待ち行列理論と情報システム性能評価

笠原　正治

本稿では，情報ネットワークやコンピュータシステムに代表される情報システムに対し，待ち行列理論を用いて性能評価やシステム設計を行うための，対象システムの問題点の捉え方やモデリングについての方法論を紹介する．具体例として，情報通信ネットワークのプロトコル階層の観点から，M/G/1 をベースにした待ち行列モデルを紹介し，情報システムの性能評価に対して待ち行列理論による解析を行う意義と留意点について言及する．

キーワード：待ち行列理論，情報システム，モデリング，性能評価

1. はじめに

待ち行列理論は，サービス施設や情報システムのような共有資源への利用要求が確率的に発生するという仮定の下で，共有資源の競合問題を定量的に解析することを目的として発展してきた [9]．20 世紀初頭，A. K. Erlang が電話交換網の設計問題に対して確率過程を応用した解析を行ったことが待ち行列理論の起源とされ，以降 1 世紀近くにわたり，待ち行列理論はそれぞれの時代に出現した通信・コンピュータのシステム設計や性能評価に関する問題を解決することで発展してきた．

近年ではコンピュータ単体の高性能化やオペレーティング・システム，ソフトウェアの高機能化，さらには情報ネットワークの高速化と無線通信技術の発展により，情報システムは未曾有の規模に巨大化し，ユーザの遍在化が進んで，提供される情報サービスは多様化の一途を辿っている．このような大規模・複雑化した情報システムに対し，構築コストとユーザ満足度のトレードオフを捉えたシステム横断的な性能評価法が益々重要になってきている．

本稿では，大規模かつ複雑な情報システムの構築に向けて待ち行列理論をどのように用いるべきか，という点に焦点を当て，システムの問題点の捉え方やモデリングのアプローチについて解説する．次に，データ通信システムの階層プロトコルを例にとり，各階層の機能に対する基本的な M/G/1 モデルを紹介する．最後にシステム評価に対して待ち行列理論を用いる意義と性能評価を行う際の留意点について言及する．

かさはら　しょうじ
奈良先端科学技術大学院大学・情報科学研究科
〒 630–0192　奈良県生駒市高山町 8916–5

2. 情報システムにおける性能評価

情報システムの性能評価とは，対象とするシステムのハードウェアやソフトウェアの構成が与えられ，システムにかかる負荷条件が与えられたときにシステムの稼働状況やユーザへのサービスの良し悪しを示す性能評価指標を定量的に明示することである [4]．

情報システムの性能を評価する意義として，以下の三点が挙げられる．

1. 定量的な特徴付け (characterization)
2. システムの挙動の理解 (understanding)
3. システムの挙動の予測 (prediction)

情報システムは情報を加工・蓄積・流通することによって，情報を活用するシステムである．そのため，情報の加工や蓄積，流通にどのくらいの時間がかかるのか，エラーはないのか，処理結果の精度はどうか，などについて定量的な評価を行うことで，システムの良し悪しを判断する．また定量的な特徴付けにより，従来システム（従来手法）と新規システム（提案手法）の差異を明確化し，新規システムの優位性を立証することは開発フェーズにおいて重要である．

次にシステムの挙動の理解であるが，一般的にシステムへの処理要求・負荷が増大すると，システムの処理性能は低下する．システムが高負荷になってきたとき，システムの稼働状況はどのようになっているのか，特にシステム全体の性能を低下させている箇所はどこなのか，処理のオーバーヘッドはないか，といったボトルネック部分を設計段階で明らかにすることは重要である．システムへの負荷条件以外にも，CPU の処理性能や通信回線速度，メモリ容量やバッファサイズといった，システム構成要素のパラメータがシステム全体に与える影響を定量的に把握すること，すなわち性

