数理解析研究所講究録
947巻 1996年 88-98

遅延に依存した割り込み優先権をもつM／G／1待ち行列システム

Delay dependent preemptive priority rules for M/G/1 queues

牧本　直樹　　白木　宏明
Naoki Makimoto　　Hiroaki Shiraki

東京工業大学　数理・計算科学専攻

音声，ファクリミリ，ファイル，動画など特性が異なるデータを単一サーバで効率よく処理するには，データの種類に応じて割り込みや非割り込みの優先権を与える必要があるが，基本的な優先方式で達成できる評価尺度（待ち時間など）の範囲はかなり限定されている．したがって，各クラスの評価尺度に応じて定まるコストを最小化したい場合，そのような優先方式だけでは最適な評価尺度が実現できない可能性が高い．本稿では，まず滞在時間に応じて優先度が決まるような割り込み優先方式をもつM/G/1待ち行列を考え，その近似解析を行う．また，その優先方式で達成可能な評価尺度の集合を調べ，最適化問題への適用について考察する．

1. はじめに

電話の音声や，ファクリミリ，ファイル転送のように，耐遅延，耐損失特性が異なるデータを単一のサーバで処理するシステムを効率よく稼働させるには，客の特性に応じた優先方式を行う必要がある．そのため，客をいくつかのクラスに分けて上位クラスの客に優先権を与える方式が考えられている．代表的な優先方式としては，先着順，割り込み優先方式，非割り込み優先方式，プロセッサーシェアリングなどが挙げられる．しかし，これらの優先方式で達成することができる評価尺度の範囲はかなり限定されている．例えば，J 個の客のクラスがある待ち行列システムで割り込み優先方式を行う場合，達成可能な平均待ち時間ベクトルは優先権の順列と同じ $J!$ 通りしか存在しない．そのため，各クラスの客の評価尺度（平均待ち時間や平均滞在時間）に応じて定まるコストを最小化する問題を考える場合，上述した基本的な優先方式だけでは最適な評価尺度ベクトルが達成できない可能性が高い．

一方で，基本的な優先方式以外にもさまざまな優先方式が提案・解析されており，本論文ではその中の1つである遅延に依存した優先権を扱う．これは，滞在時間が長くなるとクラス応じた上昇率（優先度係数）で優先度が上がる，という方式でこれまでにもいくつか研究がなされている [1, 2, 5, 7, 9, 11]．この方法では，各クラスの優先度の上昇率を変えることによって評価尺度ベクトルが変化するため，基本的な優先方式に比べて達成可能な評価尺度空間が大幅に広がることが期待される．実際，Federgruen and Gronenbelt [2] では，遅延に依存した非割り込み優先権をもつ M/G/1 について，平均待ち時間ベクトルと優先度係数ベクトルが1対1に対応することを示し，平均待ち時間ベクトルが与えられたときにそれを達成する優先度係数を求めるアルゴリズムを構築している．

本論文では [2] のモデルに割り込みを許す場合，つまり遅延に依存した割り込み優先権をもつM/G/1待ち行列を考える．まず次節でモデルを説明し，3節で平均待ち時間や平均総サービス時間に対する近似式を導出する．4節では，あるケースでは得られた近似式が厳密解と一致することを示し，さらにシミュレーションとの比較による精度の検証を行う．最後の5節では，この優先方式で達成可能な評価尺度の集合を数値的に調べ，最適化問題への適用について考察を行う．

2. 遅延に依存した割り込み優先権をもつ M/G/1

J クラスの客を単一サーバで処理する $M/G_1,\cdots,G_J/1$ 待ち行列を考える．以下，クラス p を C_p と書くことにする．C_p の客は到着率 λ_p のポアソン過程に従って到着し，一般分布に従うサービス時間だけサービスを受ける．C_p の客のサービス時間の生成確率変数を S_p とし，$E[S_p]=1/\mu_p$, $E[S_p^2]<\infty$ と仮定する．各クラスの客の到着過程およびサービス時間はすべて互いに独立であると仮定する．C_p の客のトラヒック密度を $\rho_p=\lambda_p/\mu_p$，システム全体のトラヒック密度を $\rho=\sum_{p=1}^{J}\rho_p$ と書く．以下ではシステムが安定である，すなわち $\rho<1$ である場合を考える．

