

非割込み優先処理のある網型待ち行列の近似解析

池原 悟 ・ 宮崎 正弘

(日本電信電話公社 横須賀電気通信研究所)

1. まえがき

網型待ち行列理論の発展によって、BCMP型待ち行列網の解法^{1,2)}が見い出され、以来、計算機システムの性能評価にこの型の待ち行列が広く使われるようになってきている。しかし、実際の計算機システムの動作は複雑でありBCMP型待ち行列網ではモデル化できない場合も多い。そのようなものに優先処理がある。

優先処理方式には割込み型と非割込み型があるが割込み優先方式に対しては、Reiser³⁾のセントラルサーバモデルでの近似がある。この近似では、優先呼の特性は非優先呼を無視することによって求め、非優先呼の特性は、非優先呼に対するサーバのサービス率 μ_2 と $\mu_2(1 - \rho_1)$ （但し、 ρ_1 は優先呼のサーバ使用率）に補正することによって近似的に求めている。Reiserのモデルでは異呼種間の競合は単一のサービスステーション（CPU）に限られ、他のサービスステーション（端末）での異呼種間の競合はないことが仮定されているが、Sevcik⁴⁾はこの方法を拡張し、単一の割込み優先サービスステーションをもつ待ち行列網で、他のサービスステーションでも異呼種の相互作用のある場合について解析している。

本稿では、このモデルをさらに拡張し、非割込み優先処理のある場合に対する近似解法を提案する。非割込み優先処理の行なわれるサービスステーションでは優先呼と非優先呼が互に影響し合っており、互いの影響を考慮に入れなければならない。本近似解法ではSevcikの考えを用い、待ち行列網内の非割込み優先処理サービスステーションを、優先呼のみを処理する実サーバと非優先呼のみを処理する写像サーバのサービスステーションで置換して、待ち行列網全体をBCMP型待ち行列網に帰着させている。実サーバと写像サーバのサービス率を決めるには新たに仮想保留とスーパーコールによる近似を用いている。すなわち、着目するサービスステーションのサーバ使用率と実サーバ、写像サーバのサービス率の関係を仮想保留とスーパーコールの考えを用いて求め、それらの関係とBCMP型待ち行列理論を組み合わせることで実サーバと写像サーバのサービス率を定めている。

本解法の事例としては簡単な架空のモデルによる厳密解との比較、現用システムのシステムモデルによるシミュレーションとの比較をとり上げた。それらの結果によれば本近似解法はジョブスケジューリング方式とシステム性能の関係を評価する上で、シミュレーションに代わる十分有効な方法であると期待される。

2. 解析モデル

BCMP型の待ち行列網に対して以下の制限と拡張を行なう。

- ① 呼種は2種（優先呼、非優先呼）とする。
- ② 優先呼、非優先呼の処理経路はいずれもクロストチェインで表わされ、チェーン上の呼の多重度はそれぞれ、 N_2 とする。
- ③ サービスステーションでは非割込み優先処理が行なわれるとする。即ち、
 - 1) 優先呼は非優先呼に優先してサーバを保留するが、非優先呼がサーバ保留中に到着した優先呼は処理中の非優先呼に割込まず、待ち行列に入る。
 - 2) 優先呼、非優先呼両方は共に先着順（FCFS）で処理される。

④ 非割込み優先処理の行なわれるサービスステーションこのサーバ数は1とし、サーバ保留時間は優先呼、非優先呼共に指数分布をとるものとする。

以上の解析モデルは、サービスステーションと短絡させればBCMP型待ち行列となり、従来のプログラムパッケージを直接解析できる。以下では、非割込み優先処理の行なわれるサービスステーションとまず1とし、後に複数の場合に拡張する。

3. 解析

3.1 仮想保留とサーバの分離

非割込み優先処理の行なわれるサービスステーション(以下、ステーションと呼ぶ)に着目する。このステーションの優先呼および非優先呼に対するサービス率を μ_{i1}, μ_{i2} とし、サーバ使用率を ρ_{i1}, ρ_{i2} とすると、このステーションの各呼のスループットは $\lambda_{i1} = \mu_{i1} \rho_{i1}$, $\lambda_{i2} = \mu_{i2} \rho_{i2}$ と表わされる。