能評価量に対する設計パラメータの感度を把握することも重要である．

三点目の挙動の予測は，情報システムの最適な構成を検討する上で重要である．一般に情報システムは，コストや物理的，政策的な制約の下で，ユーザが満足するような情報サービスを提供できるように構築されなければならない [8]. 一旦システムが構築されてサービスが始まると，システムの構成をすぐに変更することは容易にはできない．そのため，システムの利用要求（需要）を推測し，想定した需要の下でシステムの性能がどうなるのか，ということをできるだけ正確に見積もって，コスト的にも性能的にも最適なシステム構成を検討する必要がある．

情報システムに対する性能評価の良し悪しは，上記三項目をどれだけ反映させられるかにかかっていると言えよう．

3. モデリングのアプローチ

情報システムを構成する要素はコンピュータ機器からソフトウェア，ネットワーク機器に至るまで多種多様であり，それらが組み合わさったシステム全体を定量的に特徴付けることは大変難しい．そのため，一般的には全体のシステムを評価しやすいサブシステムに分解し，サブシステムを評価してから全体の挙動を評価する，というアプローチがとられる [5].

サブシステムへの分割方法としては，コンピュータやネットワーク・スイッチ，ルータといった機器単位の分割方法だけでなく，機器を構成する CPU やメモリなどのハードウェア，オペレーティング・システムのような基本ソフトウェア，さらにはユーザが使用するアプリケーションソフトウェアといった，機器自体の構成要素や機能面にも着目した分割を考える必要がある．

システムをサブシステムに分割する方法の一つに，マクロレベルからミクロレベルに分解する階層モデリングの方法がある [5]. 階層的なモデルを構築する際，マクロなシステムとミクロなサブシステムの相互作用（インタラクション）や依存関係が少ないほど，二つのシステムを独立なものとして扱いやすい．

マクロレベルとミクロレベルの境界の目安として，制御の時間スケールが挙げられる．ミクロなサブシステムの処理時間単位が，マクロなシステムの処理時間スケールと比べて十分に小さければ，ミクロ・マクロ間で相互に作用する事象の頻度は少なくなり，ミクロレベルのサブシステムをマクロレベルのシステムと独立なものと見なせるようになる．

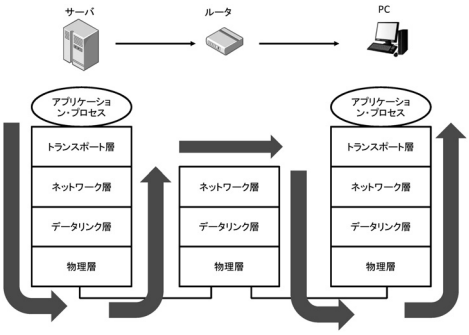

図 1　インターネット上のデータ通信

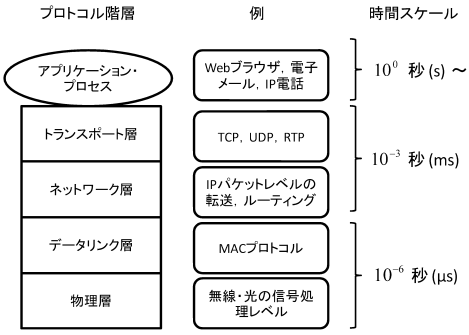

図 2　プロトコル階層と時間スケール

データ通信ネットワークは，通信機能の階層構造が明確に定義され，サブシステムとしての観察が容易なものとなっている．図 1 で示されるインターネット上のデータ通信サービスを例として見ていこう．

インターネットのデータ通信には 4 階層のモデルが用いられ，上から順にトランスポート層，ネットワーク層，データリンク層，物理層の四つに分類される．図 2 にデータ通信におけるプロトコル階層と関連するアプリケーションや機能を示す．Web ブラウザや電子メールはトランスポート層を窓口としてデータの送受信を行っている．ネットワーク層ではインターネット・プロトコル (IP) により，自律分散的なパケット転送機能を提供している．データリンク層では隣接する機器間（図 1 ではサーバ・ルータ間およびルータ・PC 間）の通信サービスを提供し，イーサネットや無線 LAN の制御がこの部分に相当する．物理層ではビット情報を光の波長や無線周波数帯域上に乗せて通信するための信号処理が行われる．各プロトコルの詳細な機能については例えば [10] を参照されたい．

