

講座

情報処理システムにおける待ち行列理論の応用 (2)\*

橋田 温\*\*

6. バッファメモリでの待ち行列

バッファメモリは情報を一時的に蓄積するメモリであり、データのアセンブル、速度変換、トラフィック変動の吸収などの役割をもつ。バッファメモリを特徴づけるパラメータとしては、(i)データの到着過程、(ii)データ単位、(iii)データの読出し方法、(iv)ハード的構成(個別か共用か)、などがある。バッファのハード構成法や容量などを定めるための尺度としては、オーバーフロー確率が良く用いられる。これは、到着データがバッファ一杯のために受け入れられない確率である。

データ長が一定か、データ長の分布を無視して個数のみで容量を定める場合は、データを客と見れば待ち行列モデルが適用できる。この場合を待ち行列形バッファと呼ぶ。これに対し、データ長の分布を考慮した場合は、1つのデータが漸次バッファを占有/解放する過程に注目すると、ダム過程(連続または不連続)に似たモデルとなる。この場合をダム形バッファと呼ぶ。

6.1 待ち行列形バッファ<sup>32)</sup>

一般に有限容量の待ち行列となるが、複雑な系に対しては、(a)無限容量待ち行列による近似、(b)損失系モデルによる近似、(c)拡散近似<sup>33)</sup>、などの近似法が用いられる。但し損失系モデルによる近似は、かなり危険側の場合があるので注意を要する。

(1) M/M/1形有限待ち行列

データの到着がポアソン過程で読み出しの処理時間が指数分布の場合は、無限容量の状態確率を求めて、それを容量の範囲で正規化すれば、有限容量の場合の状態確率が求められ、オーバーフロー確率  $P_{of}$  が計算される。

(例1) M/M/1(N): 式(3.5)から

$$P_{of} = p_N = \frac{(1-\rho)\rho^N}{1-\rho^{N+1}} \tag{6.1}$$

(例2) 共用バッファ: 図-13の如く  $n$  回線への送出力バッファを共用した場合、各回線への送出力待ちデータ数の同時確率  $p(x_1, x_2, \dots, x_n)$  は、

$$p(x_1, x_2, \dots, x_n) = \frac{\prod_{i=1}^n (1-\rho_i)\rho_i^{x_i}}{\sum_{x_1+\dots+x_n \leq N} \prod_{i=1}^n (1-\rho_i)\rho_i^{x_i}} \tag{6.2}$$

で与えられる。ここで、 $\rho_i = \lambda_i/\mu_i$ ,  $1/\mu_i$  は平均送出力時間である。これから、オーバーフロー確率は

$$P_{of} = \sum_{x_1+\dots+x_n=N} p(x_1, x_2, \dots, x_n) \tag{6.3}$$

で計算される。

(2) 処理時間が一定で周期的な場合

処理が終了した時点(周期時点の直前)におけるバッファ内のデータ数が  $j$  の平衡状態確率を  $p_j$  とすれば、有限のマルコフ連鎖の極限として求められる。すなわち、周期間の遷移に関し状態方程式が成り立ち、それを解くことにより  $P_{of}$  が求められる。

(例1) G/D(periodic)/s(N)<sup>34)</sup>: これは各処理周期で  $s$  個ずつデータを読み出す方式である。

1処理周期内に到着するデータ個数が  $k$  の確率を  $\pi_k$  とし、周期間の遷移を考えると次の状態方程式を得る。

$$\begin{cases} p_0 = a_s \pi_0 \\ p_n = a_s \pi_n + p_{n+1} \pi_{n-1} + \dots + p_{n+s} \pi_0 \quad (n \leq N-s) \\ p_n = a_s \pi_n + p_{n+1} \pi_{n-1} + \dots + p_N \pi_{N-n+s} \quad (N-s < n \leq N) \end{cases} \tag{6.4}$$

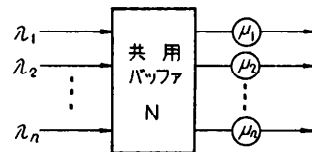


図-13 共用バッファ

\* Applications of Queuing Theory in Information Processing Systems by On HASHIDA (Musashino Electrical Communication Laboratory, N. T. T.)