このステーションでは前章の条件から、優先呼、非優先呼は互に影響し合っており、ステーションの通過時間に遅れが生じる。このような優先呼、非優先呼の処理の遅れを仮想的なサーバ保留時間 $\Delta h_{i1}, \Delta h_{i2}$ として本来の平均保留時間 $h_{i1} (= 1/\mu_{i1})$, $h_{i2} (= 1/\mu_{i2})$ に加え、平均保留時間が $h'_{i1} (= h_{i1} + \Delta h_{i1})$, $h'_{i2} (= h_{i2} + \Delta h_{i2})$ に増大したものと考えると、ステーションiは図1に示されるように2つのステーション i_1, i_2 でおきかえられる。この図で、ステーションiは優先呼のみを実行し、 i_2 は非優先呼のみを実行する。前者を実サーバ、後者を想像サーバと呼ぶ。

ステーションiと i_2 のサーバ保留時間は一般に指数分布とはならないが、これらを指数分布で近似すると、ステーションiを i_1, i_2 でおきかえた解析モデルBCMP型待ち行列網となり、従来のプログラムパッケージANET³⁾, GSC⁵⁾などで解析できる。この近似は $\Delta h_{i1}, \Delta h_{i2}$ が小さい程よく成り立つ。

3.2 実サーバと想像サーバの平均保留時間

(1) 実サーバの平均保留時間

ステーションiで優先呼が非優先呼処理を待つ時間はステーションiでは優先呼のサーバ保留時間に含まれているから、ステーションiの使用率 ρ'_{i1} はステーションiの優先呼の使用率 ρ_{i1} よりも非優先呼による仮想保留分だけ大きい。
そこでこの差を ρ_{ix} とおくと、以下の式が成り立つ。

$$\rho'_{i1} = \rho_{i1} + \rho_{ix} \quad \text{--- (1)}$$

次に、実サーバのスループットとモデルのステーションiの優先呼のスループットは等しいから、以下の式が成り立つ。

$$\rho'_{i1} / h'_{i1} = \rho_{i1} / h_{i1} = \lambda_{i1} \quad \text{--- (2)}$$

式(1), (2)から ρ'_{i1} を消去し、 h'_{i1} を求めると、以下のようになる。

$$h'_{i1} = h_{i1} (1 + \rho_{ix} / \rho_{i1}) \quad \text{--- (3)}$$

(2) 想像サーバの平均保留時間

ステーションiで非優先呼が優先呼処理を待つ時間はステーションiでは非優先呼のサーバ保留時間に含まれているから、ステーションiの使用率 ρ'_{i2} はステーションiの非優先呼の使用率 ρ_{i2} よりも優先呼による仮想保留分だけ大きい。
そこでこの差を ρ_{iy} とおくと、以下の式が成り立つ。

$$\rho'_{i2} = \rho_{i2} + \rho_{iy} \quad \text{--- (4)}$$

(1)とまったく同様の議論より、 h'_{i2} を求めると、以下のようになる。

$$h'_{i2} = h_{i2} (1 + \rho_{iy} / \rho_{i2}) \quad \text{--- (5)}$$

3.3 仮想保留状態の確率

(1) 優先呼のスーパコール近似

非割込み優先処理の行なわれるステーション i において、継続してサービスを保留する優先呼をまとめてスーパコールとする。ここで、優先呼の多重度 N_1 は 1 より十分大きいと仮定すると、スーパコールの平均呼空閑率は $1/\lambda_{i1}$ で近似できる。

次に、スーパコールの平均保留時間 W_i は以下のように近似される。 $N_1 \gg 1$ の条件のもとでは、優先呼がステーション i に到着した時他の優先呼が N_1 のステーションに無い確率は平衡状態においてステーション i に優先呼の無い確率と考えてよく、 $1 - \rho_{i1} - \rho_{ix}$ である。従って、1つの優先呼から見ると、 N_1 の優先呼がスーパコールの先頭になる確率は $1 - \rho_{i1} - \rho_{ix}$ であるから、1つのスーパコールは平均してみると $1/(1 - \rho_{i1} - \rho_{ix})$ 件の優先呼のまとまりであるとみなせる。よって、スーパコールの平均保留時間 W_i 、およびサービス率 μ_{i1}^* は以下の式で表わされる。

$$\left. \begin{aligned} W_i &= \lambda_{i1} / (1 - \rho_{i1} - \rho_{ix}) \\ \mu_{i1}^* &= 1 / W_i \end{aligned} \right\} \text{----- (6)}$$

(2) 状態方程式と N_2 の解

ステーション i のみに着目した状態方程式を得く。優先呼と同様、非優先呼の多重度 N_2 も 1 より十分大きいとする。また、優先呼はスーパコールとして扱われ、スーパコールの分布は指数分布とある。

