次世代ネットワーク，オペレーションズ・リサーチ，インターネットコマース，ウェブサーバ，マルコフ過程，拡散過程，待ち行列理論

Access-Delay Evaluation Engine for a Web-Server System with Multiple Proxy Servers in the Next Generation Network (NGN) Era
—— Modeling and Analysis via Diffusion Process ——

Yoshitaka TAKAHASHI

Abstract

It is an important and urgent OR (Operations Research) issue to evaluate and/or estimate the delay in a web-server system handling internet commerce real-time services. Usually, multiple proxy servers located separately from the web-server site enable shortening of the web-server access delay in order to guarantee the quality of real-time application services. However, there exists almost no literature analysing access delay for the web-server system with multiple proxy servers. The goal of this paper is to provide an access-delay analysis engine for a web-server system with multiple proxy servers. Our approach is based on the diffusion process technique. We derive the statistics (the squared coefficients of variation) of the individual output processes from the proxy servers. By regarding the unfinished workload in the web-server system (i.e. the incoming output processes from the proxy servers) as a diffusion process, we derive a mean-delay explicit formula. Our formula is shown to be consistent with the previously obtained result for a special case (single proxy server and Poissonian arrival model).

Key words: Next Generation Network (NGN), Operations Research (OR), internet commerce, web server, Markov process, diffusion process, queueing theory

投稿受付日 2008年1月31日
採択決定日 2008年4月30日

早稲田大学商学学術院教授

1．はじめに

元来研究者間用のクローズドなインターネットがオープン化され商業利用に開放されてからほぼ15年が経過した。現在まで個別に発展してきたインターネットサービス用IP（Internet Protocol）ネットワークと電話サービス用ネットワークをIP通信ネットワークとして統合化する次世代ネットワーク（Next Generation Network, NGN）がいよいよ実用化段階に入っている。

NGNでは，各国の通信規制が取り除かれれば，音声・Fax・テキストメール・ウェブ・静止画像・動画像・テレビ会議と云った既存サービスだけでなく，ハイビジョン放送や香り・匂いの伝送サービスも提供される予定である。従って，全てのメディアがNGNリソースにトラフィックとして加わってくるようになる。

リソース共用のプラットフォームとして，NGNは理想的である。しかし，個々のサービスが要求するネットワーク性能品質は極端に異なっている。例えば，データサービスでは多少の遅延をある程度許容してもIPパケット損失はほとんど許容されない（loss-sensitive）。一方，音声サービスでは受信側の人間が送信側の音声を明瞭に聞けるという範囲内でIPパケット損失はある程度許容されるが，IPパケット遅延はほとんど許容されない（delay-sensitive）（宮原・村田 2001参照）。

ICT（Information Communication Technology）関連機器のコスト低下と通信・放送技術がNGNを実用化たらしめているが，リアルタイム性を必要とするインターネットビジネスにおいては，ウェブサーバへのアクセス応答遅延時間を評価・予測することがオペレーションズ・リサーチ分野で重要かつ喫緊の研究課題となって来ている。

ウェブサーバへのアクセス応答遅延を短縮するためにはインターネットにおける全リンク・ルータを高速・大容量化（光ファイバー化・光ルータ化）するのが本来のあるべき姿である。しかし例えば日本の場合，中継リンクはほとんど光ファイバー化されてはいても加入者線（subscriber line）全てが光ファイバー化されているとは言い難い。トラフィックの少ない地域における加入者線を光ファイバー化するには，ネットワーク事業経営を悪化させかねない程の莫大な経費が掛かるためである。現状では，ADSL（Asymmetric Digital Subscriber Line）等の手段を用いて，電話サービス用ケーブル回線をモデムによって中速化・中容量化させている。すなわちインターネットは一朝一夕に超高速・大容量化出来ないシナリオになっている。