\*\* 日本電信電話公社武蔵野電気通信研究所交換方式研究室

ここで、 $a_i = \sum_{j=0}^s p_j$  である。  $\{p_i\}$  から

$$P_{of} = 1 - \left( \sum_{i=0}^{s-1} i p_i + s \sum_{i=s}^N p_i \right) / \lambda \quad (6.5)$$

によってオーバーフロー確率が求まる。但し  $\lambda$  は 1 周期内におけるデータの平均到着数である。

到着過程のパラメータ  $\pi_k$  を変えれば種々のモデルに適用できる。例えばポアソン到着の場合は  $\pi_k = e^{-\lambda} \times \lambda^k / k!$  である。その他、複合ポアソン到着<sup>35)</sup>、ベルヌーイ到着<sup>36)</sup>、などが解析されている。

### (3) 有限待ち行列 M/G/1(N)

構成は簡単であるが処理時間が指数分布と異なるバッファに有効なモデルである。基本的待ち行列のためすでに解析されている。結果のみを述べると<sup>37)</sup>、

$$P_{of} = \frac{\bar{h} - (1 - \rho)\bar{\gamma}_N}{\bar{h}(1 + \lambda\bar{\gamma}_N)}, \quad \rho = \lambda\bar{h} \quad (6.6)$$

ここで、 $\bar{\gamma}_N$  は次式で計算される。

$$\bar{\gamma}_N = \frac{\bar{h}}{(N-2)!} \left[ \frac{d^{N-2}}{dx^{N-2}} \frac{1}{B^* \{ \lambda(1-x) \} - x} \right]_{x=0} \quad (6.7)$$

(例 1) M/D/1(N):

$$\bar{\gamma}_N = \bar{h} \left[ \frac{1}{1-\rho} - \sum_{i=0}^{\infty} e^{-i\rho} \frac{(i\rho)^{N-i-1}}{(N-i+1)!} \right] \quad (6.8)$$

(例 2) M/E<sub>k</sub>/1(N):

$$\bar{\gamma}_N = \bar{h} \sum_{j=0}^{\min(k, N-2)} \binom{N-2-j}{j} \sum_{i=I}^{ik} \binom{ik}{N-2-i-j} \times \left( 1 + \frac{\rho}{k} \right)^{(i+1)k+i-N+2} \left( -\frac{\rho}{k} \right)^{N-2-i} \quad (6.9)$$

但し、 $I$  は  $(N-2-j)/(k+1)$  以上の最小整数を表わす。

## 6.2 ダム形バッファメモリ

データの長さを考慮した場合は、待ち行列理論がそのまま使えない。しかし各データがそれぞれ全長が一度に到着するときは、次のような近似が可能である。いま無限容量バッファ内にデータが  $n$  個あり、各データ長を  $l_1, l_2, \dots, l_n$  とすると、バッファ内の全データユニット数  $X_n$  は  $X_n = l_1 + l_2 + \dots + l_n$  である。バッファ内のデータユニット数  $X$  の母関数  $G(z)$  を求めるため、データ数分布  $P_r \{n=j\}$  の母関数を  $P(z)$ 、データ長分布  $P_r \{l=k\}$  の母関数を  $K(z)$  とすると、 $G(z) = E_x [z^X]$ 、 $P(z) = E_n [z^n]$ 、 $K(z) = E_l [z^l]$  である。但し  $E_x$  等は  $X$  について期待値をとることを表わす。  $n$  に関して条件付期待値をとると、

$$\begin{aligned} G(z) &= E_n E_{X_n} [z^{X_n} | n] \\ &= E_n E_{l_1 + \dots + l_n} [z^{l_1 + \dots + l_n}] \end{aligned} \quad (6.10)$$

となる。ここで、各  $l_i$  は独立で同じ分布に従うため、

$$E_{l_1 + \dots + l_n} [z^{l_1 + \dots + l_n}] = \prod_{i=1}^n E_{l_i} [z^{l_i}] = K(z)^n$$

となる。したがって、

$$G(z) = E_n [K(z)^n] = P\{K(z)\} \quad (6.11)$$

を得る。  $G(z)$  からデータユニット数分布  $P_r \{X=i\}$  が求められると、

$$P_{of} = P_r \{X=N\} / \sum_{i=0}^N P_r \{X=i\} \quad (6.12)$$

でオーバーフロー確率が求められる。

さて、データの到着・送出に応じてバッファ占有量が漸増・減するモデルはダム過程となり、その解析は一般に困難である。解析されたモデルとしては、コアでのダイナミックアロケーションのモデル<sup>38)</sup>、一定速度でデータが送られて来て送出が間欠的なモデル<sup>39)</sup>、入力が間欠的な有限バッファモデル<sup>41)</sup>、複数の入回線を考慮した有限バッファモデル<sup>41)</sup>、送出開始時期がデータユニットの占有量で決る在庫形バッファモデル<sup>42)</sup>、などがある。

