

サービスネットワークシステムにおける端末機制御

東京工業大学大学院システム科学専攻 潘 煥旭

(修士論文 指導教官 森村英典 教授)

1. はじめに

現代の社会は高度情報化へ向かって急速に進展しており、すでに多くのオンライン・タイムシェアリングサービスと呼ばれるサービスネットワークシステムが利用されている。たとえば、銀行の現金自動支払いサービス、図書館のオンライン情報検索などがある。これらのシステムにおいては、大型コンピュータが各地にある何百、何千台という端末機からの業務を同時に処理している。しかし計算機の容量に限度があるために、システムが超負荷状態になると、ふくそうの発生によるCPU処理率の急激な低下を考えざるを得ない。このような時、システムの能力を最大限に発揮するためには、稼働中の端末機の一部を休止させることが役立つかもしれない。本論文では、このような端末機数の調節という簡単に実現できる対策が有効でありうることを理論的に解明する。

2. サービスネットワークシステム

本論文で取扱うサービスネットワークシステムはホストコンピュータ(CPU)とM個のサービスステーションからなる。各ステーションにそれぞれm個の端末機が設置されている。

各ステーションでの客の到着は独立で同一の到着率 λ を持つポアソン過程にしたがう。客のサービス時間は入力時間、CPU処理時間、出力時間の和である。入力時間と出力時間はそれぞれ平均 $1/\mu_1$ と $1/\mu_3$ の指数分布にしたがう確率変数である。CPU処理時間はシステムの

状態によって変化するものであるが、客がCPUにもたらすサービス要求量はその客の持つ固有量で、平均 $1/\mu_2$ の指数分布にしたがうとする。CPUの処理率(単位時間内で完成するサービス量)はCPUの中にあるジョブの数、つまりCPU処理を受けている客の数 n に依存し、 $S(n)$ で表わされる。一般に $S(n)$ は図1のような形をとる。それゆえ、システムの状態に応じて、各ステーションで何台の端末を稼働させるのが最適であるかという問題が生じてくる。

3. システムの解析

ステーションの数が一般に大きいために、システム全体を1つの確率過程と見なして解析することは非常に困難である。ここでは、この大規模システムをコンピュータシステムと各ステーションにおける待ち行列システムに分解して解析する。以下、システムが平衡状態にあると仮定する。

コンピュータシステムにはCPUとすべての端末が含まれ、その状態は $n=(n_1, n_2, n_3, n_4)$ で表わされる。ここで n_1, n_2, n_3, n_4 はそれぞれ入力作業中の端末数、CPU処理中の端末数、出力作業中の端末数、暇な端末数を表わしている。明らかに、 $n_1+n_2+n_3+n_4=Mm$ 。

$\{p_i\}$ を各ステーションの系内客数の平衡分布、 $P^{(l)}(i)$ を l 個ステーションの使用末端数が i である確率とする。 $T_{jk}(n_1, \dots, n_j, \dots, n_k, \dots, n_j) = (n_1, \dots, n_j-1, \dots, n_k+1, \dots, n_j)$ 、 $C(i) = P^{(1)}(i)P^{(M-1)}(Mm-n_4-i)/P^{(M)}(Mm-n_4)$ とおいて、 n の推移率は次のように求められる。

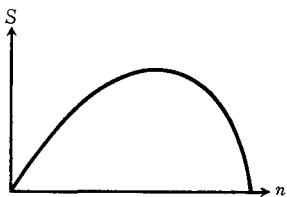


図 1

$$\begin{cases} q(n, T_{12}n) = \mu_1 n_1 \\ q(n, T_{23}n) = \mu_2 S(n_2) \\ q(n, T_{31}n) = \mu_3 n_3 C(m) \\ q(n, T_{34}n) = \mu_3 n_3 \sum_{i=0}^{m-1} C(i) \\ q(n, T_{41}n) = \lambda M \sum_{i=0}^{m-1} C(i) \end{cases} \quad (1)$$

$C(i)$ は全システムの使用末端総数 $= Mm - n_4$ の条件付きでの特定ステーションの使用末端数 $= i$ の確率を表わしているが、 M が大きければ条件のつかない確率 $p_i (i < m)$ あるいは $\sum_{k=m}^{\infty} p_k (i=m)$ で近似できよう。そうすると、 n を近似的に closed migration process [1] と見なすことができる。そこで、migration process についての既知の結果を用いて、CPU 処理中の端末総数の近似的な分布を求めると次のようになる。