そこで，通常は地理的に異なるサイトにあるプロキシを介することによりウェブサーバシステムの応答遅延時間の短縮が図られている。先行研究にはプロキシを考慮したウェブサーバの応答遅延解析はほとんど見当たらない。わずかに単一プロキシ・ポアソン到着モデルを対象にマルコフ過程論による解析を行なっている文献（Takahashi 2001）があるに過ぎない。

本稿では，上述した課題を解決するため，複数プロキシ・一般到着モデルを対象に，拡散過程論により定式化する。複数プロキシを設定しプロキシサーバからの出力過程の統計量（平方変動係数）を計算することにより，ウェブサーバへのアクセスメカニズムを解明している。ウェブサー

バシステムにおける残余仕事量を拡散過程と見做し，拡散過程の原点における境界条件として反射壁（RB）並びに基本復帰境界（ER）を設定している。これらの境界条件下の拡散方程式解より，複数プロキシ・一般到着ウェブサーバシステムにおける平均応答遅延時間の陽公式を導出している。本モデルの特殊な場合（単一プロキシ・ポアソン到着モデル）に対し，本稿で得られた結果は既存公式に一致していることが示される。

2．先行研究と研究課題

ウェブサーバシステム自体は例えば電子情報通信・情報処理・オペレーションズ・リサーチ分野で活発に研究されているが，ウェブサーバシステムに限らずインターネットにおける情報システム機器の性能解析を理論的に行なっている文献はほとんど存在しない。これは，インターネットの起源が研究者間の無償クローズドネットワークにあるため，インターネット上に流れるIPパケットのアドレス空間には現行のバージョン（v4）では課金ビットもなく，サービス品質を保証する必要がなかったからである。現に，インターネットでのサービスはベスト・エフォート（best effort）と呼ばれ「ネットワークは最善を尽くしてサービスに努めているから，品質保証は勘弁して下さい」と言っているからである。

NGNでは，IPアドレスは次世代バージョン（v6）が使われる予定で，現行のインターネットよりきめ細かいサービス管理・設計・制御・運用が可能となる。すなわち，クラス毎に課金方式を変更し，トラフィック制御方式を導入することができるようになる。このNGN時代には，現行のベスト・エフォート型サービスも残るが，情報ネットワークとしての接続品質や安定品質を保証するサービスがビジネスモデルとして出現する。従って，ユーザクラス毎に品質を保証するために，性能解析は重要かつ喫緊の研究課題になる。

閑話休題（それはさておき），筆者はTakahashi（2001）において，単一プロキシのあるウェブサーバシステムにおけるアクセス応答遅延時間解析を行なった。アプローチはマルコフ過程論（Ross 1996）を採用し，ユーザからのウェブサーバへのアクセス要求はポアソン到着，すなわち，アクセス要求間隔は互いに独立で同一の指数分布に従っていることを仮定し，サービス時間が確率的にゼロとなる変形M/G/1を解析して，平均アクセス応答遅延時間に対する公式を導出した。本稿では，次節で詳述するが，より現実に近い複数プロキシのあるウェブサーバシステムを取り扱う。ユーザからのウェブサーバへのアクセス要求間隔は独立で同一の任意の（指数分布とは限らない）一般分布に従うと仮定する。すなわち，プロキシ数とアクセス要求間隔分布の点で，既存文献（Takahashi 2001）が対象としたモデルをより一般化している。

3．複数プロキシサーバを介したウェブサーバシステムのモデル化

3.1　ウェブアクセスオペレーション

Takahashi（2001）と同様，ウェブアクセスへのオペレーションを以下のようにモデル化する。

Takahashi（2001）と異なる個所（の一つ）は，プロキシサーバ台数が複数であることを許し，N台（$N \geqq 1$）としている点である。

(1) クライアント（ユーザ，以降，ORにおける待ち行列の言葉で「客」と呼ぶ）は，ウェブから欲しいコンテンツをダウンロードしたい。このウェブサーバは1台であるが，N台のプロキシサーバを（ウェブサーバとは通常異なるサイトに）有している。