## 7. 端末制御における待ち行列

端末を制御する時間およびその制御開始時点がともに連続量である場合と、制御時間が一定でその開始時点が周期的である同期式の場合に大別される。

### 7.1 連続制御における多重待ち行列

通信制御装置が多数の端末を通信回線を介して制御する場合、各端末での要求の有無を順番に調べ、もし要求があれば処理して次の端末へ移る方法をポーリング (polling) 方式と呼ぶ。このような独立の複数個の待ち行列を順番に処理するモデルは多重待ち行列 (multiqueue) と呼ばれ、また扱者が待ち行列の間を移動する時間は移(歩)行時間 (walking time) と呼ばれる。多重待ち行列は、各端末での処理方法で次の 3 つに大別されよう。

- (a) 全処理式 (exhaustive type): 端末の要求をすべて処理してから次の端末へ移る。
- (b) ゲート式 (gating type): 制御が端末に到着したときまでに発生した要求のみを処理する。
- (c) 制限式 (limiting type): 並んでいる要求のうち先頭から一定個数までを処理する。

また端末を制御が移動する順番としては、一定方向の巡回式、優先順位による方式、確率的に移動する方式、などがある。

以下に、 $N$  端末よりなり、各端末の確率的特性が同じでかつ移行時間も端末に依らない場合、すなわち

対称形多重待ち行列に対し、巡回式の場合の結果を示す。

但し、各端末での要求の到着はポアソン過程 ( $\lambda$ )、処理時間分布を  $B(x)$ 、移行時間分布を  $U(x)$  とする。

(a) 全処理式<sup>43)</sup>:

制御が巡って来たときの端末に並んでいる要求数を  $L$ 、待ち時間を  $W$  とする。

$$\bar{L} = \lambda \bar{u} N / (1 - N\rho), \quad \rho = \lambda \bar{h} \quad (7.1)$$

$$\bar{W} = \left[ \frac{N\lambda \bar{u}}{1 - N\rho} \bar{h}^2 + \bar{u}^2 + \left\{ \frac{N(1+\rho)}{1 - N\rho} - 1 \right\} \bar{u}^2 \right] / 2\bar{u} \quad (7.2)$$

ここで、 $\bar{u}$ 、 $\bar{u}^2$  は  $U(x)$  の 1 次・2 次モーメントである。 $\bar{L}$  は端末でのバッファ容量推定に利用できる。

(b) ゲート式<sup>44)</sup>:

$$\bar{L} = \lambda \bar{u} N (1 - \rho) / (1 - N\rho) \quad (7.3)$$

$$\bar{W} = \frac{\bar{u}^2}{2\bar{u}} + \frac{(N-1)\bar{u}}{2(1-N\rho)} + \frac{N\lambda \bar{h}^2}{2(1-N\rho)} \quad (7.4)$$

(c) 制限式 (端末入力源 1)<sup>45)</sup>:

特に端末への要求発生が入力源 1 で、空き間隔が平均  $1/\lambda$  の指数分布の場合を示す。制御が或端末をはなれてから再びその端末に到着するまでの時間をサイクルタイムと呼び、その分布を  $C(x)$  とする。端末へ制御が到着したとき全端末で  $i$  個の要求がある平衡状態確率を  $q(i)$  とすれば、

$$\begin{cases} q(i) = \beta^i \prod_{t=0}^{i-1} (\alpha \beta^t - 1) q(0) \\ q(0) = \left[ 1 + \sum_{i=1}^{N-1} \binom{N-1}{i} \beta^i \prod_{t=0}^{i-1} (\alpha \beta^t - 1) \right]^{-1} \end{cases} \quad (7.5)$$

となる。但し、 $\alpha = 1/U^*(\lambda)^N$ 、 $\beta = 1/B^*(\lambda)$  である。これから、 $\bar{c}$  を  $C(x)$  の平均として、

$$C^*(s) = \sum_{i=0}^{N-1} \binom{N-1}{i} q(i) U^*(s)^N \beta^i (s)^i \quad (7.6)$$

$$\bar{c} = N\bar{u} + (N-1)\bar{h} \sum_{i=0}^{N-2} \binom{N-2}{i} q(i+1) \quad (7.7)$$

を得る。また平均待ち時間は次式で与えられる。

$$\bar{W} = \frac{\bar{c}}{1 - C^*(\lambda)} - \frac{1}{\lambda} \quad (7.8)$$

(d) 近似法<sup>46)</sup>:

多重待ち行列は、上述のモデルの他に種々考えられるが解析が困難である。簡単な結果を得るには、Leibowitz が用いた単一待ち行列で近似する方法が有効である。これは、或端末に注目して、その端末の要求の処理が終了した後サイクルタイムにおいて再びこの端末へ制御が戻ってくる単一待ち行列を考え、サイクルタイムの中に他端末での処理時間と移行時間をす

べて含める方法である。このサイクルタイムは、一般に他端末の状態に依存するが、他端末は平衡状態の確率で各状態にあると近似する。そうすると、例えば全処理式では、

$$C^*(s) = U^*(s)^N \cdot [G\{B^*(s)\}]^N \quad (7.9)$$

となる。ここで、 $G(z)$  は近似された単一待ち行列において制御到着時点の要求数分布に対する母関数であり、次の関係を満たす。

$$G(z) = U^*(\lambda(1-z))^N [G\{B^*(\lambda(1-z))\}]^N \quad (7.10)$$

これから平均待ち時間等が求められる。

$$\bar{L} = N\lambda \bar{u} / (1 - N\rho) \quad (\text{式(7.1)と同じ})$$

$$\bar{W} = \frac{1 + \lambda \bar{h}}{2} \cdot \frac{\bar{c}^2}{\bar{c}} \quad (7.11)$$

但し近似法では、処理時間の分散が大きいたまは、 $\bar{W}$  を過少評価する傾向がある。

## 7.2 同期式多重待ち行列

時間が一定長 (=1) のタイムスロットに分割され、各タイムスロットでは 1 データユニットずつ処理されるとする。この方式には、(i)一定周期ごとのタイムスロットが特定の端末へ割当てられている周期的時分割方式 (STDM, synchronous time division multiplexing)、(ii)空きタイムスロットを自由に使用できる非周期的時分割方式 (ATDM, asynchronous TDM)、(iii)ポーリング式でタイムスロットを順番に使わせる方式、などがある。これらは、賭博者破産問題 (gambler's ruin problem) の解を用いて、統一的に解析されている<sup>47)</sup>。以下に、端末  $i$  に制御が到着する直前の平均要求数 (端末  $i$  の)  $\bar{L}_i$  と、長さ  $m$  ユニットのデータの平均系内時間  $\bar{T}_i(m)$  について結果を示す。

(i) STDM:

各端末ごとに単一待ち行列となる。端末  $i$  にタイムスロット内に到着するデータ長の平均と分散を  $\mu_i$ 、 $\sigma_i^2$  とすれば、

$$\bar{L}_i = \frac{N\sigma_i^2}{2(1-N\mu_i)} + \frac{N\mu_i}{2} \quad (7.12)$$

$$\bar{T}_i(m) = \frac{N+1}{2} + m(N-1) + N \left( \bar{L}_i - \frac{(N+1)\mu_i}{2} \right) \quad (7.13)$$

(ii) ATDM:

例えばループ網で端末を接続する場合などのように、そのままでは端末からの要求が衝突しないようにタイムスロットを使用させる順番を定める必要がある。端末番号の若いものほど制御装置に近い、すなわち使用順位が上として、対称な場合を考えると、

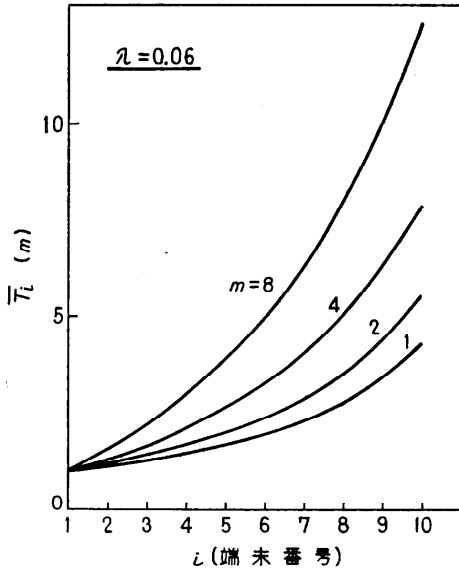


図-14 ATDM における Built-in Priority

$$\bar{L}_i = \frac{\lambda(i-1)\lambda^2 + \{1-(i-1)\lambda\}^2 + 1}{2(1-i\lambda)\{1-(i-1)\lambda\}^2} + \frac{1}{2} \cdot \frac{\lambda}{1-(i-1)\lambda} \quad (7.14)$$

$$\bar{T}_i(m) = \frac{m + i\lambda/2(1-i\lambda) - i\lambda/2 + (i-1)\lambda}{1-(i-1)\lambda} - (m-1) \quad (7.15)$$