さて、ステーション i の状態を表わす状態変数 g, n を以下の通り定める。

$$g = \begin{cases} 0: \text{ステーション } i \text{ にスーパコールなし} \\ 1: \text{スーパコールが仮想保留状態 (待ち状態)} \\ 2: \text{スーパコール実行中} \end{cases}$$

n : ステーション i 内の非優先呼件数

ここで、 $g = 0, 2$ の時 $0 \leq n \leq N_2$ 、 $g = 1$ の時 $1 \leq n \leq N_2$ である。平衡状態において、ステーション i が状態 (g, n) となる確率を $P_i(g, n)$ で表わすと、規格化条件として以下の式が成り立つ。

$$P_i(0, 0) + P_i(2, 0) + \sum_{g=0}^2 \sum_{n=1}^{N_2} P_i(g, n) = 1 \text{----- (7)}$$

図2にステーション i の状態遷移図を示す。ここで $\rho_{i1}, \rho_{i2}, \rho_{ix}, \rho_{ig}$ は以下の式で表わされる。

$$\rho_{i1} = \sum_{n=0}^{N_2} P_i(2, n) \text{----- (8)}$$

$$\rho_{i2} = \sum_{g=0}^1 \sum_{n=1}^{N_2} P_i(g, n) \text{----- (9)}$$

$$\rho_{ix} = \rho_{i2} - \sum_{n=1}^{N_2} P_i(0, n) = \sum_{n=1}^{N_2} P_i(1, n) \text{----- (10)}$$

$$\rho_{ig} = \rho_{i1} - P_i(2, 0) = \sum_{n=1}^{N_2} P_i(2, n) \text{----- (11)}$$

図2から、以下の状態方程式が成り立つ。

$$(\lambda_{i1} + \lambda_{i2}) P_i(0, 0) = \mu_{i1} P_i(0, 1) + \mu_{i1}^* P_i(2, 0) \text{----- (12)}$$

$$(\lambda_{i2} + \mu_{i2}) P_i(1, 1) = \lambda_{i1} P_i(0, 1)$$

$$(\lambda_{i2} + \mu_{i2}) P_i(1, 2) = \lambda_{i1} P_i(0, 2) + \lambda_{i2} P_i(1, 1)$$

$$\mu_{i2} P_i(1, N_2) = \lambda_{i1} P_i(0, N_2) + \lambda_{i2} P_i(1, N_2 - 1)$$

$$\left. \begin{aligned} & \text{----- (13)} \end{aligned} \right\}$$

$$(\mu_{i1}^* + \lambda_{i2}) P_i(2,0) = \lambda_{i1} P_i(1,0) + \mu_{i2} P_i(1,1) \text{ --- (14)}$$

式(13)の右辺を加え合せ、式(10)を用いると β_{i2} は以下の式で表わされる。

$$\beta_{i2} = \lambda_{i1} h_{i2} \beta_{i2} / (1 + \lambda_{i1} h_{i2}) \text{ --- (15)}$$

次に、式(12)、式(13)の第1式、式(14)の3式と $P_i(1,0) = 1 - \beta_{i1} - \beta_{i2}$ を用いて、 $P_i(2,0)$ を求めると以下のようになる。

$$P_i(2,0) = \frac{\lambda_{i1} (\lambda_{i1} + 2\lambda_{i2} + \mu_{i2}) (1 - \beta_{i1} - \beta_{i2})}{\lambda_{i1} \mu_{i1}^* + (\lambda_{i2} + \mu_{i1}^*) (\lambda_{i2} + \mu_{i2})} \text{ --- (16)}$$

よって式(11)から β_{i2} は以下の式で表わされる。

$$\beta_{i2} = \lambda_{i1} \left\{ \frac{1}{\mu_{i1}} - \frac{(\lambda_{i1} + 2\lambda_{i2} + \mu_{i2}) (1 - \beta_{i1} - \beta_{i2})}{\lambda_{i1} \mu_{i1}^* + (\lambda_{i2} + \mu_{i1}^*) (\lambda_{i2} + \mu_{i2})} \right\} \text{ --- (17)}$$

従って、 β_{i1} , β_{i2} が求まれば式(3), (5), (6), (15), (17)より h_{i1}' , h_{i2}' が求まる。

3.4 漸近近似解析手順

クロースド系では β_{i1} , β_{i2} はあらかじめ一意に定めることができない。そこで、以下のような漸近近似解析を行なう。

Step1: <リ返し表示> $\rho = 0$ とする。ステーション i のサービス規律 PS (Processor Sharing) とした時の解析モデルを BCMP 型待ち行列理論で解き、得られたステーション i の使用率を $\beta_{i1}^{(0)}$, $\beta_{i2}^{(0)}$ とする。

