

資源要求のある待ち行列網のモデル化と評価

3B-2

木下 俊之
(株)日立製作所
システム開発研究所

高橋 幸雄
東京工業大学

梅原 元
東京工業大学

1. はじめに

計算機システム内には、一般に複数のジョブが存在し、システムの性能はこれらのジョブ間の相互作用によって影響される。例えば、あるジョブがファイルを更新する際、関連するファイルを排他的に使用しなければならない。つまり、あるジョブがこれらのファイルを使用している間は、他のジョブからのアクセスは禁止される。このようなアクセスのぶつかりは、システムの性能を低下させる要因となる。このように複数のジョブから排他的に使用される計算機の要素を、「排他的使用資源」あるいは単に「資源」と呼ぶ。

これまで待ち行列網では、この資源を近似的にしか扱えなかった¹⁾。発表者は、この資源要求のある待ち行列網をマルコフ連鎖で記述し、その平衡方程式を数値的に厳密に解くことで、資源がシステムのボトルネックになる限界について解析した²⁾。この前回の結果をさらに詳細化して、本発表では、ひとつのジョブの動きをマルコフ連鎖で記述し、資源へのアクセスのぶつかりがジョブの応答時間に与える影響について解析した。一般に、資源へのアクセスのぶつかりはマルコフ性を持たない。そこで近似のマルコフ連鎖を導入して、資源待ちにより応答時間が著しく悪化することを、定量的に確認した。

2. 資源要求のあるセントラルサーバモデル

資源要求のあるセントラルサーバモデルでは、CPU処理の完了時に、資源を獲得したり解放するかを確率的に決定し、さらに次に遷移するノードを確率的に決定する。そして資源獲得の要求時に、既に他のジョブによって資源が取られていれば、後から資源を要求したジョブは資源待ちキューに入って資源が解放されるのを待つ。その後、資源の解放時に資源を獲得して、後続の処理を実行する(図1)。

以下、次の記号を用いる。

N : ジョブ数

M : I/Oノード数(ゆえに、全ノード数は $M+1$)

(CPUノード: $m=0$, I/Oノード: $m=1 \sim M$)

μ_0 : CPUノードでのサービス率

μ_m : I/Oノード m でのサービス率 ($m=1 \sim M$)

$p^{nn}, p^{nr}, p^{rn}, p^{rr}$: CPUノードでのサービス終了前後でジョブが資源を獲得/解放する確率。二重添え字の、前者はCPUサービス中のジョブの資源確保の状態を、後者はCPUサービス終了後の資源確保の状態を表わし、'n'は資源を保持していないことを、'r'は保持していることを表わす。

p_m^n, p_m^r : CPUノードからノード m への遷移確率 ($m=0 \sim M$)。右上の添え字 'r', 'n' は、CPUサービス終了時の資源の有無を表わす。

CPU → CPU の遷移が起きた時、それまでジョブは

終了し、新たなジョブが生成されるものとする。従って p_0^n はジョブの完了確率である。また、 $p_0^r=0$ とする。

全てのノードと資源待ちキューで、FCFS (First Come First Served)スケジュールが行なわれる。またサービス時間は負の指数分布に従い、互いに独立である。

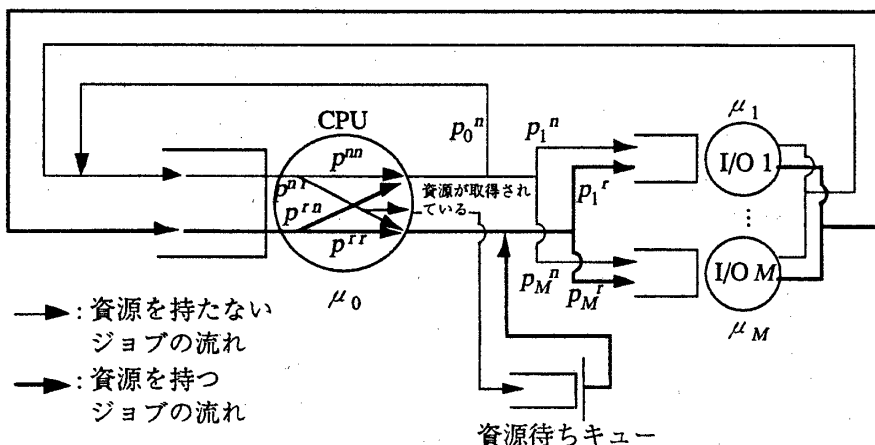


図1 資源要求のあるセントラルサーバモデルの構成 (資源数=1 の場合)