となる。  $\bar{T}_i(m)$  の数値例を図-14 に示す。この場合、制御装置に近い端末ほど系内時間が短かく有利となるため、結果的に優先権が付き built-in priority とも呼ばれる。

(iii) ポーリング方式:

対称な場合で、端末間の移行時間を  $r$  タイムスロットとすると、

$$\bar{L}_i = \frac{\sigma^2}{2(1-N\mu)} + \frac{Nr\mu(1-\mu)}{2(1-N\mu)} \quad (7.16)$$

$$\bar{T}_i(m) = \frac{N\sigma^2}{2(1-N\mu)} + \frac{Nr(1-\mu)}{2(1-N\mu)} + \frac{1-r}{2} \quad (7.17)$$

同期的多重待ち行列のモデルとしてはこの他、両方向のトラフィックを運ぶ場合<sup>48)</sup>、1 タイムスロット当り複数個のデータユニットを処理する場合<sup>49)</sup>、端末でのバッファモデル<sup>50)</sup>などが解析されている。

### 8. 総合システムにおける待ち行列

情報処理システムを総合的に評価するには、CPU、記憶装置、I/O デバイス、などすべての resource を含めたシステムの分析が必要である。そのような総合



図-15 直列形待ち行列ネットワーク



図-16 循環形待ち行列ネットワーク

システムでは各 resource で待ち行列が形成され、それらが有機的に結合されたモデルとなり、待ち行列ネットワーク (queueing network) と呼ばれる。待ち行列ネットワークは、その形態上から見て以下のように呼ばれている。

(i) 直列形 (tandem queue): 図-15 に示す如く、待ち行列が直列に並んでいて、各段からの退去が次段への到着となっている場合。

(ii) 循環形 (cyclic queue): 図-16 の如く、直列形で最終段のサービスを受けた客が再び最初の待ち行列に到着する場合。

(iii) 帰還形 (feedback queue): ある段でサービスを受けた客の一部が、前段の待ち行列へ帰還して再びサービスを受ける場合。

(iv) 閉鎖形 (closed queue): ネットワークの外から客は到着せず、系内客数が有限の場合。また系外より客の到着がある場合を開放形 (open queue) と呼ぶ。

帰還形の中には閉鎖形と開放形があり、また循環形は常に閉鎖形である。例えば、端末から発生したジョブが、通信制御装置、記憶装置、CPU などで帰還しながら処理を受けて端末へ戻り、利用者の思考時間を経て再びジョブが発生する系を考えると、端末まで含めた閉鎖形、帰還形の待ち行列ネットワークとなる。

#### 8.1 マルコフ形待ち行列ネットワーク

各 resource での待ち行列が M/M/形の場合には、系の平衡方程式が成り立ち、平衡状態での同時確率が求められる。閉鎖形で、系内客数が  $N$ 、段数 (待ち行列数) が  $K$ 、 $i$  段のサービス率および窓口数が  $\mu_i$  と  $r_i$ 、 $i$  段のサービスを終了した客が次に  $j$  段へ移る先天的確率が  $q_{ij}$  のモデルを考えよう<sup>51)</sup>。  $n_i$  を  $i$  段の系内客数として、同時確率  $p(n_1, n_2, \dots, n_K)$  に対し次の平衡方程式が成り立つ。

$$\begin{aligned} & \left\{ \sum_{i=1}^K \varepsilon(n_i) \alpha_i(n_i) \mu_i \right\} p(n_1, \dots, n_K) \\ &= \sum_{i=1}^K \sum_{j=1}^K \varepsilon(n_j) \alpha_j(n_j+1) \mu_j q_{ji} \\ & \quad \times p(n_1, \dots, n_j-1, \dots, n_j+1, \dots, n_K) \quad (8.1) \end{aligned}$$

ここで、 $\epsilon(n_i)=0(n_i=0), =1(n_i \neq 0)$ , および  $\alpha_i(n_i) = n_i(n_i \leq r_i), = r_i(n_i \geq r_i)$  である。式(8.1)の解は次の積形式で与えられる。

$$\begin{cases} p(n_1, \dots, n_K) = \prod_{i=1}^K x_i(n_i) / G(N) \\ G(N) = \sum_{n_1 + \dots + n_K = N} \prod_{i=1}^K x_i(n_i) \end{cases} \quad (8.2)$$