Step2: $\beta_{i1} = \beta_{i1}^{(0)}$, $\beta_{i2} = \beta_{i2}^{(0)}$ より h_{i1}' , h_{i2}' を求める。

Step3: ステーション i の平均保留時間が h_{i1}' , h_{i2}' の実サーバステーション i, 写像サーバステーション j でおきかえた解析モデルを BCMP 型待ち行列理論で解き、実サーバ, 写像サーバの使用率 $\beta_{i1}^{(1)}$, $\beta_{i2}^{(1)}$ を求める。

Step4: β_{i1}' , β_{i2}' の値を $\beta_{i1}^{(1)}$, $\beta_{i2}^{(1)}$ とした時の β_{i1} , β_{i2} の値 $\beta_{i1}^{(2)}$, $\beta_{i2}^{(2)}$ を式(1), (4), (15), (17)より求める。なお、解は陽には得られないので、二分探索によって求める*。

Step5: $\beta_{i1}^{(2)}$, $\beta_{i2}^{(2)}$ が $\beta_{i1}^{(0)}$, $\beta_{i2}^{(0)}$ に必要とする精度で一致すれば解析を終了し、そうでない時は $\rho = \rho + 1$ として Step2 に戻る。

3.5 本解法上の近似と精度

本近似解法では2章で規定された待ち行列網を解く過程で以下の近似を行なっている。

- ① 実サーバ, 写像サーバの保留時間を指数分布で近似した。
- ② 優先呼の多重度 N_1 が1より十分大きいことを仮定し、優先呼に対してスーパコール近似を行なった。
- ③ 非優先呼の多重度 N_2 も1より十分大きいとして、ステーション i への非優先呼の到着率が網の状態に独立であると仮定した。

これらのうち①の影響を考える。実サーバ, 写像サーバの保留時間はそれぞれ優先呼, 非優先呼の本来のサーバ保留時間とお互いが待ち合せする時間(仮想保留時間)を加え合せたものであるから、実サーバ, 写像サーバの保留時間は指数分布に比べると変動係数が大きい。これらと①の通り近似しているため、近似解法で得られた応答時間は正しい値に比べて小さくなると思われる。この傾向は仮想保留が増えるにつれて大きくなるから優先呼の保留時間 h_{i1} , 又は多重度 N_1 が増えるにつれて非優先呼の応答時間 T_2 の近似誤差が増大し、逆に非優先呼の保留時間

* β_{i2} は式(4), (15)より β_{i1}' , β_{i2}' , β_{i1} より一意に定まるので、一次元の二分探索を行なえばよい。

間 h_{i2} 、又は多重度 N_2 が増すにつれて、優先呼の応答時間 T_1 の近似誤差が増大することが予想される。

次に、②③の近似は優先呼、非優先呼の多重度 N_1, N_2 が大きい程よく成り立つから、これらの近似誤差は N_1, N_2 が大きくなるにつれて小さくなると考えられる。

以上まとめると、近似解析では一般に、以下のことがわかる。

- ④ 優先呼の多重度 N_1 が大きい程、優先呼の応答時間 T_1 の近似誤差は小さく、非優先呼の多重度 N_2 が大きい程、非優先呼の応答時間 T_2 の近似誤差は小さい。
- ⑤ 優先呼、非優先呼の保留時間 h_{i1}, h_{i2} が小さい程、それぞれ非優先呼、優先呼の応答時間 T_2, T_1 の近似誤差は小さい。

3. 6 本近似解法の拡張

(1) 複数の非割込み優先処理ステーションへの拡張

この場合は、非割込み優先処理の行なわれるステーション $i, j, \dots (i+j+\dots)$ に対して同時に3.4の解法を適用すればよい。ただし、本モデルでは系内呼損はなく、各ステーションの優先呼および非優先呼の使用率はいずれも比例関係にあるので、Step 4で $g_{i1}^{(k+1)}, g_{i2}^{(k+1)}, g_{j1}^{(k+1)}, g_{j2}^{(k+1)}, \dots$ を求める時、ある特定のステーション i に対して得られた $g_{i1}^{(k+1)}, g_{i2}^{(k+1)}$ から他のステーション j の使用率 $g_{j1}^{(k+1)}, g_{j2}^{(k+1)}, \dots$ を求めればよい。従って、複数の非割込み優先処理ステーションのある場合も、Step 4で必要な二分探索は高々1回である。