(2) 客は（認識せずに）加入者線（Subscriber line）を通して，ISPにおけるプロキシサーバにコンテンツを要求する。このプロキシサーバを第i番目という意味で，プロキシ#iと呼ぶ（$1 \leqq i \leqq N$）。

(3) プロキシ#iは，そのサーバ内キャッシュ（cache，記憶装置の一種）内に，客の要求コンテンツがあれば，直ぐ当該コンテンツを客へダウンロードする（客の待ち時間はゼロ）。

(4) プロキシ#iのキャッシュ内に当該コンテンツが無ければ，ウェブサーバに当該コンテンツを要求しダウンロードする。（客のサービス時間はプロキシがウェブサーバに当該コンテンツを要求してからそれをウェブサーバから得る迄の時間に相当する）

(5) このときのリソース競合は，ウェブサーバで生じる。ウェブサーバには単一バッファ（待ち行列）があり，全プロキシ#i（$1 \leqq i \leqq N$）から要求されてくるコンテンツの各プロキシへの配送を先着順（First-Come, First-Service）で行うものとする

3.2 待ち行列モデル化

Takahashi（2001）では，サービス時間として，ある確率（プロキシ内キャッシュに客の要求したコンテンツがある確率）でサービス時間がゼロとなることを許し，客の到着過程をポアソン到着，すなわち，相続く客の到着間隔は互いに独立で同一の指数分布に従うとして，マルコフ過程により定式化し，平均待ち時間や平均系内客数に対する陽公式を導出した。

さてTakahashi（2001）のアプローチを，複数プロキシサーバのある場合に適用すると解析が複雑になる。各プロキシ内に客の要求するコンテンツのある確率が，一般には異なるからである。本稿では見方を変え，応用確率論の言葉で言えば，到着過程をthinningする。すなわち，サービス時間がゼロの到着は，後段に控えているウェブサーバシステムへの到着にはなかったことにする。このアプロ―チは以下のように纏められる。記号の導入とともに述べておく。

(1) 相続く客のプロキシ#iへの到着間隔（A_i）は互いに独立で同一分布に従う（independent and identically distributed, iid）ものとする（$1 \leqq i \leqq N$）。確率変数A_iの平均と平方変動係数は次式で定義される：

$$E(A_i) = \frac{1}{\lambda_i} \qquad (i = 1, 2, \ldots, N) \tag{1}$$

$$C_{Ai}^2 = \frac{V(A_i)}{E(A_i)^2} \qquad (i = 1, 2, \ldots, N) \tag{2}$$

ここで，任意のiid確率変数Xに対し，E(X)はXの期待値，V(X)はXの分散，C_XはXの変動係数（＝標準偏差/平均）を表わしている。以降もこの記法を用いる。

(2) 客の要求コンテンツがプロキシ#iキャッシュ内にある確率をp_iとし，この確率でウェブサーバシステム外に退去するものとする。

(3) 客の要求コンテンツがプロキシ#iキャッシュ内にない場合（この場合の確率は$1-p_i$），ウェブサーバシステム内における単一バッファ（待ち室）に入り，到着順で処理される。バッファ容量は無限にあるものとしバッファオーバフローはないものと仮定する。

(4) ウェブサーバシステムにおける処理は，単一サーバで行われるものと仮定する。その処理時間(B)は，プロキシがウェブサーバに当該コンテンツを要求をしてから当該コンテンツをウェブサーバから得る迄の時間と見做す。処理時間間隔は（ほとんどのコンピュータネットワークの確率統計的な解析（Allen 1990, Robertazzi 2000）と同様）iidと仮定する。処理時間(B)の平均と平方変動係数は次式で定義される：

$$E(B)=\frac{1}{\mu} \tag{3}$$

$$C_B^2=\frac{V(B)}{E(B)^2} \tag{4}$$

注意1：プロキシ#iキャッシュ内には人気の高い順にコンテンツを揃えておく。このような人気順に並べた場合の確率関数（プロキシ#iサーバ内で計算作成されるヒストグラム）$\{f_k : k=1, 2, \cdots, D\}$はZipfの法則が成り立つことが知られている。ここでDはコンテンツの総数を表す。すなわち，第k番目に人気のあるコンテンツを要求する確率f_kは次式で与えられる：ある数αが存在して，

$$f_k=\frac{k^\alpha}{\sum_{k=1}^{D} k^\alpha} \qquad k=1, 2, \cdots, D \tag{5}$$