ここで、 $x_i(n_i)$  は次式で定義される。

$$x_i(0) = 1, \quad x_i(k) = x_i(k-1) y_i / \alpha_i(k) \quad (8.3)$$

但し、補助パラメータ  $y_i$  はつぎの  $K$  個の連立方程式の根である。

$$\sum_{i=1}^K y_i q_{ij} = y_j \quad j=1, 2, \dots, K \quad (8.4)$$

このモデルで  $N \rightarrow \infty$  とし、1 段目のサービス時間間隔を到着時間と見れば、 $(K-1)$  段の開放形(無限入力源)となる。したがって、このモデルで表-3 の如く補助パラメータを定めれば、マルコフ形の種々の待ち行列ネットワークに対する結果を得る。

同時確率が求められると、周辺分布  $P_r\{n_i=k\}$ , 平均系内客数  $\bar{n}_i$ ,  $i$  段のスループット  $\bar{\theta}_i$  などが計算される。例えばすべての段で窓口が  $1(r_i=1)$  の場合は、 $x_i(k) = y_i^k$  となり、

$$P_r\{n_i=k\} = \frac{y_i^k}{G(N)} [G(N-k) - y_i G(N-k-1)] \quad (8.5)$$

$$\bar{n}_i = \sum_{k=1}^N y_i^k \frac{G(N-k)}{G(N)} \quad (8.6)$$

$$\bar{\theta}_i \triangleq \sum_{k=1}^N \alpha_i(k) P_r\{n_i=k\} = y_i \frac{G(N-1)}{G(N)} \quad (8.7)$$

で計算される。上式の計算に必要な  $G(1), \dots, G(N)$  については、計算アルゴリズムも提案されている<sup>56), 57)</sup>。

このような閉鎖形・掃選形の待ち行列ネットワークを用いた多重プログラミングシステムの分析例は幾つ

表-3 各種モデルに対する補助パラメータの値

モデル	系内人数 $N$	窓口数 $r_i$	移動確率 $q_{ij}$	補助パラメータ $y_i$
循環形 <sup>52)</sup>	有限	1	$q_{ij} = \begin{cases} 1(j=i+1(\text{mod } K)) \\ 0(j \neq i+1) \end{cases}$	$\mu_1$
直列形 <sup>53)</sup>	無限	$r_i = 1$ $r_i \geq 1$ ( $i > 1$ )	$q_{ij} = \begin{cases} 1(j=i+1) \\ 0(j \neq i+1) \end{cases}$	$\mu_i r_i$
最終段階掃選形 <sup>54)</sup>	任意	1	$q_{ij} = \begin{cases} 1(i \neq K, j=i+1) \\ 0(i \neq K, j \neq i+1) \\ q_{Kj} = q_j > 0 \end{cases}$	$\frac{\mu_1(1-q_K-q_{K-1}-\dots-q_i)}{(1-q_K-q_{K-1}-\dots-q_2)}$
網形 <sup>55)</sup>	無限	$r_i = 1$ $r_i \geq 1$ ( $i > 1$ )	任意	式(8.4)

か報告されており、かなり良い精度で応答時間などを推定できることが確かめられている<sup>4), 58)</sup>。

### 8.2 局所平衡の成り立つ待ち行列ネットワーク<sup>59)</sup>

式(8.1)は式(3.4)と同じ意味の平衡方程式である。

一方、サービス段  $i$  と状態  $S_j$  に対し、

[  $i$  段への到着により  $S_j$  へ遷移する割合 ]

= [  $i$  段からの退去により  $S_j$  から遷移する割合 ] (8.8)

が成り立つ場合を局所平衡(local balance)と呼び、式(8.8)を局所平衡方程式(local balance equation)という。状態  $S_j$  を固定して局所平衡方程式を加えると、通常の平衡方程式(これを全域平衡方程式, global balance equation と呼ぶ)を得る。したがって、局所平衡方程式は全域平衡方程式に対する十分条件となっていて、平衡方程式の解を探すのに便利である。また客にクラスがある場合に、クラスごとの特性を求めるのにも有効である。

局所平衡が成り立つ場合の到着過程としては、ポアソン過程( $\lambda$ )か、クラスごとに独立なポアソン過程( $\lambda_r$ )であり、系内客数に依存するパラメータでも良い。系に入ったクラス  $r$  の客が最初に  $j$  段のサービスを受ける確率を  $b_{jr}$ , クラス  $r$  の客が  $i$  段のサービス終了後クラス  $s$  の客となって  $j$  段へ移行する確率を  $p_{i,r;j,s}$  とする。後者は、以前に受けたサービスの種類によってその後の遷移先が異なるモデルに有効である。