本論文で扱う優先方式は次の通りである．時点 τ でクラス p の客が到着したとき，この客の優先度関数を

$$q_{(t)}=\alpha_p(t-\tau) \tag{1}$$

によって定める．サーバは，各時点 t においてシステムに滞在している客の中から優先度関数の値が最大の客の処理を行う．もし，ある客のサービス中に優先度関数が逆転した場合は割り込みが可能で，より優先度関数の大きな客の処理を先に行い，割り込まれた客は自分より優先度関数の大きな客がすべて退去した後，残りの処理を再開する．以下では一般性を失うことなく，添字の小さなクラスの客が高い優先度をもつ，つまり $\alpha_1\geq\alpha_2\geq\cdots\geq\alpha_J\geq 0$ とする．また $\alpha_p=0$ の場合は，他のクラスの客は C_p の客に対して通常の割り込み優先権をもつことになる．

3. 近似解析

本論文で扱うモデルのように，優先権が時間の関数として動的に変化する待ち行列システムには，サービス時間分布，割り込みの有無，優先度関数の形状などによっていくつかのバリエーションがある．到着間隔とサービス時間がともに指数分布に従う場合は [1, 5, 7, 10] で研究されており，各クラスの平均待ち時間などが求められている．また [2] では，サービス時間が一般分布に従う非割り込み優先モデルを解析し，(1) と同じ線形の優先度関数について平均待ち時間を導出している．本論文のモデルは [2] のモデルで割り込みを許すようにしたものであるが，厳密な解析は困難であるため，以下では近似解法を提案する．

本論文のモデルでは割り込みを許すため，サービスを開始してから退去するまでの時間とサービス時間は必ずしも一致しない．そのため，実際にサービスを受けている時間を**サービス時間**，サービスを開始してから退去するまでの時間を**総サービス時間**と呼んで区別することにする．また，到着してから初めてサービスを受けるまでの時間を**待ち時間**と呼ぶ．サービスを始めてから割り込まれることによって新たに待つ時間は，総サービス時間に含まれ待ち時間には含まれない．定常状態における C_p の客の総サービス時間，待ち時間を表す確率変数をそれぞれ R_p, W_p と書く．

3–1. 平均総サービス時間

時点 0 で C_p の客 A が到着したとする．$I_0^{(p)}$ を A のサービス時間とし，$I_{0i}^{(p)\prime}$ を $I_0^{(p)}$ の間に A に割り込むことのできる C_i の客の到着時刻の範囲の長さとする．また，$I_0^{(p)}$ の間に割り込む C_i の客のサービス時間の合計を $I_{1i}^{(p)}$ とし

$$I_1^{(p)}=\sum_{i=1}^{p-1}I_{1i}^{(p)}$$

とする．この段階で A の総サービス時間は $I_1^{(p)}$ だけ増え，そのため新たに A に割り込む客が生じる．それらの客による A の総サービス時間の増加分は，

$I_{ji}^{(p)'}$: $I_j^{(p)}$ の間に割り込むことのできる C_i の客の到着時刻の範囲の長さ

$I_{ji}^{(p)}$: $I_{j-1}^{(p)}$ の間に割り込んだ C_i 客のサービス時間の合計

とおくと

$$I_j^{(p)} = \sum_{i=1}^{p-1} I_{ji}^{(p)}$$

によって再帰的に与えられる．したがって，求めたい R_p は

$$R_p = \sum_{j=0}^{\infty} I_j^{(p)} \tag{2}$$

となる．図 1 より，$\alpha_p I_j^{(p)} = \alpha_i (I_j^{(p)} - I_{ji}^{(p)'})$ が成り立つから，

$$I_{ji}^{(p)'} = \left(1 - \frac{\alpha_p}{\alpha_i}\right) I_j^{(p)} \tag{3}$$

を得る．客の到着はポアソン過程に従っているから，$I_{j-1,i}^{(p)'}$ の間に到着する C_i の客数の平均は $\lambda_i I_{j-1,i}^{(p)'}$，それらの客のサービス時間の合計の平均は $\lambda_i I_{j-1,i}^{(p)'} \times 1/\mu_i = \rho_i I_{j-1,i}^{(p)'}$ となる．これと (3) より

$$E[I_{ji}^{(p)} | I_{j-1}^{(p)}] = \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i I_{j-1}^{(p)}$$

を得る．両辺の平均をとると

$$E[I_{ji}^{(p)}] = \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i E[I_{j-1}^{(p)}] \tag{4}$$