(2) 割込み優先処理の混在する場合への拡張

前項(1)の解析モデルをさらに拡張し、ネットワーク内に非割込み型と割込み型の優先処理が混在する場合を考える。

割込み型優先処理の行なわれるステーション i と前と同じ、実サーバと写像サーバのステーションに分離すると、これらの平均保留時間 h_{i1}', h_{i2}' とサーバ使用率 g_{i1}', g_{i2}' はそれぞれ以下の式で表わされる。

$$h_{i1}' = h_{i1} \quad \text{--- (18)}$$

$$h_{i2}' = h_{i2} / (1 - g_{i1}) \quad \text{--- (19)}$$

$$g_{i1}' = g_{i1} \quad \text{--- (20)}$$

$$g_{i2}' = g_{i2} / (1 - g_{i1}) \quad \text{--- (21)}$$

式(20)、(21)と g_{i1}, g_{i2} について解けば以下の式を得る。

$$\left. \begin{aligned} g_{i1} &= g_{i1}' \\ g_{i2} &= g_{i2}' (1 - g_{i1}') \end{aligned} \right\} \quad \text{--- (22)}$$

従って、この場合は割込み型の優先処理の行なわれるステーションについても3.4のStep 1 ~ Step 5を適用すればよい。ただし、Step 2で式(18)(19)より h_{i1}', h_{i2}' を求め、Step 4では二分探索の代りに式(22)を用いられればよい。

4. 事例検討

前章で提案した漸近近似解法の精度および実用性を確かめるため2つの事例に適用する。第1の例は2ステーションからなるモデルであり、本近似解法の結果と数値的に状態方程式を解いて得た厳密解との比較を行なう。

第2の例は5ステーションからなるモデルで、現用のシステムをモデル化したものである。このモデルでは近似解法とシミュレーション結果とを比較し、非割込み優先方式と割込み優先方式と先着順サービス、PSサービス方式と比べる時の評価手段として本近似解法の有効性を確かめる。

4. 1 厳密解との比較(2ステーションの場合)

4.1.1 解析モデル

図3に示すような2つのステーションからなるモデルを解析する。本モデルはステーション2のサーバ数が無限大であり、このステーションで待ち合せは生じないため、出線1, 入線は優先呼側 N_1 , 非優先呼側 N_2 の有限線群のモデルに等価である。

4.1.2 数値解析の方法

ステーション1の状態に着目して状態変数 q, n_1, n_2 と以下のように定める。

$$q = \begin{cases} 0 & \text{サービス中の呼なし} \\ 1 & \text{優先呼処理中} \\ 2 & \text{非優先呼処理中} \end{cases}$$

n_1 : ステーション1上の優先呼の数

n_2 : ステーション2上の非優先呼の数

系の状態を状態ベクトル (q, n_1, n_2) で表わし、平衡状態において系が (q, n_1, n_2) とある確率を $P(q, n_1, n_2)$ で表わす。すると $2N_1N_2 + N_1 + N_2 + 1$ 個の状態方程式が得られるからこれを連立させて数値的に解き、各状態確率 $P(q, n_1, n_2)$ を求め、 $P(q, n_1, n_2)$ から各種特性を計算する。

4.1.3 結果の比較

図3のモデルと3.4で述べた漸近近似解法と4.1.2の方法で解き求めた大数時間特性を図4, 図5に示す。図3のステーション2の呼の平均通過時間は一定であるから、図4, 5ではステーション1の平均通過時間を示した。

これらの図から以下の観察が得られる。

i) 近似解析による優先呼の応答時間 T_1 の精度は非優先呼の多重度 N_2 が増大するにつれて悪化し、非優先呼の応答時間 T_2 の精度は優先呼の多重度 N_1 が増大するにつれて悪化する。いずれも optimistic である。

ii) 近似解析による優先呼の応答時間 T_1 の近似誤差は N_1 を大きくすると小さく、非優先呼の応答時間 T_2 の近似誤差は N_2 を大きくすると小さくなる。

これらの傾向は3.5で予測した傾向と一致する。また図の結果から、この例では3.5で述べた近似のうち②③に比べて①の誤差が支配的であることがわかる。すなわち、非割込の優先処理方式では優先呼の多重度を増すと非優先呼の仮想保留が増し、逆に非優先呼の多重度を増すと優先呼の仮想保留が増大するため、実サーバ、写像サーバの保留時間の変動係数が増大する。これは3.5①のように近似しているため①の結果となり、結果は optimistic である。

またii)の傾向は①の傾向が N_1 大、又は N_2 大で弱められることを示している。すなわち、 N_2 を固定し、 N_1 を増大すると、優先呼がサーバを使用する割合が大きくなり、個々の優先呼からみた時の仮想保留は小さくなるため優先呼のサーバ保留時間は指数分布に近づく。また、 N_1 を固定した時は N_2 を大きくするにつれて非優先呼のスループットが増して、非優先呼当りの仮想保留が減少し、非優先呼のサーバ保留時間が指数分布に近づくことを意味している。

