7T-6 仕事量保存則を利用した フィードバック待ち行列の解析法

住田修一

ＮＴＴ電気通信研究所

1．まえがき　仕事量保存則（文献(1) p.172）は、リトルの公式と同様に、種々の待ち行列システムに対して広く成り立つ性質の１つである．最近，この保存則を利用して解析する方法が、ある種のモデルの解析（厳密解や近似解の導出など）に有効であることが示され、(2)~(5) 待ち行列モデルの新しい解析手法として注目されつつある．本稿では、仕事量保存則をフィードバックのある待ち行列の解析に応用し、待ち行列モデルの解析に如何に有用であるかを示す．

2．モデル　本稿で考察するフィードバックのある待ち行列モデルを図１に示す．モデルを以下に説明する．(i) 客は、到着率λのポアソン過程に従い単一サーバの系に到着する．(ii)系に到着した客は、第１回目のサービスを受けるため待ち行列Ｑに並ぶ．第１回目のサービスが終了すると、第２回目のサービスを受けるため直ちにフィードバックする．第２回目のサービスが終了すると、系から退去する．フィードバックに要する時間は０とする．(iii) 第１回目、第２回目のサービス時間は、互いに独立で、それぞれ、平均h_1、h_2の指数分布に従う．(iv)待ち室の容量は無限とする．(v) 定常状態の存在する条件　$\rho_1+\rho_2<1$ を仮定する．ここで、$\rho_1=\lambda h_1$、$\rho_2=\lambda h_2$である．(v) サーバが用いるサービス規律は仕事量保存形（work-conserving）とする．先着順、割込み継続形優先権、非割込み形優先権などは仕事量保存形である．詳細は文献(1) p.173 を参照．

サービス規律により待ち行列構成は異なり、図１は、フィードバックした客が待ち行列の最後部に並ぶ場合を図示している．

3．仕事量保存則を利用した解析法

3.1 フィードバック待ち行列の仕事量保存則

本節では、節２で述べたモデルに対して仕事量

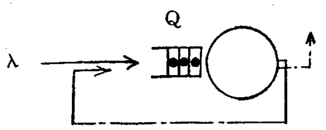

図１　フィードバック待ち行列（フィードバックが１回）

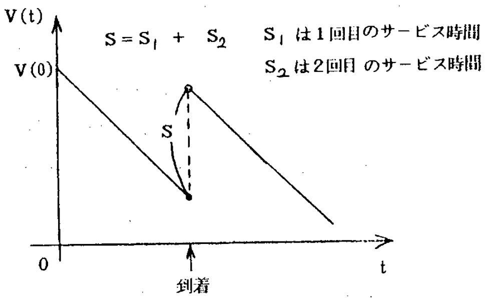

図２　系内仕事量の変化

保存則を利用した解析方法を提案する．ここで、仕事量保存則とは、客の到着過程とサービス時間分布が与えられたとき、平均系内仕事量（サービス中の客の残余サービス時間と待合せ中のすべての客の残余サービス時間の和の平均）が、どのような仕事量保存サービス規律に対しも常に一定の値をとる性質のことを言う．すなわち、平均系内仕事量は、サービス規律が仕事量保存形である限り、客の到着過程とサービス時間分布にのみ依存する量であり、サービス規律とは独立であることを意味する．

図１のモデルではフィードバックがあるため、このシステムの系内仕事量を以下のように定義する．まず、系内にいるある客に着目する．この客が第２回目のサービスを終了して系から退去するまでにサーバから受ける残りのサービス時間を、この客の残余サービス時間と定義する．１度もサービスを受けていない客の残余サービス時間は、第１回目と第２回目のサービス時間の和となる．残余サービス時間をこのように定義すると、系内仕事量は、系内にいるすべての客の残余サービス時間の和と定義される．

以上のように定義した系内仕事量の変化を考える．時刻ｔにおける系内仕事量をV(t)とする．V(t)の変化の一例を図２に示す．系外からの客の到着時点で、この客の第１回目と第２回目のサー

A method of analyzing a single-server queue with feedback based on the conservation law
Shuichi SUMITA
NTT Electrical Communications Laboratories

ビス時間の和だけ増加する．その他の時点では、傾き－１で減少する．サービス規律が仕事量保存形である限り、これらの変化はサービス規律によらない．すなわち、V(t)は、サービス規律によらず同じ値をとる．

図１のフィードバック待ち行列では、サービス時間は指数分布に従うので、仕事量保存則は次式のように表わされる．

$$(h_1+h_2)L_1(\theta)+h_2L_2(\theta)=V \qquad (1)$$

ここで、$L_1(\theta)$、$L_2(\theta)$ は、それぞれ、サービス規律 θ を用いたときの第１回目、第２回目のサービスを待っている客の平均数である．ただし、サービス中の客も含める．V は、平均系内仕事量である．平均系内客数は、サービス規律に依存するのに対して、平均系内仕事量は、サービス規律には依存しないことに注意されたい．