サーバからユーザ PC へのデータ通信フローを定量的に評価するために，個々のプロトコル階層に着目してモデル化することを考える．四つの階層はそれぞれ

独立した通信機能を持っているが，ある階層の機能が実行されるとその影響は他の階層にも影響を与えることに注意しよう．例えばデータリンク層で Go-Back-N や Selective Repeat のような誤り再送制御 (ARQ) が働いているとき，物理層でビットエラーによるパケットロスが発生すると，リンク層の ARQ が機能して廃棄パケットを再送する．これはネットワーク層から見ると IP パケットの転送時間が長くなることに相当する．同様に，アプリケーション・プロセスがトランスポート層プロトコルとして TCP を使って通信をしているとき，ルータで IP パケットが廃棄されると，TCP が廃棄されたデータ・セグメントを再送する．アプリケーション・プロセスにはこの再送制御が一つのセグメントの送信に時間がかかったように見える．このように，あるプロトコル階層の機能が他の階層に影響を与えるため，プロトコル階層間の依存性を注意深く観察してモデルを考える必要がある．

図 2 の右側には，各処理のおよその時間スケールを示している．リンク層のレベルではおよそマイクロ秒オーダ，ネットワーク層の経路制御やトランスポート層レベルではミリ秒のオーダ，アプリケーションレベルでは秒から分のオーダである[1]．隣接するプロトコル階層間の時間スケールはおよそ 100 倍から 1,000 倍ほど異なっているが，ある程度の従属関係があるため，この部分をどのようにモデルに取り込むかがモデリングの鍵となる．4.1 節で，そのようなモデルの例を取り上げる．

4. 情報ネットワークの確率モデル

ここでは，プロトコルの主な機能の基本特性を理解するという観点から，遅延の評価に特化した待ち行列モデルを紹介する．プロトコル階層はデータリンク層，ネットワーク層，トランスポート層の三階層に着目し，待ち行列モデルは基本的な M/G/1 に限定して話を進めることにする．

4.1 データリンク層の M/G/1 モデル

データリンク層は，隣接するノードに対してデータ・パケットを送信するためのプロトコルであり，パケットは一時的にノードのバッファに蓄積されて転送が行われる．ここではパケットの遅延を待ち行列で評価することを考えよう．

[1] 無線 LAN の IEEE 802.11n（データリンク層相当）の最小制御時間スロットは 9 マイクロ秒，TCP では OS にも依存するが設定可能パラメータの時間単位はミリ秒となっている．

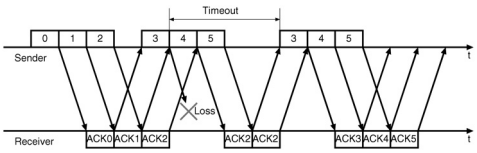

図 3 Go-Back-N 誤り再送制御

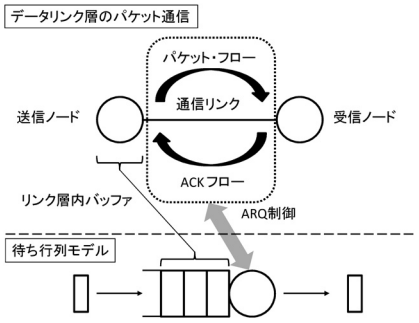

図 4 ARQ 制御と待ち行列モデルの対応関係

今，データリンク層へのパケットの到着が率 λ のポアソン過程に従っており，パケットの転送時間 S が独立同一な確率分布に従うものとする．データリンク層の送信バッファ容量に制限がないと仮定すると，送信バッファ部分は M/G/1 待ち行列としてモデル化できる．このとき，パケットのバッファ内待ち時間 W の平均は次のポラチェック・ヒンチンの平均値公式で与えられる．