となる．$I_{0i}^{(p)'}$ の間に到着する C_i の客数を N_i とすると，$E[I_1^{(p)}] = \sum_{i=1}^{p-1} E[S_i]E[N_i]$ となり

$$\begin{aligned} E[I_j^{(p)}] &= \sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i E[I_{j-1}^{(p)}] = \left\{\sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i\right\}^{j-1} E[I_1^{(p)}] \\ &= \left\{\sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i\right\}^{j-1} \sum_{i=1}^{p-1} E[S_i]E[N_i]. \end{aligned} \tag{5}$$

A のサービス開始前の C_i の到着がポアソン過程に従っていると仮定すると $E[N_i] \simeq \lambda_i E[I_{0i}^{(p)'}]$ と近似できて，これを (5) に代入すると

$$\begin{aligned} E[I_j^{(p)}] &= \left\{\sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i\right\}^{j-1} \sum_{i=1}^{p-1} E[S_i]E[N_i] \simeq \left\{\sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i\right\}^{j-1} \sum_{i=1}^{p-1} \rho_i E[I_{0i}^{(p)'}] \\ &= \left\{\sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i\right\}^{j} E[S_p] \end{aligned} \tag{6}$$

となる．(2) の両辺の平均をとり (6) を代入し $\sum_{i=1}^{p-1}(1 - \alpha_p/\alpha_i)\rho_i < 1$ が成り立つことを考慮すれば，

$$E[R_p] = \sum_{j=0}^{\infty} E[I_j^{(p)}] \simeq \sum_{j=0}^{\infty} \left\{\sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i\right\}^{j} E[S_p] = \frac{E[S_p]}{1 - \sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i} \tag{7}$$

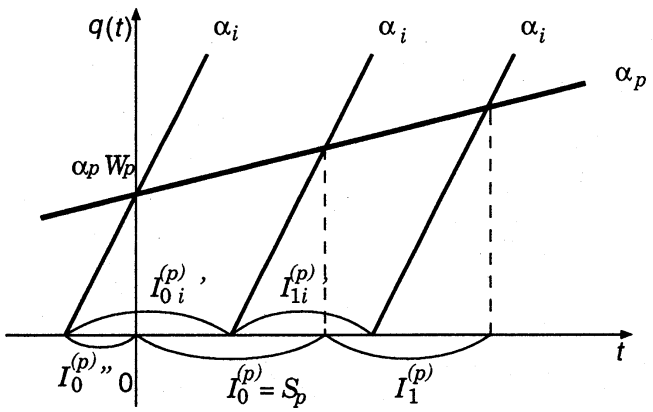


図 1: C_pの客と C_iの客の優先度関数 $q(t)$

となる.

3–2. 平均待ち時間

C_p の客 B が時点 0 に到着したとして，$U_0^{(p)}$ を B の到着以降新たな到着がないと考えた場合の待ち時間とする．また，$U_{0i}^{(p)'}$ を $U_0^{(p)}$ の間に割り込むことのできる C_i の客の到着時刻の範囲の長さとし，$U_0^{(p)}$ の間に割り込む C_i の客のサービス時間の合計を $U_{1i}^{(p)}$ とする．

$$U_1^{(p)} = \sum_{i=1}^{p-1} U_{1i}^{(p)}$$

は $U_0^{(p)}$ の間に割り込む客のサービス時間の合計を表す．この段階で B の待ち時間は $U_0^{(p)}$ だけ増加し，そのため B の後に到着して先にサービスを受ける客が発生する．それらの客による B の待ち時間の増加分は

$U_{ji}^{(p)'}$: $U_j^{(p)}$の間に割り込むことのできるC_iの客の到着時刻の範囲の長さ

$U_{ji}^{(p)}$: $U_{j-1}^{(p)}$の間に割り込んだC_i客のサービス時間の合計

として

$$U_j^{(p)} = \sum_{i=1}^{p-1} U_{ji}^{(p)} \tag{8}$$

によって再帰的に与えられる．したがって，求めたい W_p は

$$W_p = \sum_{j=0}^{\infty} U_j^{(p)} \tag{9}$$

となる．E[R_p] を求めたときと同様の方法を用いると

$$E[U_j^{(p)}] = \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i E[U_{j-1}^{(p)}]$$