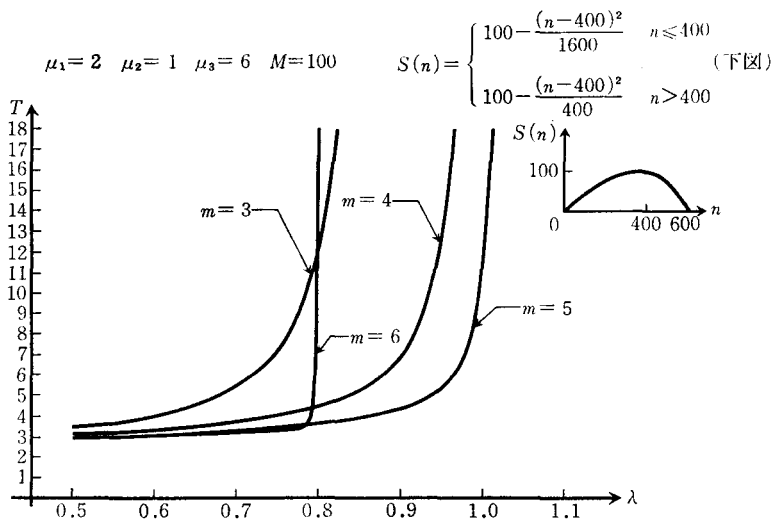


図 2

$$\pi_2(n) \approx \frac{(\lambda M)^n}{\mu_2^n \prod_{r=1}^n S(r)} \sum_{k=0}^{M-m-n} \frac{1}{k!} \left[\lambda M \left(\frac{1}{\mu_1} + \frac{1}{\mu_3} \right) \right]^k \quad (2)$$

各ステーションにおける待ち行列システムは端末機をサービス窓口とする $M/G/m$ 型システムと見なせる。それで、Kimura による近似式 [2] を利用して、近似的な平均待ち時間を次式で計算する。

$$W \approx \frac{1+C_v^2}{2} \left[\frac{1-C_v^2}{1+(1-\rho)(m-1)} \frac{\sqrt{4+5m-2}}{16m\rho} + C_v^2 \right]^{-1} \times \frac{m^{m-1} \rho^m}{\mu m! (1-\rho)^2} \left[\sum_{n=0}^{m-1} \frac{\rho^n}{n!} m^n + \frac{m^m \rho^m}{m! (1-\rho)} \right]^{-1} \quad (3)$$

ここで、 C_v と ρ はそれぞれサービス時間の変動係数と端末機の利用率を表わし、(2) で計算できる。

4. 数 値 例

図 2 に $\mu_1=2$, $\mu_2=1$, $\mu_3=6$, $M=100$,

$$S(n) = \begin{cases} 100 - \frac{(n-400)^2}{1600}, & n \leq 400 \\ 100 - \frac{(n-400)^2}{400}, & n > 400 \end{cases}$$

の時の $m=3, 4, 5, 6$ のそれぞれについて、客の平均システム時間 T を λ に対して示す。 T を評価基準とすれば、 λ が大きい時、 $m=5$ が最適であることがわかる。

5. 近似精度の検討

式 (2) の近似誤差を具体的に評価することはできない

が、システムが平衡状態にあるならば、全システムへの到着率と退去率とは一致するはずであるという事実にもとづいて、(2) を用いて計算した CPU スループット $R = \sum_{n=0}^{M-m} [\pi_2(n) \mu_2 S(n)]$ を総到着率 λM と比較することにより、その近似精度を間接的に確かめることができる。表 1 は (2) で計算した前例の ρ と R を示しているが、 R と λM が大体一致していることはわかる。(システムが平衡状態にあるという前提の下で解析を行なったため、表 1 中のかっこのない数値だけに注目する)

6. $\rho \approx 1$ と $\rho \geq 1$ の場合

表 1 で $\rho \approx 1$ のところでは、 R と λM がややはなれていますが、ここで $\rho \approx 1$ および $\rho \geq 1$ の場合について、前の解析を改善する。

(1) を導くところで、到着あるいは退去の直後時点に暇な端末機があるかどうかを考える必要があったが、 $\rho \approx 1$ と $\rho \geq 1$ の場合はほとんど暇な端末機がないから、コン