なお、図3においてステーション1のサービス率 μ_1, μ_2 を小さくするにつれて近似誤差が大きくなるが、この傾向は3.5⑥の考察の通りである。

4.2 シミュレーションとの比較(5ステーションの場合)

4.2.1 解析モデル

図6に示すような5つのステーションからなる網型待ち行列を解析する。

各パラメータの値を表1に示す。本モデルは2つのプロセッサ、すなわちCPU、CCP(通信制御プロセッサ)をもつ現用の科学技術計算システムをモデル化したものである。

3.4で提案して漸近近似解法の適用性を調べるため、近似解法とシミュレーションとの結果を比べる。また同時に、他のスケジュール方式と非割込み優先処理方式と比較する際の評価方法として本近似解法の有用性をみるため、他のスケジュール方式との評価結果を合せて示す。

すなわち、図6のモデルにおいてプロセッサ(CPU, CCP)のスケジュール方式として、以下の4つの場合と比較する。

- ① 先着順サービスの場合(シミュレーション)
- ② PSの場合(BCMP型待ち行列解析パッケージのBSEによる数値解)
- ③ 割込み優先処理の場合(本近似解法とシミュレーション)
- ④ 非割込み優先処理の場合(本近似解法とシミュレーション)

なお、シミュレーションではGPSを用いた。

4.2.2 評価結果と考察

評価結果を図7、図8に示す。これらの結果から以下のことが観察される。

(1) 本近似解法とシミュレーションとの比較

非割込み優先方式に対する近似解析とシミュレーション結果は、シミュレーション誤差も考慮に入れるとかなりよく合っていると考えられる。例えば、図7からは会話処理(優先呼)について前の例題と同様の傾向があり、3.5④⑤の考察と一致する。

なお、本モデルでは近似解法は3桁の精度で解が収束するのに5〜7回の繰返し計算を必要とした。計算時間は3MIPSのコンピュータを用いて、ポイントデータ当り、シミュレーションが約3600秒要したのに対して、近似解析はいずれも数秒程度であった。

次に割込み優先方式についてみると、この場合も本近似解法の結果はシミュレーションとかなりよく一致していることがわかる。この場合、収束も早く、3桁の精度の解が2〜3回の繰返し計算で得られた。

(2) スケジュール方式比較

図7、図8をみると、非割込み優先方式に対する本近似解析の結果は、優先呼の多重度の小さい領域では先着順方式に近く、 N_i の大きい領域では割込み優先方式に近づいている。これは多重度の小さい領域では仮想保留が小さいため、非優先呼がサーバをつかむチャンスが大きいが、多重度が大きくなると、優先呼が連続して処理される確率が増大し、非優先呼がサーバをつかむチャンスが急速に小さくなることを示している。

このように近似解析の結果は他のスケジュール方式と比べ、非割込み優先方式の特性を十分表わしていることがわかる。従って、本例からも本近似解法が方式比較をする上で、従来のシミュレーションに代わる手法として有効であることが確かめられた。

5. あとがき

異なるワークロードの2呼種を処理するクローズド型の待ち行列網において、任意の数のサービスステーションで非割込み優先処理の行なわれる場合をとりあげ、その漸近近似解法を提案した。本近似解法の特徴は以下の点に要約される。

- ① 非割込み優先処理の行なわれるサービスステーションを優先呼のみを処理する実サーバステーションと非優先呼のみを処理する写像サーバステーションに分離したこと。
- ② 分離して設定された各サービスステーションの平均サーバ保留時間とそとのサービスステーションの優先呼、非優先呼の使用率との関係を仮想保留、サーバコールの概念を用いて求めたこと。
- ③ ①によりBCMP化された解析モデルで、②で得た関係を用いて実サーバと写像サーバの平均サーバ保留時間を漸近的に決定する方法を示したこと。

本近似解法はその適用領域として優先呼、非優先呼の多重度が1に比べて十分大きいことを仮定しているが、これらの多重度が1に近い時でも、ジョブ当りの負荷（呼当りの全処理時間に対する非割込み処理のサービスステーション上での処理時間の比）が小さい時にはかなり良い精度が得られることがわかった。その結果、先着順、FS、割込み優先処理など他のスケジュール方式との性能比較を行なう場合も十分有効な結果が得られる。