$$E[W] = \frac{\lambda E[S^2]}{2(1-\lambda E[S])} \tag{1}$$

S はパケットの転送時間を表す確率変数であり，データリンク層の制御を S の確率分布にいかにして反映させるか，という点がモデル化のポイントとなる．

今，パケットロスに対する誤り再送制御がデータリンク層で機能している場合を考えよう．ここでは文献 [1] の Go-Back-N 誤り再送制御の遅延評価を紹介する．Go-Back-N では，送信ノードはパケットを連続して N 個送出し，受信ノードはパケットを受信すると対応する送達確認応答 (ACK) パケットを送信ノードに返送する．もしパケットロスが発生すると，Go-Back-N では廃棄パケットから送信をやり直す．図 3 は $N = 3$ のときの Go-Back-N の挙動を表している．

図 3 のようなパケット送信メカニズムを単一サーバ待ち行列でモデル化したものが図 4 である．送信ノードのデータリンク層に着目し，リンク層から物理層に向けた出力バッファを待合室，サービス時間はパケットを送出してから再送制御を経て送信が成功するまで

の時間，一度に送信されるパケットの数は一つ，といった特徴をモデル化したものになっている．3 章の最後で紹介した下位層の挙動を考慮したモデルになっていることに注意してほしい．

ここで以下の仮定を置く．

- パケット長は一定で 1 パケットの伝送時間を単位時間（スロット）とする．
- パケットは率 λ のポアソン過程に従って送信側に到着する．
- 受信側では受け取ったパケットに誤りがあればそのパケットを廃棄する．
- 伝送路でパケットに誤りが発生する確率を p とし，伝送誤りは互いに独立に発生する．
- 送信側でパケットの伝送が終了してから $n-1$ スロット経過してもその ACK が返ってこない場合は，そのパケットの再送を行う．

このように仮定すると，一つのパケットが送信に成功するまでの時間 S は次式で与えられる．

$$\Pr\{S = 1 + kn\} = (1-p)p^k, \quad k = 0, 1, \ldots \quad (2)$$

ここで k はパケットの再送回数である．(2) 式より，S の平均と 2 次モーメントは

$$E[S] = 1 + \frac{np}{1-p} \quad (3)$$

$$E[S^2] = 1 + \frac{2np}{1-p} + \frac{n^2p(1+p)}{(1-p)^2} \quad (4)$$

で与えられ，データリンク層でのバッファでの平均送信待ち時間は，(3)，(4) を (1) 式に代入することで，陽な形で得ることができる．

文献 [7] では，Stop-and-Wait, Go-Back-N, Selective-Repeat, の三種類の ARQ 方式に対して M/G/1 モデルを用いた遅延解析を行い，パケットのラウンド・トリップ時間や誤り発生確率が遅延に与える影響について比較評価を行っている．

4.2 ネットワーク層の M/G/1 モデル

ネットワーク層は，複数の中間ノード（ルータ）を経てパケットを目的ノードに送り届ける経路制御の役割を担っている．パケットはルータ内のバッファに蓄積され，適切な出力リンクに向けて送出される．この部分の基本的な遅延評価は，前節で紹介した M/G/1 のポラチェック・ヒンチンの平均値公式で行うことができる．ここではネットワーク層で通信品質の差別化を図るしくみが提供されている場合を考えよう．

インターネットで通信品質を保証するための技術として，DiffServ と呼ばれるサービス差別化技術が提案されている．DiffServ のサービスクラスは優先転送 (Expedited Forwarding: EF)，帯域保証転送 (Assured Forwarding: AF)，ベストエフォートの三種類が定義されており，優先転送クラスのパケットは他クラスのパケットよりも優先してパケットの転送処理が行われる．帯域保証転送クラスはさらに複数のサブクラスが定義されており，それらに対して重み付け公平待ち行列 (Weighted Fair Queueing: WFQ) による転送処理が行われる．ベストエフォートクラスは何も制御が行われないクラスである（図 5 参照）．