さて，K_iをプロキシ#iのキャッシュサイズ（K_iコンテンツまで記憶可能）とする。本節3.2で提案したモデル化において，客の要求コンテンツがプロキシ#iキャッシュ内にある確率p_iはZipfの法則を適用した場合，次式で与えられる：

$$p_i=\sum_{k=1}^{K_i} f_k=\frac{\sum_{k=1}^{K_i} k^\alpha}{\sum_{k=1}^{D} k^\alpha} \tag{6}$$

メモリ価格の低廉化によりキャッシュサイズK_iをかなり大きくすることが技術的に可能となったが，サイバー空間における所謂ロングテール現象（梅田 2006）によって，Dに比べるとK_iはまだ小さい。

4．ウェブサーバシステムへの到着過程解析

前節で行なったモデル化の解析を進めるに当たり，まず，各i（$i=1, 2, \cdots, N$）に対し，プロキシ #i からの出力間隔を統計的に特徴付ける。個々の出力過程のN個の重畳（superposition）がウェブサーバシステムへの到着過程である。プロキシ #i への客の到着間隔A_iのラプラス・スティルチェス変換（Laplace-Stieltjes Transform, LST）を $A_i^*(s)$ とおく。すなわち，

$$A_i^*(s) = \int_0^\infty e^{-st} dP(A_i \le t) \tag{7}$$

実際に，ウェブサーバシステムにやって来る実質的な客（substantial customers）は，プロキシ #i 内では要求コンテンツが見出せなかった訳である。その客の到着間隔を A_{is} とおくと，プロキシ #i 内に見出せなかった回数はパラメータ p_i の幾何分布に従う故，A_{is} の LST $A_{is}^*(s)$ は次式で与えられる。

$$A_{is}^*(s) = \sum_{n=0}^{\infty} (1-p_i) p_i^n [A_i^*(s)]^{n+1} \tag{8}$$

式(8)において，無限和の収束を仮定して計算すると，次式を得る：

$$A_{is}^*(s) = \frac{(1-p_i)A_i^*(s)}{1-p_i A_i^*(s)} \tag{9}$$

LST の性質から，確率変数 A_{is} の１次・２次モーメントは，式(9)をsについてそれぞれ１階・２階微分して$s=0$を代入して求められる。

$$E(A_{is}) \equiv \frac{1}{\lambda_{is}} = \frac{E(A_i)}{1-p_i} = \frac{1}{\lambda_i(1-p_i)} \tag{10}$$

$$E(A_{is}^2) = \frac{(1-p_i)E(A_i^2) + 2p_i[E(A_i)]^2}{(1-p_i)^2} \tag{11}$$

よって，実質到着間隔 A_{is} の平方変動係数を C_{is}^2 と略記すると，次式となる：

$$C_{is}^2 \equiv \frac{V(A_{is})}{E(A_{is})} = (1-p_i)C_{Ai}^2 + p_i \tag{12}$$

繰り返し，ウェブサーバシステムにやって来る実質客の到着過程は，式(10)，(12)で特徴付けられた平均と平方変動係数を持つ再生過程の重畳（superposition）である。