局所平衡が成り立つ場合のサービス窓口としては、次の4タイプのうちどれかである。

(タイプ1) FCFS- $\cdot/M/1$ : この場合、平均サービス率は系内客数の関数でも良い。

(タイプ2) PS- $\cdot/G_r/1$ : ここで  $G_r$  は、LST(ラプラススチェルチェス変換)が有理関数であるサービス時間分布であり、クラスごとに異なっても良い。

(タイプ3)  $\cdot/G_r/\infty$ :

(タイプ4) 割込み継続形 LCFS- $\cdot/G_r/1$ :

ここで一般の  $G_r$  に対しても局所平衡が成り立つのは、有理関数の LST をもつサービス時間分布が、指数分布の窓口から成る副サービス段の結合で表わされるためである。 $G_r$  の中には、アーラン分布、超指数分布なども含まれる。

局所平衡が成り立つ場合には、平衡状態での同時確率は次の積形式で与えられることが証明されている。

$$P_r\{S=(x_1, x_2, \dots, x_K)\}$$

$$=Cd(S)f_1(x_1)f_2(x_2)\dots f_K(x_K) \quad (8.9)$$

ここで、 $x_i$  は  $i$  段の系内客数ベクトル (各クラスについての)、 $S$  は系の状態、 $C$  は正規化定数、 $d(S)$  は  $S$  での系内客数の関数、 $f_i$  は  $i$  段の  $x_i$  による関数である。

簡単のために、開放形で到着率が状態に依らない場合について全クラスをまとめた  $i$  段の系内客数  $n_i$  に関する同時確率を求めると、

$$P_r\{S=(n_1, \dots, n_K)\} = Cd(S)h_1(n_1), \dots, h_K(n_K) \quad (8.10)$$

である。ここで、 $d(S) = \lambda^{n_1+n_2+\dots+n_K}$  で、 $i$  段がタイプ1のとき  $h_i(n_i) = (\sum_r e_{ir})^{n_i} (1/\mu_i)^{n_i}$ 、タイプ2または4のとき  $h_i(n_i) = (\sum_r e_{ir}/\mu_{ir})^{n_i}$ 、タイプ3のとき  $h_i(n_i) = (1/n_i!) (\sum_r e_{ir}/\mu_{ir})^{n_i}$  であり、 $1/\mu_{ir}$  は  $i$  段のクラス  $r$  に対する平均サービス時間、 $e_{ir}$  は次の連立方程式の根である。

$$\sum_{i,r} e_{ir} p_{ir; j, s} + b_{j, s} = e_{j, s} \quad (8.11)$$

すなわち、 $e_{ir}$  は帰還も含めて  $i$  段へ到着する客の相対的な到着率となっている。これから周辺分布を求めると表-4 のようになる。

### 8.3 その他の待ち行列ネットワーク

非マルコフ形の待ち行列ネットワークの解析は、一般に困難なため近似法が用いられる。サービス時間に関しては、アーラン分布か超指数分布に当てはめれば局所平衡が成り立つ場合となるが、到着がポアソン過程でないとか、待ち合わせ室が有限のために閉塞現象 (blocking) が起こる場合に対しては局所平衡が成り立たない。

マルコフ形の変形も含めて、待ち行列ネットワークのその他の解析には、拡散近似を用いる方法<sup>60)</sup>、電気回路論の Norton の定理に似た性質を用いる簡易法<sup>61)</sup>、閉塞のある有限待ち行列ネットワークの解析<sup>62), 63)</sup>、直接数値的に平衡方程式を解く方法<sup>64)</sup>、などがある。

## 9. コンピュータ通信における待ち行列

近年、複数のコンピュータ間を通信回路で結んで、コンピュータ間で処理の分散を行うコンピュータネットワークの研究・実用化が盛んとなっている。これまでの情報処理システムと異なる点は、センター中心ではなく各コンピュータがそれぞれ独自性をもって処理を行うことである。この場合、コンピュータ間を直接接続せずに、データの交換・伝送を行う専用の通信網を介して接続した方が、標準化・仮想化などの点で望

表-4 開放形待ち行列ネットワークの周辺分布

$i$ 段のタイプ	$P_r(n_i=k)$	$\rho_i$
1	$(1-\rho_i)\rho_i^k$	$\lambda \sum_r e_{ir}/\mu_i$
2		$\lambda \sum_r e_{ir}/\mu_{ir}$
4		
3	$\frac{\rho_i^k}{k!} e^{-\rho_i}$	