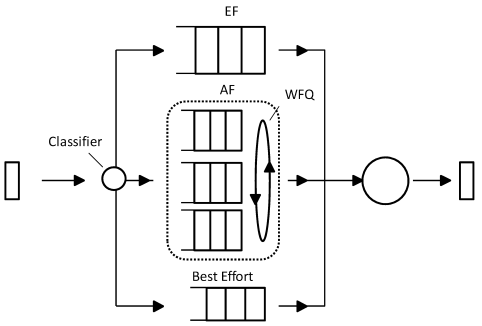

図 5 DiffServ によるパケット優先転送処理

三種類のサービスクラスに対する遅延を評価するモデルとして，非割込み優先権サービス規範を持つ M/G/1 を考えることができる．今，客は n 個の優先クラスに分類され，$i < j$ のとき，クラス i の客はクラス j よりも高優先であると仮定する．クラス k $(= 1, 2, \ldots, n)$ の客は率 λ_k のポアソン過程に従ってシステムに到着し，クラス k の客のサービス時間 S_k は独立同一な一般分布に従うものと仮定する．非割込み優先権サービスでは，現在サービス中の客よりも高優先の客が到着したとしても，現在の客のサービスは割り込まれない．

このとき，高優先クラス 1 の待ち時間 W_1 の平均は

$$E[W_1] = \frac{\sum_{i=1}^{n} \lambda_i E[S_i^2]}{2(1 - \lambda_1 E[S_1])} \quad (5)$$

で与えられ，優先クラス k $(= 2, 3, \ldots, n)$ については次式で与えられる．

$$E[W_k] = \frac{\sum_{i=1}^{n} \lambda_i E[S_i^2]}{2(1 - \sum_{i=1}^{k-1} \rho_i)(1 - \sum_{i=1}^{k} \rho_i)} \quad (6)$$

ただし $\rho_i = \lambda_i E[S_i]$ である．

DiffServ での EF クラスは一番高い優先権を持っているため，EF クラスの遅延は (5) 式で評価でき，また AF クラスおよびベストエフォートクラスのパケットは (6) 式で遅延を見積もることができる．

文献 [6] では，DiffServ サービスを提供するルータでの遅延について，ここで紹介した形の遅延解析を行っている．また，文献 [11] では，DiffServ の EF クラスとベストエフォートクラスに着目し，ネットワーク全体の容量設計問題の遅延制約条件に非割込み優先権付き M/G/1 の結果を利用している．

4.3 トランスポート層の M/G/1 モデル

トランスポート層はアプリケーションがネットワーク通信を行うための窓口として機能するプロトコルであり，インターネットでは TCP がエンド・ホスト間で高信頼な通信を提供する役目を持っている．TCP はネットワークの混み具合を推測してパケットの送信速度を適応的に調節する輻輳制御の機能を有している．TCP の輻輳制御の詳細については [10] を参照されたい．

TCP の輻輳制御は，通信リンクの帯域を公平に配分するように機能することが知られている．この特徴に着目した複数コネクションの帯域共有モデルとして，プロセッサシェアリングサービス規範を持つ M/G/1 がある．プロセッサシェアリングでは，系内に n 人の客が存在するとき，サーバはそれらの客に対して $1/n$ の能力でサービスを同時に提供する．

今，客の到着が率 λ のポアソン過程に従い，サービス時間 S の平均が $E[S]$ で与えられるプロセッサシェアリング M/G/1 待ち行列を考える．系の安定条件 $\lambda E[S] < 1$ が成立するとき，系内客数 N の定常分布はサービス時間 S の分布には依存せず，平均 $E[S]$ のみで定まり，M/M/1 と同じ次式の幾何分布で与えられることが知られている [8]．

$$\Pr\{N = k\} = (1-\rho)\rho^k, \quad k = 0, 1, \ldots$$

ここで $\rho = \lambda E[S]$ である．これより，平均系内客数は

$$E[N] = \frac{\rho}{1-\rho}$$

で与えられ，客の滞在時間 T の平均はリトルの公式より次式で与えられる．

$$E[T] = \frac{E[N]}{\lambda} = \frac{E[S]}{1-\rho} \tag{7}$$

このように，系内客数分布がサービス時間分布の平均のみに依存し，分布には依存しない性質のことを，サービス時間分布に対する不感性 (insensitivity) という．不感性の性質は M/G/c/c や M/G/∞ の系内客数分布にも見られることが知られている．