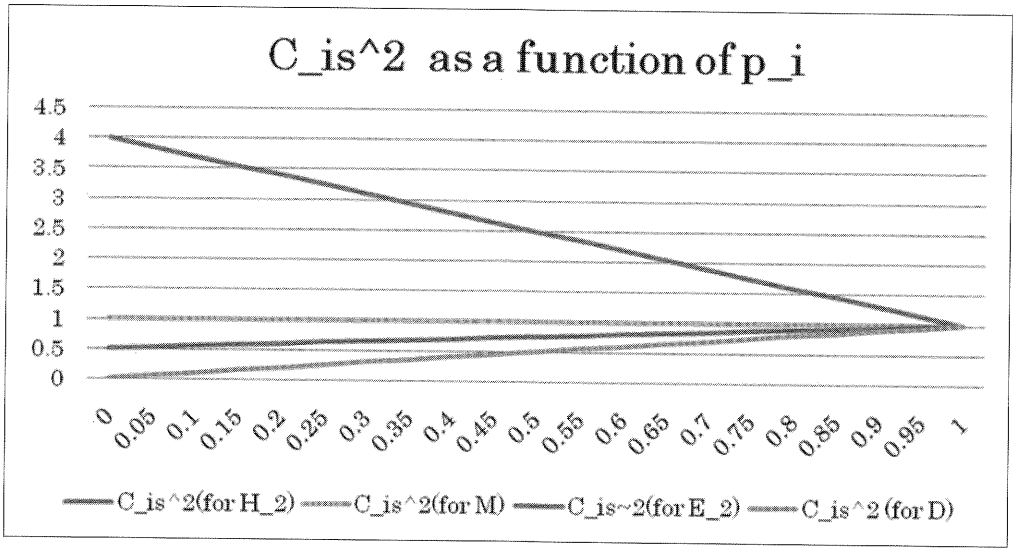
注意２：各 i（$i=1, 2, \cdots, N$）に対し，式(12)より次のことが分かる。

a) プロキシ #i が介在しない（$p_i=0$）ときは，$C_{is}=C_{Ai}$ となる（for some $i=1, 2, \cdots, N$）。

b) 既存文献（Takahashi 2001）で仮定している「単一プロキシ（$N=1$）・ポアソン到着（$C_{Ai}=1$）モデル」のときは，プロキシ #i のキャッシュ内ヒット確率（p_i）に関係なく実質客の到着過程もポアソン到着（$C_{is}=1$）となる。

図1．プロキシ #i キャッシュ内に要求コンテンツがヒットする確率（p_i）を関数としたウェブサーバシステムにやって来る実質客の到着間隔の平方変動係数（C_{is}^2）のグラフ

Fig.1 The squared coefficient of variation (C_{is}^2) of the out-put process from proxy #i as a function of the probability (p_i) that a request finds its content in the proxy cache.



c) 図1に，式(12)による数値例を示す。横軸はプロキシ #i キャッシュに要求コンテンツがヒットする確率 p_i を，縦軸はウェブサーバシステムにやって来る実質客の到着過程に対する平方変動係数 C_{is}^2 を示す。プロキシ #i への元々の到着間隔が2次超指数分布（H_2, $C_{Ai}^2=4$），指数分布（M, $C_{Ai}^2=1$），2次アーラン分布（E_2, $C_{Ai}^2=0.5$），単位分布（D, $C_{Ai}^2=0$）に従うときを示している。プロキシ #i からの出力間隔（ウェブサーバシステムにやって来る実質客の到着間隔）の平方変動係数 C_{is}^2 が p_i に関して線型的にポアソン到着に対する平方変動係数（$C_{Ai}^2=1$）に収束している様子が分かる。すなわち，プロキシ #i のキャッシュ内ヒット確率（p_i）が十分大きいときは，プロキシ #i への元々の到着過程が何であろうと

$$C_{is}^2 \to 1 \quad (as\ p_i \to 1)$$

となり，ウェブサーバシステムにやって来る実質客の到着間隔はほぼ指数分布（要するにポアソン到着）とみなしてよいことが分かる。逆に，確率 p_i が小さいときは，プロキシ #i への元々の到着間隔分布の影響が大きいことが分かる。