本近似解法はまた、割込み優先処理ステーションと非割込み優先処理ステーションの混在する待ち行列網に対しても容易に拡張されることを示した。

ステーション数が少なく多重度の小さい領域では直接数値解析を試みることも比較的可能であるが、今後は全領域をカバーする解法が確立する必要がある。

参考文献

- (1) F. Baskett, K.M. Chandy, R.R. Muntz, and F.G. Palacios, Open Closed and Mixed Networks of Queues with Different classes of Customers, J. of ACM, Vol. 22, No. 2, p. 248~260 (1975)
- (2) M. Reiser and H. Kobayashi, Queueing Networks with Multiple Closed Chains; Theory and Computational Algorithms, IBM J. of Res. and Dev., Vol. 19, No. 3, p. 283~294 (1975)
- (3) M. Reiser, Interactive Modeling of Computer Systems, IBM SYST J., Vol. 15, No. 4, p. 309~329 (1976)
- (4) K.C. Sivrik, Priority Scheduling Disciplines in Queueing Networks Models of Computer Systems, IFIP Congress preceedings, p. 565~570 (1977)
- (5) 池原, 山田, 汎用ネットワーク型待ち行列解析プログラムとEC, 通研実報, Vol. 29, No. 5, p. 1051~1077 (1980)

Table 1. Parameter Value Table

	Device Name	No. of I/O Units r_i		Average Server Holding Time h_{ij}	Average No. of visits to station i /job	Transition Probability P_{ij}
Service Station 1	CPU	1	class 1	13.9	23.9	—
			class 2	21.2	809	—
Service Station 2	Drum	5	class 1	17.2	11.5	0.494
			class 2	17.2	251	0.310
Service Station 3	Disk Pack	14	class 1	61.7	8.4	0.361
			class 2	61.7	410	0.506
Service Station 4	Magnetic Tape	9	class 1	55.9	2.4	0.102
			class 2	55.9	147	0.182
Service Station 5	Communication Control Processor	1	class 1	92.8	1.0	0.043
			class 2	2280	1.0	0.0012

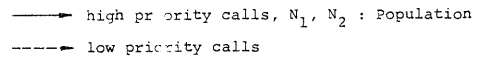
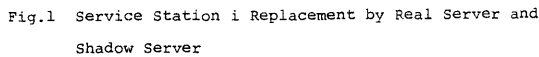

$$\begin{aligned} \mu_{11} &= 20, \mu_{21} = 2 \\ \mu_{12} &= 12, \mu_{22} = 1 \end{aligned}$$

Fig.3 Example with 2 Service Stations

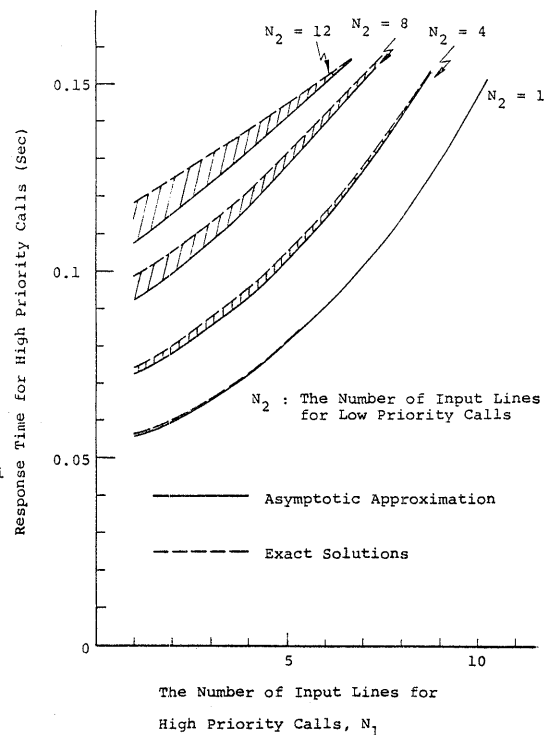
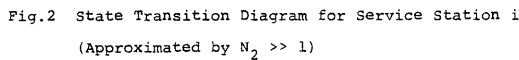
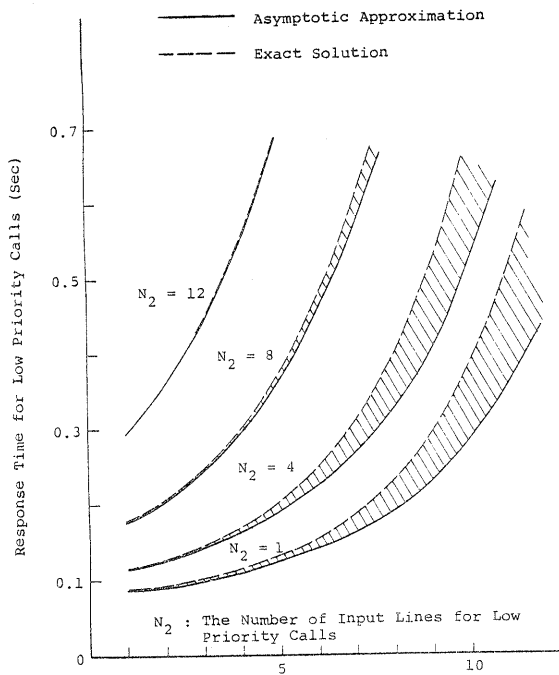