TCP のフローレベルの性能評価として，図 6 のネットワークを考えよう．図 6 はダンベルモデルと呼ばれ，

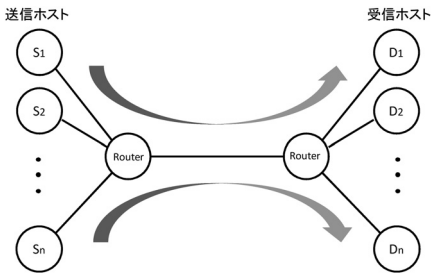

図 6 ダンベル・トポロジー・ネットワーク

TCP の挙動を調べる際の基本モデルとして広く利用されている．図 6 では，n 台の送信ホストからの TCP フローが左のルータに集約され，1 本の通信リンクを介して右のルータから受信ホストに送られている状況を表している．

今，TCP のコネクション要求が率 λ のポアソン過程に従って発生し，送信するデータ量の平均が b ビット，通信リンクの回線速度を C ビット毎秒と仮定する．1 本の TCP フローが回線を占有するときのコネクション持続時間がサービス時間 S に相当し，その平均は $E[S] = b/C$ で与えられる．TCP による帯域の共有状況をプロセッサシェアリング M/G/1 でモデル化すると，$\lambda b/C < 1$ のとき，TCP フロー 1 本当たりの持続時間 T は (7) 式より

$$E[T] = \frac{b}{C - \lambda b}$$

で見積もることができる．

ここではもっとも単純な単一ボトルネックリンクに対する TCP フロー競合モデルを紹介したが，TCP の挙動やネットワークの環境を考慮した拡張モデルについては [3] を参照されたい．プロセッサシェアリング待ち行列は資源を公平かつ同時に共有するようなシステムのモデル化に適しており，他の応用例として Web サーバシステムがある．詳しくは [5] および [5] に挙げられている参考文献を参照されたい．

5. ネットワークシステム性能評価の意義

前節では M/G/1 の応用を主眼においてモデルを紹介したが，モデルの仮定は現実の状況やシステムの動作を正確に反映していないため，本当に役に立つのかという疑問が残る．前章の M/G/1 モデルに対して例えば以下の不備が挙げられよう．

- バッファ容量は有限である．
- パケットの到着はポアソン過程ではない．

- ARQ 再送制御の再送回数は有限である．
- リンクの伝送エラーはバースト的である．
- TCP ウィンドウフロー制御のアルゴリズムが反映されていない．

以下では，待ち行列理論が役に立つ性能評価とはどういうものか，という点について考えてみたい．

5.1 待ち行列理論の特徴

待ち行列理論は，入出力システムを確率モデルとして捉え，性能評価量を解析する手法である．具体的には客の到着過程やサービス時間の変動を考慮してシステムの状態推移を支配するエルゴード的マルコフ連鎖を構築し，客数分布や待ち時間分布を解析する．

待ち行列理論を用いた情報システムの性能評価の利点としては，性能評価量の平均や高次モーメントの真値が得られること，および導出結果は数式や数値計算アルゴリズムで与えられるので結果の再現性が高いことが挙げられる．もし性能評価量の解析結果が陽な形で導出されていれば，システムパラメータが性能評価量に与える影響を把握することもできる．一方，待ち行列理論の欠点としては，待ち行列モデルの記述能力が低いため，対象システムの細部をそぎ落とした抽象度の高いモデルを構築せざるを得ないこと，定常分布や極限分布の導出が中心のため，システムの過渡的な特性を把握することが難しいこと，などが挙げられる．