5．拡散過程による解析

第3. 2節で述べたように，ウェブサーバシステムを単一サーバとして待ち行列モデル化し，そのサーバの残余仕事量過程 $\{V(t)\}$ を拡散過程と見做す。すなわち，$f(x, t)$ を $V(t)$ の確率密度関数（probability density function, pdf）とすると，

$$f(x,t)\,dx \equiv P(x \le V(t) < x+dx) \tag{13}$$

と略記出来る。待ち行列モデルにおける残余仕事量過程を拡散過程と見做す場合，負にはならないため，原点（$x=0$）における境界条件が必要である。境界条件として，反射壁（reflecting barrier, RB）境界と基本復帰（elementary return, ER）境界が知られている。

5.1 反射壁（RB）境界を有する拡散方程式

原点（$x=0$）に反射壁を置くと，$V(t)$ の pdf $f(x,t)$ は次式を満たす：

$$\frac{\partial f}{\partial t} = -\alpha\frac{\partial f}{\partial x} + \frac{\beta}{2}\frac{\partial^2 f}{\partial x^2} \tag{14}$$

$$0 = \left[-\alpha f + \frac{\beta}{2}\frac{\partial f}{\partial x}\right]_{x=0} \tag{15}$$

ここで，α，β は拡散パラメータと呼ばれ，次式で定義される。

$$\alpha \equiv \lim_{t\to\infty}\frac{E(U(t))}{t} \tag{16a}$$

$$\beta \equiv \lim_{t\to\infty}\frac{V(U(t))}{t} \tag{16b}$$

ここで，原点（$x=0$）に境界を置かない（時間の経過とともに負の値を取る）残余仕事量過程 $\{U(t)\}$ を考えている。

式(13)，(14)を満たす（RB 境界を有する拡散方程式に対する）過渡解は Heyman（1975）により得られている。解の適切性や函数解析的構造については文献（吉田・伊藤 1976，伊藤 1979）参照。特に定常状態における $f(x,t)$，$V(t)$ を $f(x)$，V 等と略記すると，

$$f(x) = -\frac{2\alpha}{\beta}e^{\frac{2\alpha}{\beta}x} \tag{17}$$

拡散パラメータ α，β の決定は，第5.3節で行う。

5.2 基本復帰（ER）境界を有する拡散方程式

第4章の結果から，原点（$x=0$）に基本復帰境界を置くと，$V(t)$ の pdf $f(x,t)$ は次式を満たす：

$$\frac{\partial f}{\partial t} = -\alpha\frac{\partial f}{\partial x} + \frac{\beta}{2}\frac{\partial^2 f}{\partial x^2} + \sum_{i=1}^{N}\lambda_{is}\pi_0(t)\frac{dB(x)}{dx} \tag{18}$$

$$\frac{d\pi_0(t)}{dt} = -\sum_{i=1}^{N}\lambda_{is}\pi_0(t) + \left[-\alpha + \frac{\beta}{2}\frac{\partial f}{\partial t}\right]_{x=0} \tag{19}$$

$$\lim_{x\to 0} f(x,t) = \lim_{x\to\infty} f(x,t) = 0 \tag{20}$$

ここで，拡散パラメータ α，β は反射壁境界のときと同様，式(16)で定義され，決定は次節で行

なう。式(18)，(19)に表れる $\pi_0(t)$ はシステム空き（system idle）確率，$dB(x)/dx$ はルベーグ積分論におけるラドン・ニコディム密度関数（伊藤 1963）であり，第3. 2節で定義した処理時間（B）が連続型確率変数のときは，通常の確率密度関数 pdf $dB(x)/dx = b(x)$ に等しい。