が成り立つから，(8) を用いて再帰的に計算すると

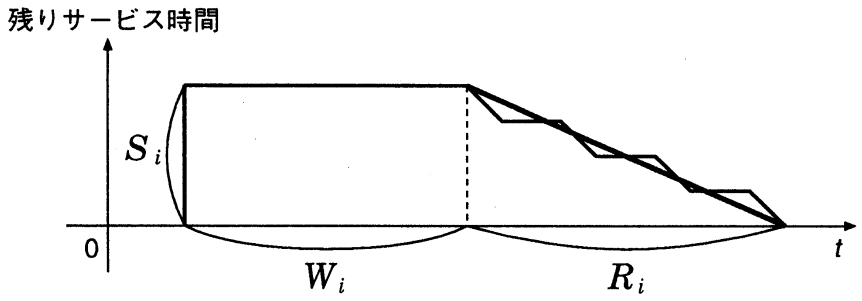


図 2: 客の 1 人あたりのシステムに持ち込む仕事量

$$E[U_j^{(p)}] = \sum_{i=1}^{p-1}\left(1-\frac{\alpha_p}{\alpha_i}\right)\rho_i E[U_{j-1}^{(p)}] = \cdots = \left\{\sum_{i=1}^{p-1}\left(1-\frac{\alpha_p}{\alpha_i}\right)\rho_i\right\}^j E[U_0^{(p)}] \tag{10}$$

を得る．(9) の両辺の平均をとり (10) を代入し，$\sum_{i=1}^{p-1}(1-\alpha_p/\alpha_i)\rho_i < 1$ が成り立つことを考慮すれば

$$E[W_p] = \sum_{j=0}^{\infty} E[U_j^{(p)}] = \sum_{j=0}^{\infty}\left\{\sum_{i=1}^{p-1}\left(1-\frac{\alpha_p}{\alpha_i}\right)\rho_i\right\}^j E[U_0^{(p)}] = \frac{E[U_0^{(p)}]}{\sum_{i=1}^{p-1} 1-\left(1-\frac{\alpha_p}{\alpha_i}\right)\rho_i} \tag{11}$$

となる．

$U_0^{(p)}$ は，C_p の客が割り込まれない場合の待ち時間であるから，システムの残余仕事量から自分より優先度の低い客に対して割り込む仕事量を差し引いた量に一致する．特に，C_J の客は割り込むことがないので $U_0^{(J)}$ はシステムの残余仕事量と等しく

$$E[U_0^{(J)}] = \frac{1}{2(1-\rho)}\sum_{i=1}^{J}\lambda_i E[S_i^2]$$

で与えられる [8]．図 2 は C_i の客の残余仕事量のサンプルパスで，水平部分は客が列にならんでサービスを待っている状態，斜めになっている部分はサービスを受けている状態に対応している．サービスが開始／再開してから割り込まれるまで時間，および割り込まれてから次にサービスを再開するまでの時間は互いに独立で同一の分布に従っているから，残余仕事量は平均的にはある程度一様に減少していくと考えてよい．そこで，残余仕事量が図 2 の太線で示したように直線で減少すると考えて近似を行うと，C_i の客がシステムに持ち込む仕事量は

$$S_i \times (W_i + \frac{1}{2}R_i)$$

となる．W_p と S_p が独立であることと (7)，(11) を用いると，この仕事量の平均は

$$E[S_i]E[W_i] + \frac{1}{2}E[S_iR_i] = E[S_i]E[W_i] + \frac{E[S_i^2]}{2-2\sum_{j=1}^{i-1}\left(1-\frac{\alpha_i}{\alpha_j}\right)\rho_j}$$

となる．自分がシステムに到着したとき，待っている C_i の客が自分より先にシステムから退去する確率を f_{ip} とすると

$$f_{ip} = \begin{cases} 1 & (i \geq p) \\ \dfrac{\alpha_i}{\alpha_p} & (i < p) \end{cases}$$

となる [7]．到着率を考慮に入れると，以上の結果より

$$E[U_0^{(p)}] \simeq \frac{1}{2(1-\rho)} \sum_{i=1}^{J} \lambda_i E[S_i^2] - \sum_{i=p+1}^{J} \left(1 - \frac{\alpha_i}{\alpha_p}\right) \left\{ \rho_i E[W_i] + \frac{\lambda_i E[S_i^2]}{2 - 2\sum_{j=1}^{i-1} \left(1 - \frac{\alpha_i}{\alpha_j}\right) \rho_j} \right\} \tag{12}$$