待ち行列理論が取り扱う確率過程はエルゴード的マルコフ連鎖という限られたクラスであることに注意すれば，2 章で挙げた意義の中でも「挙動の予測」という点で待ち行列理論が性能評価で役立つのは非常に稀であると言わざるを得ない．予測の意味でシステムの設計に役立ったのは，回線交換網におけるアーラン呼損式であろう[2]．しかしながら，インターネットに見られるパケット交換網においては，マルコフ的な到着過程ではパケットトラヒックの持つ相関構造を捉えることができないため，高精度な予測を行える待ち行列モデルを構築することは非常に難しい．

このような点を踏まえると，待ち行列理論を用いて情報システムの性能を評価をする意義として次の三点が考えられるであろう．

5.1.1 モデリングを通しての対象の本質の理解

対象をエルゴード的マルコフ連鎖を用いて評価する場合，マルコフモデルの作成が極めて重要である．マルコフ連鎖の確率的挙動を特徴付けるのは推移確率・推移率であり，対象の特徴や挙動のモデル化を通して推移確率・推移率を慎重に決定する必要がある．この際，対象の確率的挙動にある程度のマルコフ性を仮定することが必要になってくるため，モデルを構成する際には，マルコフ性の仮定を大雑把に行うのではなく，対象の挙動を詳細に分析した上でマルコフ性の仮定を行う必要がある．この分析を行うことは，対象の本質をより深く理解することにつながっている．

5.1.2 対象の定性的性質の把握

対象の確率的挙動が本質的にマルコフ的ではない場合，対象を詳細に分析して慎重に構成したマルコフモデルをもってしても，得られる結果は定量的な意味では一致しない．しかしながら，対象の挙動を支配する主要因パラメータをモデルに取り込んでいれば，そのパラメータが性能に与える影響を定性的に把握することが可能である．負荷に対する性能評価量の変化や，有限容量待ち行列系における棄却率に対する容量の感度などが代表的例として挙げられる．

5.1.3 対象の平均的挙動の把握

一般に，システムのサービス能力に対してユーザの需要が相対的に小さいとき，または変動が大きくても長期の挙動を観察するとき，性能評価量の平均が予測値として使える可能性が高くなる．大規模・複雑な情報システムを構築する初期段階においては，月・年単位の中長期的な平均的性能を見積もってシステム構成要素のスペックを決定する必要があり，このようなときには待ち行列理論のアプローチが有用である．

5.2 性能評価の目的の変遷

図 7 は，対象システムの認知度と待ち行列理論で扱うマルコフモデルの有用性が，時間の経過とともにどのように変化するのかを概念的に表したものである．この図では，時間軸の原点で対象システムの概念が生まれ，時間とともに研究開発が進み（時間軸の中央付近），製品となって世の中に出回る（時間軸の右側）様子を表している．対象システムに対する認知度は時間の原点近くでは極めて小さく，時間とともに認知度・

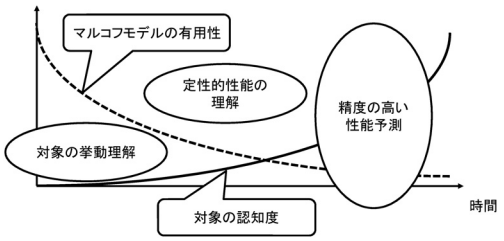

図 7 マルコフモデルの有用性

[2] 予測に適したのは M/G/c/c のもつサービス時間分布に対する不感性に依るところが大きい．情報通信網におけるアーラン B 式の詳細については [9] を参照されたい．

理解度が増大していく．このように見ていくと，性能評価の目的も時間とともに以下のように変化していくことが考えられる．

- 対象システムの黎明期：システムの挙動の理解
- 対象システムの開発初期：どのパラメータが性能に影響を与えるのか，という定性的な性質の理解
- 対象システムの開発後期：最適設計に向けた高精度な性能予測

このように考えると，基本的な待ち行列モデルはシステムの黎明期に近いときほど有用であり，開発が進むにつれて精密なモデルが求められるようになることも自然に理解できる．しかしながら，解析可能であれば，複雑なモデルでも OK というわけでもない．以下に待ち行列モデルを構築する際に留意すべき事項を記す．