Fig.4 Response Time for High Priority Calls



The Number of Input Lines for High Priority Calls $> N_1$

Fig.5 Response Time for Low Priority Calls

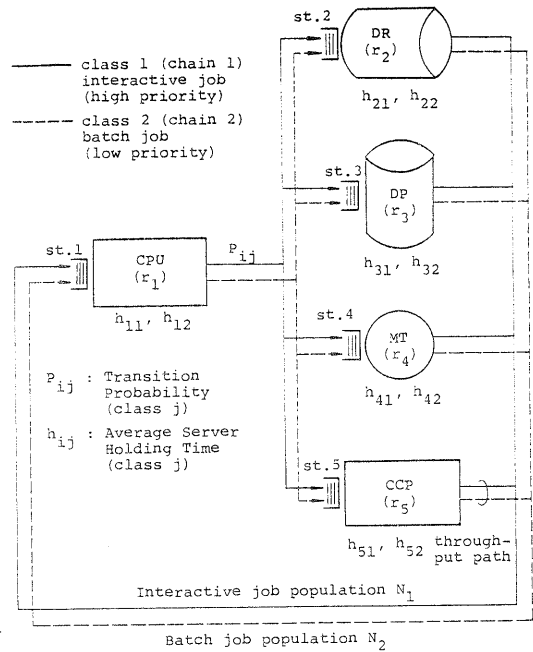


Fig.6 System Model with Interactive and Batch Jobs.

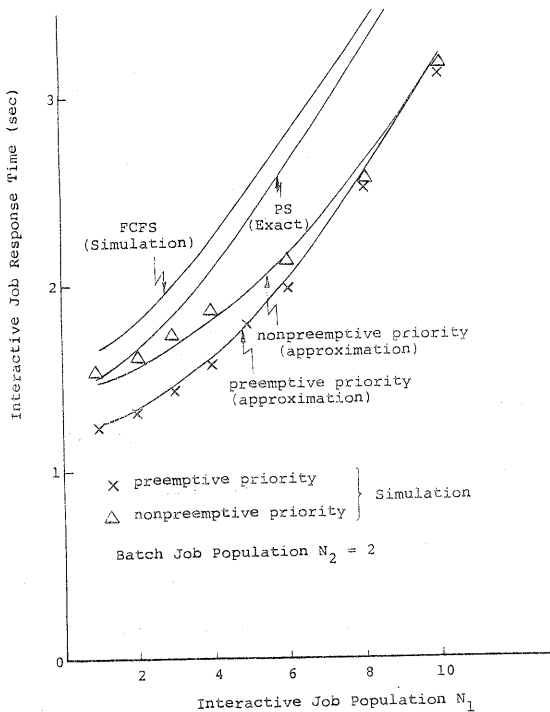


Fig.7 Response Time Comparison with Simulation

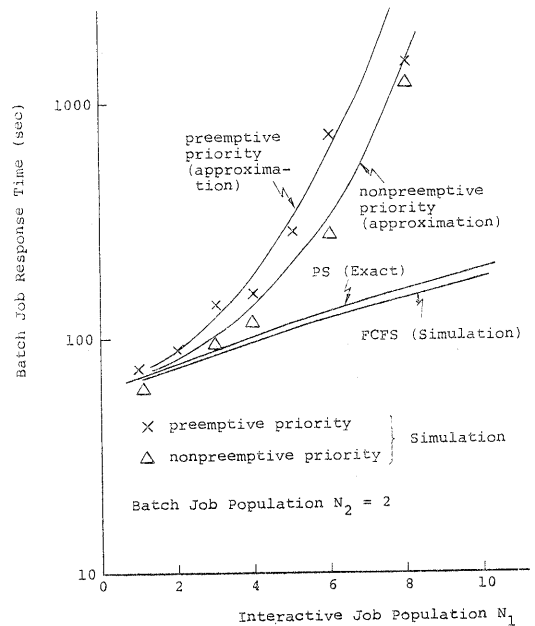


Fig.8 Response Time Comparison with Simulation