が得られる．よって (11) に (12) を代入することにより

$$E[W_p] \simeq \frac{\frac{1}{2(1-\rho)} \sum_{i=1}^{J} \lambda_i E[S_i^2] - \sum_{i=p+1}^{J} \left(1 - \frac{\alpha_i}{\alpha_p}\right) \left\{ \rho_i E[W_i] + \frac{\lambda_i E[S_i^2]}{2 - 2\sum_{j=1}^{i-1} \left(1 - \frac{\alpha_i}{\alpha_j}\right) \rho_j} \right\}}{1 - \sum_{i=1}^{p-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i}$$

となる．この $E[W_p]$ に関する再帰式は

$$E[W_J] = \frac{\frac{1}{2(1-\rho)} \sum_{i=1}^{J} \lambda_i E[S_i^2]}{\sum_{i=1}^{J-1} \left(1 - \frac{\alpha_p}{\alpha_i}\right) \rho_i}$$

から出発して順に計算するすることができる．

3–3. リーグ分け

これまで考えてきた優先方式では 3 クラス以上の場合に通常の割り込み優先方式を実現することができない．そこで，[2] で提案されているリーグ分けの方法を利用することで，割り込み優先方式も含むように拡張し，その場合の各クラスの客の待ち時間や総サービス時間に対する近似式を導出する．

リーグ分けとは，J 個のクラス $C_1, \cdots, C_p$ を L 個のリーグ $E_1, \cdots, E_L$, $E_k = \{C_{i_k}, \cdots, C_{i_{k+1}-1}\}$ に分割し，

- 同じリーグに属するクラスの客に対しては遅延に依存する割り込み優先方式を適用
- 異なるリーグ E_k, E_ℓ $(k < \ell)$ に属するクラスの客に対しては，E_k に属するクラスの客が E_ℓ に属するクラスの客に割り込み優先権をもつ

という方式である．

以下では，1 リーグの場合の解析を利用して各クラスの客の平均待ち時間と平均総サービス時間の近似式を求める．E_ℓ に属する C_p の客 C に対して，$I_0^{(p)}$ を C のサービス時間，$U_0^{(p)}$ を C 以降到着がなかった場合の待ち時間とし

$I_j^{(p)}$ ： $I_{j-1}^{(p)}$ の間に C に割り込む客のサービス時間の合計

$U_j^{(p)}$ ： $U_{j-1}^{(p)}$ の間に C に割り込む客のサービス時間の合計

とする．異なるリーグ間では割り込み優先であることに注意すると，(4) と同様の考え方で

$$\mathrm{E}[I_j^{(p)}] \simeq \left\{ \sum_{i<p, i\in E_\ell} \left(1-\frac{\alpha_p}{\alpha_i}\right)\rho_i + \sum_{i<p, i\notin E_\ell} \rho_i \right\}^j \mathrm{E}[S_p]$$

$$\mathrm{E}[U_j^{(p)}] \simeq \left\{ \sum_{i<p, i\in E_\ell} \left(1-\frac{\alpha_p}{\alpha_i}\right)\rho_i + \sum_{i<p, i\notin E_\ell} \rho_i \right\}^j \mathrm{E}[U_0^{(p)}]$$

という近似式が導出できる．これらの式を (7)，(11) に代入すれば，リーグ分けがある場合について

$$\mathrm{E}[W_p] \simeq A_p \left[\frac{\sum_{i\in B_\ell}\lambda_i \mathrm{E}[S_i^2]}{2\left(1-\sum_{i\in B_\ell}\rho_i\right)} - \sum_{i>p, i\in E_\ell} \left(1-\frac{\alpha_i}{\alpha_p}\right) \left\{ \rho_i \mathrm{E}[W_i] + \frac{1}{2}\lambda_i \mathrm{E}[S_i^2] A_i \right\} \right]$$

$$\mathrm{E}[R_p] \simeq A_p \mathrm{E}[S_p]$$

が得られる．ここで

$$A_p = \left[1 - \left\{ \sum_{i<p, i\in E_\ell} \left(1-\frac{\alpha_p}{\alpha_i}\right)\rho_i + \sum_{i<p, i\notin E_\ell} \rho_i \right\} \right]^{-1}$$

$$B_\ell = \bigcup_{k=1}^{\ell} E_k$$

である．

4. 近似精度の検証

この節では，3 節で導出した近似式が，(i) 先着順，(ii) 2 クラス割り込み優先に対しては厳密解と一致することを示し，またシミュレーションとの比較によって精度を検証する．