- 性能評価量を計算する際の計算量が莫大であると，時間消費型のシミュレーションに対する優位性がなくなるため，計算量が少なくなるようなモデルを構築すべきである．
- 待ち行列モデルの適用領域を明確にする．例えば遅延評価で無限バッファモデルを用いるとき，どの程度の負荷のときに有効か，というように，モデルを構築する際に設けた仮定がどのようなときに妥当であるかについては注意を払う必要がある．

6. まとめ

本稿では，情報システムに対して待ち行列理論を用いた性能評価やシステム設計を行うための，対象システムの問題点の捉え方やモデリングについての方法論を概説した．また，情報ネットワークのプロトコル階層に対応した基本的な M/G/1 待ち行列モデルを紹介し，待ち行列理論を性能評価で用いる際の意義についても著者なりの見解を紹介させていただいた．

情報システムの性能評価研究で待ち行列理論を用いるときの留意事項を以下に補足する．

- 待ち行列モデルの解析自体に困難さがあるか？
- 非常に興味のある対象に対して待ち行列モデルを考えているか？
- 性能評価の数値例はシミュレーションで得られる結果よりも示唆に富んだものとなっているか？

上記はすべて必要というわけではなく，少なくとも一つは満足する必要があるという意味で捉えていただきたい．

近年情報ネットワークではパケットレベルのサービス品質 (Quality of Service: QoS) から，ユーザ自身の体感品質 (Quality of Experience: QoE) が重要視されるようになってきた．また，コンピュータシステムでは分割・処理・統合の繰り返しによる大規模タスクのワークフロー評価が強く求められるようになってきている．言い換えると，「情報」の処理の流れ，処理フローに対する評価が強く求められるようになってきており，そこには単純な「待ち」がない．情報システムに対する性能評価には，これまで以上に対象システムに対するモデル化の工夫と新しい理論成果の応用が求められており，この分野の今後の発展が期待される．

最後に待ち行列理論とシステムの性能評価に関する最近の書籍を紹介して結びとさせていただきたい．文献 [5] は，情報システムの性能評価を念頭に待ち行列理論の網羅的な紹介を行っている良書であり，対象を待ち行列でモデル化する際に役立つ内容が，豊富な例とともに記されている．[2] はコンピュータシステムの性能評価に向けた待ち行列モデルを解説した書籍であり，データセンターに見られる複数サーバ待ち行列やジョブのスケジューリングについて豊富な解説が記載されている．

参考文献

[1] D. Bertsekas and R. G. Gallager, *Data Networks 2nd Ed.*, Prentice Hall, 1992.

[2] M. Harchol-Balter, *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*, Cambridge University Press, 2013.

[3] 石橋圭介，川原亮一，「TCP フロー制御における帯域共有のモデル」，オペレーションズ・リサーチ，**49** (2004), 438–442.

[4] 紀一誠，「情報システムの性能評価 (1)〜(3)」，オペレーションズ・リサーチ，**40**, 6 月号–8 月号 (1995).

[5] H. Kobayashi and B. L. Mark, *System Modeling and Analysis: Foundations of System Performance Evaluation*, Pearson Education, 2008.

[6] H. Lee, “Anatomy of Delay Performance for the Strict Priority Scheduling Scheme in Multi-Service Internet,” *Computer Communications*, **29** (2005), 69–76.

[7] 宮原秀夫，尾家祐二，『コンピュータネットワーク』，共立出版，1999.

[8] 滝根哲哉，伊藤大雄，西尾章治郎，『ネットワーク設計理論』，岩波書店，2001.

[9] 滝根哲哉，村田正幸，「通信網における待ち行列―理論の応用と課題―」，オペレーションズ・リサーチ，**43** (1998), 264–271.

[10] アンドリュー・S・タネンバウム，デイビッド・J・ウェザロール，『コンピュータネットワーク 第 5 版』，日経 BP 社，2013.

[11] K. Wu and D. S. Reeves, “Capacity Planning of DiffServ Networks with Best-Effort and Expedited Forwarding Traffic,” *Telecommunication Systems*, **25** (2004), 193–207.