式(18)－(20)を満たす（ER 境界を有する拡散方程式に対する）定常解は Takahashi（1986）により得られている。この定常解を用いて

$$f(x) = \frac{2\pi_0}{\beta} \sum_{i=1}^{N} \lambda_{is} \int_0^x (1 - B(y)) e^{-\frac{2\alpha y}{\beta}} dy \bullet e^{\frac{2\alpha x}{\beta}} \tag{21}$$

$$\pi_0 = -\alpha \tag{22}$$

なお，定常状態が存在するためのには $\pi_0 > 0$でなければならない（次節注意３参照）。

5. 3　拡散パラメータの決定

中心極限定理（鈴木・山田 2001）を用いると，任意の（再生間隔の平均 m，分散 σ_M^2をもつ）再生過程 $\{M(t)\}$ に対して，$t \to \infty$ のとき

$$M(t) \sim N\left(\frac{1}{m}t, \frac{\sigma_M^2}{m^3}t\right) \tag{23}$$

であることが直接的に示される。ここで，$N(\mu, \sigma^2)$ は平均 μ，分散 σ^2を持つ正規分布であることを示す。

従って，式(23)と第４章（プロキシ #i からの出力過程）の解析を用いて，Takahashi（1986）第3. 3節の議論を繰り返すと，式(16)で定義される拡散パラメータは次式のように求められる：

$$\alpha = \sum_{i=1}^{N} \frac{\lambda_{is}}{\mu} - 1 \tag{24a}$$

$$\beta = \sum_{i=1}^{N} \frac{\lambda_{is}}{\mu^2}(C_{is}^2 + C_B^2) \tag{24b}$$

注意３：式(22)に式(24a)を代入すると，

$$\pi_0 = 1 - \sum_{i=1}^{N} \frac{\lambda_{is}}{\mu}$$

である。上式はリトルの公式（Allen 1990，Wolff 1989）からも導かれる（G/G/1待ち行列モデルにおける厳密解である）。定常状態が存在するためには

$$\sum_{i=1}^{N} \frac{\lambda_{is}}{\mu} < 1$$

が必要条件であることが分かる。

5. 4　平均システム性能評価尺度

平均残余仕事量 $E(V)$ 次式で与えられる：

RB 境界のとき

$$E(V)=\int_0^\infty xf(x)dx=-\frac{\beta}{2\alpha} \tag{25-RB}$$

ER 境界のとき

$$E(V)=\int_0^\infty xf(x)dx=\frac{1}{2}\sum_{i=1}^{N}\lambda_{is}\left[E(B^2)-\frac{\beta}{\alpha}\frac{1}{\mu}\right] \tag{25-ER}$$

平均待ち時間 $E(W)$ は所謂，客平均（customer average）である。上で求めた平均残余仕事量 $E(V)$ は時間平均（time average）である（Wolff 1989）。客平均と時間平均の一般的な関係式（Brumelle 1971）を用いて，次式を得る。

$$E(W)=\frac{E(V)\dfrac{\sum_{i=1}^{N}\lambda_{is}}{2}E(B^2)}{\sum_{i=1}^{N}\dfrac{\lambda_{is}}{\mu}} \tag{26}$$

リトルの公式（Allen 1990，Wolff 1989）により，平均待ち客数 $E(N)$ は次式で与えられる：

$$E(N)=\sum_{i=1}^{N}\lambda_{is}E(W) \tag{27}$$

プロキシ #i を介して要求したウェブアクセスの平均応答遅延時間を $E(D_i)$ とおくと，確率 p_i で応答遅延時間がゼロであるため，次式を得る：

$$E(D_i)=(1-p_i)\left[E(W)+\frac{1}{\mu}\right] \tag{28}$$

注意 4：式(26)における $E(V)$ に，式(25-RB)，(25-ER)を代入して，最終的には式(26)(28)から平均応答遅延時間公式を得る。既存文献（Takahashi 2001）で仮定している「単一プロキシ（$N=1$）・ポアソン到着（$C_{Ai}=1$）モデル」のとき，式(25-RB)，(25-ER)両方とも既存公式と一致していることが分かる。実際，プロキシ番号を表わす添え字（index）i を省略すると

$$E(V)=E(W)=\frac{\lambda(1-p)E(B^2)}{2\left[1-\dfrac{\lambda(1-p)}{\mu}\right]} \tag{29}$$