4–1. 先着順

各クラスの優先度係数がすべて同じであるとすると，本論文で扱う優先方式は先着順になる．近似式 (7)，(11) において $\alpha_1 = \cdots = \alpha_J$ とすると

$$\mathrm{E}[W_p] + \mathrm{E}[R_p] \simeq E[S_p] + \frac{1}{2(1-\rho)} \sum_{i=1}^{J} \lambda_i E[S_i^2]$$

となる．一方，先着順の場合ここでのモデルは客が 1 クラスの M/G/1 と見なせるからポラチェック–ヒンチンの公式より

$$\mathrm{E}[W_p] + \mathrm{E}[R_p] = E[S_p] + \frac{1}{2(1-\rho)} \sum_{i=1}^{J} \lambda_i E[S_i^2]$$

を得る．したがって，求めた近似式が厳密解と一致することが確かめられた．

4–2. 2 クラス割り込み優先方式

2 クラスの場合に $\alpha_2 = 0$ とすれば通常の M/G/1 の割り込み優先方式となる．近似式に $\alpha_2 = 0$ を代入して計算すると

$$\begin{aligned} \mathrm{E}[W_1] + \mathrm{E}[R_1] &\simeq E[S_1] + \frac{\lambda_1 E[S_1^2]}{2(1-\rho_1)} \\ \mathrm{E}[W_2] + \mathrm{E}[R_2] &\simeq \frac{1}{1-\rho_1}\left\{E[S_2] + \frac{\lambda_1 E[S_1^2] + \lambda_2 E[S_2^2]}{2(1-\rho)}\right\} \end{aligned}$$

が得られる．一方，2 クラスで割り込み優先方式の M/G/1 の平均系内滞在時間は

$$\begin{aligned} \mathrm{E}[W_1] + \mathrm{E}[R_1] &= E[S_1] + \frac{\lambda_1 E[S_1^2]}{2(1-\rho_1)} \\ \mathrm{E}[W_2] + \mathrm{E}[R_2] &= \frac{1}{1-\rho_1}\left\{E[S_2] + \frac{\lambda_1 E[S_1^2] + \lambda_2 E[S_2^2]}{2(1-\rho)}\right\} \end{aligned}$$

で与えられる [4]．よって，この場合も近似式は厳密解と一致することが確認できた．

4–3. シミュレーションとの比較

クラス数，サービス時間分布，到着率，優先度係数などを変化させてシミュレーションを行い，近似式の精度を調べた結果の一例を示す．表 1 は 4 クラスの M/M,E_2,H_2,D/1 において

- $\lambda_1 : \lambda_2 : \lambda_3 : \lambda_4 = 3 : 2 : 3 : 4$
- $\mu_1 = 1.0,\ \mu_2 = 2.0,\ \mu_3 = 1.0,\ \mu_4 = 2.0$
- $\rho = 0.5, 0.7, 0.9$

と設定した場合の平均系内滞在時間を示している．シミュレーション結果の欄は，中央の数字がシミュレーションの結果，左右が 95% 信頼区間をそれぞれ表している．

ρ	$\mathrm{E}[W_1]+\mathrm{E}[R_1]$		$\mathrm{E}[W_2]+\mathrm{E}[R_2]$	
	近似値	シミュレーション結果	近似値	シミュレーション結果
0.5	1.82	[1.82 – 1.82 – 1.83]	1.50	[1.49 – 1.49 – 1.49]
0.7	2.77	[2.76 – 2.77 – 2.78]	2.67	[2.66 – 2.67 – 2.68]
0.9	6.86	[6.73 – 6.82 – 6.92]	7.78	[7.62 – 7.73 – 7.84]
ρ	$\mathrm{E}[W_3]+\mathrm{E}[R_3]$		$\mathrm{E}[W_4]+\mathrm{E}[R_4]$	
	近似値	シミュレーション結果	近似値	シミュレーション結果
0.5	3.41	[3.38 – 3.39 – 3.40]	3.84	[3.80 – 3.82 – 3.84]
0.7	6.42	[6.34 – 6.37 – 6.41]	9.75	[9.59 – 9.65 – 9.71]
0.9	19.88	[19.38 – 19.68 – 19.99]	44.92	[43.48 – 44.32 – 45.16]