リトルの公式を使うと式(1) は更に次のように変形される.

$$(\rho_1+\rho_2)D_1(\theta)+\rho_2D_2(\theta)=V \qquad (2)$$

ここで，$D_i(\theta)$ は、第 i 回目(i=1,2) のサービスを待ち始めてから、そのサービスが終了するまでの時間の平均である．以下では遅延時間と呼ぶ．

3.2 平均系内仕事量の計算

本モデルでは，サービス時間は指数分布に従うのでV は次式により計算できる．

$$V=L_1(\theta)(h_1+h_2)+L_2(\theta)h_2 \qquad (3)$$

θ は仕事量保存形であれば何でもよいが、ここでは、プロセッサシェアリング(PS)に対してV を計算する．これは，θ をPSとすると本モデルはBCMP形待ち行列網（文献(1)p.94)の特別な場合とみなせ、既知の結果が使えるからである．BCMP形待ち行列網の結果より L_i(PS) は次式で与えられる．

$$L_i(PS)=\rho_i/(1-\rho) \qquad i=1,2 \qquad (4)$$

ここで、$\rho=\rho_1+\rho_2$．

式(3),(4) よりV として次式を得る.

$$V=(\rho_1h_1+\rho_1h_2+\rho_2h_2)/(1-\rho) \qquad (5)$$

3.3 平均遅延時間の導出法

式(2) を利用するとサービス規律が次の２つの場合に $D_i(\theta)$ がV を用いて表現できる．V は既に式(5) で与えられているので、結局 $D_i(\theta)$ が導出されたことになる.

(1) 第１回目のサービスが第２回目のサービスに対して割込み優先権を持つ場合(PP)

系に到着してから第１回目のサービスが終了するまでの平均時間 D_1(PP) は、M/M/1 の公式より

$$D_1(PP)=h_1/(1-\rho_1) \qquad (6)$$

式(6) を式(2) に代入すると

$$D_2(PP)=(V-(\rho_1+\rho_2)h_1/(1-\rho_1))/\rho_2 \qquad (7)$$

式(5) を式(7) のV に代入すると次式を得る．

$$D_2(PP)=(\rho_1h_1+h_2)/((1-\rho)(1-\rho_1)) \qquad (8)$$

(2)フィードバックした客が待ち行列の最後に並び先着順処理(FCFS)される場合

系外から到着する客の到着過程はポアソン過程に従うので、この客の到着直前の平均系内仕事量はV に等しい．[6] このV には、第１回目のサービスを待っている客の第２回目のサービス時間も仕事量として含まれている．このことに注意すると，サービス規律は先着順処理であるから、第１回目のサービスを受けるまでの待ち時間は、系内仕事量V から第１回目のサービスを待っているすべての客の第２回目のサービス時間を引いたものに等しい．第１回目のサービスを待っている客の平均をL_1(FCFS)とすると次式が成り立つ.

$$D_1(FCFS)=V-L_1(FCFS)h_2+h_1. \qquad (9)$$

式(9) のL_1(FCFS)にリトルの公式を適用すると次式を得る．

$$D_1(FCFS)=V-\rho_2D_1(FCFS)+h_1.$$

よって、次式を得る．

$$D_1(FCFS)=(V+h_1)/(1+\rho_2) \qquad (10)$$

式(5) を式(10)に代入すると

$$D_1(FCFS)=(\rho_2h_2+h_1)/((1-\rho)(1+\rho_2)) \qquad (11)$$

式(10)を式(2) に代入すると次式を得る.

$$D_2(FCFS)=\{(1-\rho_1)V-(\rho_1+\rho_2)h_1\}/\{(1+\rho_2)\rho_2\} \qquad (12)$$

式(5) を式(12)に代入すると

$$D_2(FCFS)=(\rho_1h_1+h_2)/((1-\rho)(1+\rho_2)) \qquad (13)$$

4．むすび

今後、更に本手法の適用域を拡張する．御意見を頂いた当研究所トラヒック研究室、橘田特別研究室の諸兄及びOR学会Q部会のメンバーの方々に深謝します．

文献：(1)E.Gelenbe & I.Mitrani:"Analysis and Synthesis of Computer Systems," Academic Press,1980. (2)G.Barberis:"A useful tool in the theory of priority queueing," IEEE Trans. Commun. ,COM-28, 9, pp.1757-1762, 1980. (3) S.Sumita:"A method for analyzing single server queueing models having two classes of customers,"Trans.IECE Japan, E68, 9, pp.555-556,1985.(4)S.Sumita:"An application of the conservation law to analyzing single server queueing systems with two independent input streams,"Trans.IECE Japan, E.69, 5, pp.628-637, 1986.(5)D.Everitt:"Simple approximation for token rings,"IEEE. Trans.Commun.,COM-34, 7, pp.719-721, 1986.(6)R.W.Wolff:"Poisson arrivals see time averages,"Oper. Res., 30, 2, pp.223-231,1982.