となる。客平均待ち時間 $E(W)$ が時間平均残余仕事量 $E(V)$ と一致している。即ち，[Wolff (1982)] がマルチンゲールの収束定理を用いて証明した Poisson Arrivals See Time Averages (PASTA) が成り立っていることを確認出来る。

6．おわりに

リアルタイム性を必要とするインターネットコマースにおいて重要なウェブサーバ性能評価法の確立をモチベーションとして，複数プロキシを考慮した一般到着・一般サービス待ち行列モデ

ルを新たに提案した。ラプラス・スティルチェス変換を用いて複数プロキシサーバからの出力過程を特徴付ける統計量（平方変動係数）を解析した。ウェブサーバシステムにおける残余仕事量を拡散過程と見做したモデル化を行ない，上述した出力過程解析と合わせて，システム平均応答遅延時間公式を導出した。本公式が特殊な場合（単一プロキシ・ポアソン到着モデル）に対する既存公式に一致していることを示した。今後の課題としては，Takahashi et al.（2007）の様に，提案した拡散過程モデルの有効性を計算機実験により実証することが挙げられる。また，ウェブでは応答遅延が長いとユーザがリロード（更新）ボタンをクリックしてトラフィック輻輳を齎す。この場合は，髙橋（2004）と同様な再送問題を論考することが必要になる。

謝辞

本研究の一部は，早稲田大学商学部徳井研究振興基金ならびに産業経営研究所リサーチプロジェクトの助成を受けて行われたものである。関係各位に深謝する。

参考文献

Allen, A. O. (1990). *Probability, Statistics, and Queueing Theory with Computer Science Applications.* Academic Press. New York.

Brumelle, S. L. (1971). On the Relationship between Customer and Time Averages in Queues. *J. Appl. Prob.* 8(3): 508-520.

Heyman, D. P. (1975). A Diffusion Model Approximation for the GI/G/1 Queue in Heavy Traffic. *Bell System Technical J.* 54(9): 1637-1644.

伊藤清三（1963）.『ルベーグ積分入門』東京：裳華房.

伊藤清三（1979）.『拡散方程式』東京：紀伊國屋書店.

宮原秀夫・村田正幸（2001）.『インターネットがもたらすマルチメディア社会』大阪大学出版会.

Robertazzi, T. G. (2000). *Computer Networks and Systems: Queueing Theory and Performance Evaluation.* Springer-Veralg. New York.

Ross, S. M. (1996). *Stochastic Processes.* New York: John Wiley & Sons.

鈴木武・山田作太郎（2001）.『数理統計学』東京：内田老鶴圃.

Takahashi, A., Y. Takahashi, S. Kaneda, and N. Shinagawa (2007). Diffusion Approximations for the GI/G/c/K queue. *Proceedings of the 16th IEEE International Conference on Computer and Communication Networks.* 681-686.

Takahashi, Y. (1986). Diffusion Approximation for the Single-Server System with Batch Arrivals of Multi-Class Calls. *IECE Transactions.* J69-A (3): 317-324.

Takahashi, Y. (2001). Performance Modeling a Web Server Access Operation with Proxy Server Caching Mechanism. *The Waseda Commercial Review.* No. 389: 125-137.

髙橋敬隆（2004）.「再呼のある状態依存入力トラヒックモデル：確率解析と情報ネットワークへの応用」『早稲田商学』No. 402: 39-55.

Takahashi, Y. (2005). A Single-Server Queueing System with Modified Service Mechanism: An Application of the Diffusion Process to the System Performance Measure Formulas. *The Waseda Business & Economic Studies.* No. 41: 19-28.

梅田望夫（2006）.『ウェブ進化論』東京：筑摩書房.

吉田耕作・伊藤清三（1976）.『函数解析と微分方程式』東京：岩波書店.

Wolff, R. W. (1982). Poisson Arrivals See Time Averages. *Operations Res.* 30: 223-231.

Wolff, R. W. (1989). *Stochastic Modeling and the Theory of Queues.* New Jersey: Prentice-Hall.