表 1: M/M,E_2,H_2,D/1

他のシミュレーション結果も含めて検討すると，比較を行ったかなりのケースで近似値は信頼区間内に入っており，またシミュレーション結果との相対誤差は最大で 4 % 程度と概ね良好な精度であった．一般に，サービス分布の変動係数の小さい場合は近似値が信頼区間に含まれるが，変動係数の大きい場合は近似値の方がやや大きい値をとるようである．またトラヒック密度が高いと，値そのものが大きくなるため精度が上がる反面，トラヒック密度がさほど高くないときは近似値がやや大きい値をとる傾向が観察された．

5. 最適化問題への応用

本節では，これまでに得られた結果の，遅延に依存した割り込み優先方式をもつ M/G/1 待ち行列システムにおける最適化問題への応用について述べる．

J クラスの客を単一サーバで処理する M/G/1 待ち行列システムにおいて

$\xi_p(R)$: 優先方式 R のもとでの C_p の客の評価尺度

$C(\xi_1(R),\cdots,\xi_J(R))$: システム全体のコスト関数

とする．実現可能な優先方式の集合を $\mathcal{R}$ とし，最適化問題

$$\left|\begin{array}{ll}\text{minimize} & C(\xi_1(R),\cdots,\xi_J(R)) \\ \text{subject to} & R\in\mathcal{R}\end{array}\right.$$

を考える．

例　平均系内滞在時間を評価尺度として，線形コスト関数

$$C(\xi_1(R),\cdots,\xi_J(R))=\sum_{p=1}^{J} c_p\,\xi_p$$

を最小化する問題を考える．実現可能な優先方式 $\mathcal{R}=$（仕事量を保存する優先方式）の場合は，c_p/λ_p が大きな順に各クラスに対して割り込み優先権を与える方式が最適となる [2].

この結果は，(i) $R\in\mathcal{R}$ ならば $\sum_p \lambda_p\,\xi_p(R)=$（一定），(ii) 線形関数は凸包の端点で最小値をとる，という事実に基づいているため，例えばコスト関数が凸関数で内点で最適解をとるような場合には，単純な割り込み優先方式だけでは最適な評価尺度を実現できない．

[2] では，遅延に依存した非割り込み優先方式をもつ M/G/1 において

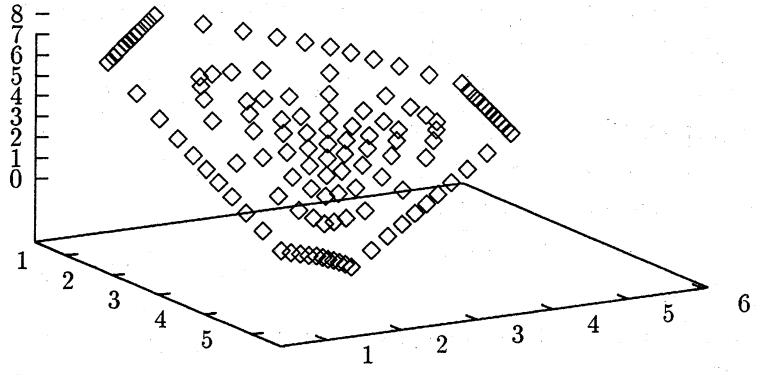
1. リーグ分け／優先度係数を変化させて達成できる各クラスの客の平均待ち時間ベクトルは，J 次元実数空間における $J-1$ 次元超平面上の凸包をなす
2. この凸包の頂点は通常の非割り込み優先方式によって達成される
3. 達成可能な平均待ち時間ベクトルとリーグ分け／優先度係数は 1 対 1 に対応する

ことを示し，さらに達成可能な平均待ち時間ベクトルが与えられたとき，それを実現するリーグ分け／優先度係数を求めるアルゴリズムを提案している．

同様な結果が，遅延に依存した割り込み優先方式をもつ M/G/1 でも成り立つことが期待されるが，[2] の証明は平均待ち時間に対する厳密解を用いているためここでは適用できない．そこで，得られた近似式をもとに優先度係数 α_p を変化させたときの平均系内滞在時間の範囲を数値的に調べてみた．仕事量を保存する優先方式では，$\xi_p=C_p$ の客の平均系内滞在時間 に対して

$$\sum_{p=1}^{J}\lambda_p\,\xi_p=(\text{一定}) \tag{13}$$

という関係が成り立つ [7]. したがって，平均系内滞在時間ベクトルは (13) で表される $J-1$ 次元超平面上に存在する.