表 1 利用率 ρ (上) とスループット R (下)

m	3	4	5	6
λ				
0.7	0.760 70.000	0.570 70.000	0.456 70.000	0.380 70.000
0.8	0.912 79.182	0.688 80.000	0.550 80.000	(1.427) (78.900)
0.9	(1.050) (82.922)	0.837 89.923	0.671 90.000	(78.326) (1.439)
1.0	(1.174) (83.805)	0.999 95.735	0.907 98.600	(87.084) (1.436)

ピュータシステムの状態を $n=(n_1, n_2, n_3, n_4)$ のかわりに $n=(n_1, n_2, n_3)$ のように取っても差支えない。この n がまた closed migration process となっている。そして、平衡分布 $\pi_2(n)$ は次式で求められる。

$$\pi_2(n) = \frac{\frac{1}{\mu_2^n \prod_{r=1}^n S(r) (Mm-n)!} \left(\frac{1}{\mu_1} + \frac{1}{\mu_3}\right)^{Mm-n}}{\sum_{n_2=0}^{Mm} \frac{1}{\mu_2^{n_2} \prod_{r=1}^{n_2} S(r) (Mm-n_2)!} \left(\frac{1}{\mu_1} + \frac{1}{\mu_3}\right)^{Mm-n_2}} \quad (4)$$

表2に(4)で計算した前例の R を示す。表1の場合と違って、 R が λ に依存しない。 R を評価基準とすれば、 $m=4$ が最適で、 $m=6$ では、かなり悪い状況になる。

7. おわりに

サービスネットワークシステムを理論的に解析したが、端末機制御が有効であることがわかった。以下に、最適端末機数算出の手順をまとめる。

μ_i ($i=1, 2, 3$), M と $S(n)$ は既知とする。与えられた λ に対し、端末機数候補のそれぞれについて、(2)で ρ

表2 $\rho \approx 1$ と $\rho \geq 1$ の時の R

m	3	4	5	6
R	84.671	97.330	96.656	1.414

を計算する。 ρ が小さい (たとえば、 $\rho < 0.95$) となる候補があれば、その中で、 T を最小にする端末機数を最適数とする。すべての候補に対し、 $\rho \approx 1$ あるいは $\rho \geq 1$ なら、(4)で計算した R を最大にする端末機数が最適である。システムを常に良好状況に保つために、 λ の変化に応じて運転端末機を適宜休止させることが適切であろう。

参考文献

- [1] F. P. Kelly: Reversibility and Stochastic Networks, John Wiley & Sons, 1979
- [2] T. Kimura: A Two-Moment Approximation for the Mean Waiting Time in the GI/G/S Queue, Research Reports on Information Sciences, T. I. T., B-154 (1984)

〈書 評〉

日科技連OR実習部会 編

経営改善のシナリオ — 3つのケーススタディー

日科技連出版社 昭和60年8月 184頁 定価2000円

本書は著者らが10余年間、日科技連の主催するOR教育コース総合実習の中で教材として使用してきたものをまとめたものである。著者らはシナリオライティングの手法を使って、はじめに企業の中で問題意識が発生し、最後にシステムの改善計画が報告書にまとめられるまでのプロセスを何人かの人物の活動や会議の記述を通していきいきと浮かびあがらせることに成功している。

最近では、OR実施理論の研究も盛んになり、ORが企業の中でどのように導入され、定着し、組織革新を実現していくのかの科学理論の体系化も行なわれてきている。そのような状況で、本書をOR実施理論との関連で読んでみるののおもしろいかもしれない。そこで以下では昨年10月の秋季OR学会での松田武彦先生による講演、「OR実施理論研究の基本方向」のフレームワークで本書を論評することにする。

本書は、以下の各章で構成されている。第1章 修理サービス体制の改善、第2章 取替再生システムの改善

第3章 部品供給システムの改善、第4章 問題を解決する方法。

第1章は、N重工業が販売し建設機械の修理サービス部門の改善をテーマとしている。まず「問題設定過程」における「問題意識発生段階」をみてみよう。話は取締役会での経営刷新の方針を受けて総務部が招集した部長会の場面からはじまる。各長の発言から、問題が徐々に姿を見せはじめる。次に与件と政策変数を明確にする「問題領域の想定段階」はどうか。著者らは政策変数として、在庫管理水準、サービスマン数、サービスセンタ数を上げている。ここで、政策変数に何をとりかに対応して、「①現状の体制をあまり変えず、コストもかけずに効果のあるもの、②ある程度コストはかけても改善効果の大きいもの、③将来に向けての先行投資的に多少コストをかけるもの」を区別することにより、著者らは現実の問題の解にはコスト・ベネフィットによるレベルがあることを主張している。