3. 応答時間解析のための近似マルコフ連鎖

図1の待ち行列網を記述するマルコフ連鎖を M_r とし、 M_r の諸量を次のようにおく。

u_{EC} : ひとつのジョブが、CPU → I/O(→ CPU)と遷移する回数を表わす確率変数

u_{RQ} : ひとつのジョブが、資源待ちキューにつながれた回数を表わす確率変数

t : ジョブの応答時間を表わす確率変数

T_{EC} : ジョブが CPU → I/O → CPU と一巡するのに要する平均時間

T_0^n : 資源を持たない1回の平均CPU滞在時間

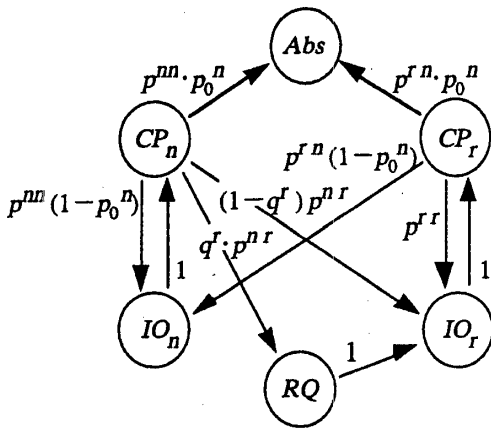
T_{RQ} : 1回の平均資源待ちキュー滞在時間

q^r : CPUサービス終了時にジョブが資源を要求した時、資源待ちキューにつながれる条件付き確率
この q^r は、次のように定義される。

$$q^r = \frac{\beta}{\alpha} \left\{ \begin{array}{l} \alpha: M_o \text{で、資源を持たないジョブがCPUノードの先頭にいる状態確率の合計} \\ \beta: M_o \text{で、資源を持たないジョブがCPUノードの先頭において、かつ他のジョブのひとつが資源を保持している状態確率の合計} \end{array} \right.$$

このマルコフ連鎖 M_o を解くことで、 T_{EC} 、 T_0^n 、 T_{RQ} ならびに q^r を求めることができる。

さて、ジョブの応答時間に資源へのアクセスのぶつかりがどう影響するかを調べため、資源待ちキューにつながれた回数で条件付けられた平均ジョブ応



CPn: 資源なしでCPUにいる状態
 CPr: 資源を持って
 IOn: 資源なしでI/Oのひとつに
 IOr: 資源を持って
 RQ: 資源待ちキューにいる状態
 Abs: 吸収状態(ジョブの完了)

図2 近似マルコフ連鎖の遷移図

表1 資源待ちキューにつながることの応答時間への影響
 (N=6, M=2, $\rho_{RQ}=0.60$)

k	0	1	2	3	4
$P(u_{RQ}=k)$	0.727	0.217	0.045	0.0094	0.0019
$E(u_{EC} u_{RQ}=k)$	0.726	0.215	0.046	0.0101	0.0022
$E(u_{EC} u_{RQ}=k)$	2.42	6.59	11.80	17.01	22.22
$E(t u_{RQ}=k)$	2.44	6.71	12.00	17.23	22.57
$E(t u_{RQ}=k)$	9.70	35.24	64.34	93.45	122.55
	10.05	34.80	63.69	92.42	121.75

(上段: 近似解析結果、下段: シミュレーション結果)

答時間 $E(t|u_{RQ}=k)$ ($k=0,1,2,\dots$) を求めたい。そこで、ひとつのジョブの動きをマルコフ連鎖で記述したいが、あるジョブの資源待ちキューへの訪問と出発は、他のジョブのふるまいに依存する。つまり資源へのアクセスのぶつかりに遭遇した時のジョブの動きを、そのジョブ自身のふるまいに基づくマルコフ連鎖で記述することは、厳密にはできない。しかし上記の q^r を用いて図2のようなマルコフ連鎖を考えると、資源アクセスのぶつかり時を含むジョブの動きを、マルコフ連鎖で近似的に記述することができる。

この近似マルコフ連鎖を解くことにより、資源待ちキューにつながれた回数による u_{EC} の条件付き平均値 $E(u_{EC}|u_{RQ}=k)$ が求められるので、近似的に

$$E(t|u_{RQ}=k) \approx T_{EC} \cdot E(u_{EC}|u_{RQ}=k) + T_0^n + T_{RQ} \cdot k \dots (*)$$

($k=0,1,2,\dots$)

と表わされる。

4. 数値解析

式(*)に基づく数値解析結果と対応するシミュレーション値を、表1に示す。両者の値は良く一致しており、上記の近似マルコフ連鎖が良い精度を持つことを示している。この例は資源がボトルネックになり始めのものであるが、これより次のことが判かる。

- (1) 約73%という多くのジョブが資源アクセスのぶつかりに遭遇しないにもかかわらず、資源がボトルネックとなり得る。
- (2) 一方、いったん資源待ちキューにつながると、ジョブの応答時間は著しく大きくなる。これが、資源がボトルネックとなる直接の原因である。

5. おわりに

資源要求のある待ち行列網において、資源要求のぶつかりがジョブの応答時間に及ぼす影響を調べた。今後は、資源が複数の場合を扱う予定である。

[参考文献]

- 1) C.H.Sauer: Approximate Solution of Queueing Networks with Simultaneous Resource Possession, IBM Journal of Res.& Dev., Vol.25, No.6, (1981.11)
- 2) 木下、高橋: 資源要求のある待ち行列網のモデル化の一提案、情報処理学会第51回全国大会論文集(4)、平成7年後期(1995.9)