グラフ 1：優先度係数を変化させたときの平均系内滞在時間ベクトル（3 クラス）

実際，3 クラスの結果をプロットした 図 3 を見ると，平均系内滞在時間ベクトルは各クラスに割り込み優先権を与えたときのベクトルを端点とする凸包上を埋めるように分布していることが確かめられる．直観的には，優先度係数を連続的に変化させると評価尺度も連続的に変化すると考えられるから，凸包上の任意のベクトルはリーグ分け／優先度係数をうまく選ぶことで実現できると予想されるが，厳密な証明はまだ得られていない．待ち行列の連続性の議論等が必要になるであろう [13].

最後に，平均総処理時間 $\mathrm{E}[R_p]$ のベクトルが与えられたとき，それが（近似式に関して）達成可能かどうかを判定し，可能ならば対応する優先度関数を求める簡単なアルゴリズムについて説明する．準備として，

$$f_p(x) = \mathrm{E}[R_p]\left\{\sum_{i=1}^{p-2}\left(\prod_{k=i+1}^{p-1} r_k\right)\rho_i\right\}x + \left(1-\sum_{i=1}^{p-1}\rho_i\right)\mathrm{E}[R_p] - \mathrm{E}[S_p]$$

とする.

Input : λ_p, $\mathrm{E}[S_p]$, $\mathrm{E}[R_p]$, $(p=1,\cdots,J)$

Output : α_p $(p=1,\cdots,J)$

Algorithm :

1. $f_p(x)=0$ を満たす r_p $(p=2,\cdots,J)$ を計算
2. $r_p \notin [0,1]$ なる p が存在 $\Longrightarrow$ $\mathrm{E}[R_p]$ は実現不可能
   $r_p \in [0,1]$ $(p=2,\cdots,J) \Longrightarrow$ $\mathrm{E}[R_p]$ は実現可能
   (a) $r_{i_\ell}=0$ なる i_ℓ で $E_\ell=\{C_{i_\ell},\cdots,C_{i_{\ell+1}-1}\}$ にリーグ分け
   (b) $\alpha_p \in E_\ell$ を
   $\alpha_{i_\ell}=1$
   $\alpha_p=\prod_{j=i_\ell}^{p} r_j, \quad p=i_\ell+1,\cdots,i_{\ell+1}-1$
   で求める

それ以外の評価尺度について，達成可能な評価尺度ベクトルが与えられたときにそれを達成する優先度係数を求める方法は見つかっていない．評価尺度は一般に優先度係数に関して単調であるから，その性質を用いて優先度係数の空間を分割するなどの方法が考えられるであろう．

参考文献

[1] Bagchi, U. and Sullivan, R., "Dynamic nonpreemptive priority queues with general linearly increasing priority function," *Oper. Res.*, **33**, 1278-1299, 1985.

[2] Federgruen, A. and Groenevelt, H., "M/G/c queueing systems with multiple customer classes," *Management Science*, **28**, 1121-1138, 1988.

[3] Federgruen, A. and Groenevelt, H., "Characterization and control of achievable performance in general queueing system," *Oper. Res.*, **36**, 1988.

[4] Heyman, D.P. and Sobel, M.J., *Stochastic Models in Operations Research, Vol.1*, McGraw-Hill, New York, 1982.

[5] Jackson, J., "Some problems in queueing with dynamic priorities," *Nav. Res. Log. Quart.*, **7**, 235–249, 1960.

[6] Jaiswal, N.K., *Priority Queues: Mathematics in Science and Engineering*, **50**, Academic Press, New York, 1968.

[7] Kleinrock, L., "A delay dependent queue discipline," *Nav. Res. Log. Quart.*, **11**, 329–341, 1964.

[8] Kleinrock, L., "A conservation law for a wide class of queueing disciplines," *Nav. Res. Log. Quart.*, **12**, 181–192, 1965.

[9] Kleinrock, L. and Finkelstein, R., "Time dependent priority queues," *Oper. Res.*, **15**, 104–116, 1967.

[10] Kleinrock, L., *Queueing Systems, Vol.2,* John Wiley, New York, 1976.

[11] Netterman, A. and Adiri, A., "A dynamic priority queue with general concave priority functions," *Oper. Res.*, **27**, 1088–1100, 1979. New York, 1976.

[12] Takagi, H., *Queueing Analysis: A Foundation of Performance Evaluation, Vol.1, Vacations and Priority Systems*, Elsevier Science Publisher, North-Holland, 1991.

[13] Whitt, W., "Continuity of generalized semi-Markov Processes," *Math. Oper. Res.*, **5**, 494–501, 1980.