表-5 交換方式の比較

交換方式	回線交換	メッセージ交換	パケット交換
比較尺度	回線交換	メッセージ交換	パケット交換
サービス品質	損失率 回線設定時間	遅延時間	遅延時間
遅延特性	送出時間	短メッセージに不利	短メッセージに有利
所要バッファ量	0	大	中
回線効率	短メッセージに不利	大	中

ましい。この通信網はサブネットワークと呼ばれ、主としてデータの交換を行うノードと伝送を行う通信回線より成る。このサブネットワークでは、前章までと異なった以下のような待ち行列問題が提起されている。

#### (1) ノードにおける交換方式の選定

データを交換する方法としては、回線交換・メッセージ交換・パケット交換に大別される。これらの定性的比較は表-5 の通りであるが、定量的比較には待ち行列理論が用いられる。しかし方式の差が大きいため比較尺度を同じにすることが困難で、充分な評価がされているとは言い難い<sup>65)-68)</sup>。

#### (2) サブネットワークの統合・分離

情報処理システムごとに、通信されるデータの特性、トラヒック特性、要求条件などが異なり、それらと同じサブネットワークで統合するか、別々にするかという問題である。例えば、2つのシステムにおいて通信するデータ長が互いに非常に異なっている場合、通信回線を共用するかどうかという問題をモデル化すると、図-17 の如き待ち行列系の比較となる<sup>69)</sup>。このモデルから、遅延時間やメモリ量などの評価尺度が比較される。

#### (3) データの遅延特性

メッセージ交換またはパケット交換では、各ノードで待ち行列が形成され、待ち行列ネットワークとな

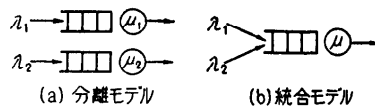


図-17 分離・統合の待ち行列モデル

る。データ(メッセージまたはパケット)が各ノードを中継されて着ノードからコンピュータへ送出されるまでがサブネットワークの遅延時間である。遅延時間はデータの経由するルートに依存し、複雑な要因より成るため解析が困難である。そのため、各ノードでは同じデータでも新たにデータ長分布の中から選んだデータ長をもつと仮定すれば、ノード間の影響を少なくできる。これを独立仮設(independence assumption)と呼んでいる<sup>70)</sup>。独立仮設のもとで、ノードへのデータの到着がポアソン過程で、 $K$  段のリンクを経てコンピュータへ送出されるまでの平均遅延時間  $\bar{T}$  は、Pollaczek-Khinchin の公式を用いて次式で近似される<sup>71)</sup>。

$$\bar{T} = \sum_{i=1}^K \frac{\lambda_i}{\gamma} \left[ \frac{\lambda_i \bar{l}^2 / C_i}{2(1 - \lambda_i \bar{l} / C_i)} + \frac{l}{C_i} + P_i + \tau_c \right] + \tau_c \quad (9.1)$$

ここで、 $C_i$  はリンク  $i$  の回線速度、 $\bar{l}$  と  $\bar{l}^2$  は信号データも含めた全データに対するデータ長の1次・2次モーメント、 $l$  は注目しているデータの長さ、 $P_i$  は伝播遅延、 $\tau_c$  はノードでの処理遅延、 $\lambda_i$  は回線  $i$  へ加わる単位時間当りのデータ数、 $\gamma$  はサブネットワークへ加わる単位時間当りの全データ数である。この式を用いて、回線速度の最適割付け、ルーティングの最適化、などのマクロ的検討がなされている<sup>72), 73)</sup>。

#### (4) スループット特性

スループットとは、単位時間当り通信されたデータ数であり、サブネットワークの総合特性を評価する尺度の1つである。通常の待ち行列系では、スループット対入力(加わるデータ数/単位時間)特性は単調増大で飽和する特性を示すが、サブネットワークでは閉塞などの現象によりスループットが下る場合がある。

例えば、無線回線を用いたパケット網である ALOHA 網では、パケットが衝突することによりスループットが減少する<sup>74)</sup>。送出時間の一定長パケットがポアソン過程(パラメータ  $\lambda$ ) で発生するとき、図-18 より衝突確率は  $1 - e^{-2\lambda}$  である。但し  $\lambda$  は  $\tau$  時間当りの総発信パケット数である。したがって、 $\lambda$  を単位時間としたスループット  $\theta$  は  $\theta = \lambda e^{-2\lambda}$  となり、 $\lambda = 1/2$  で  $\max \theta = 1/2e$ 、 $\lambda > 1/2$  で  $\theta$  は減少する。

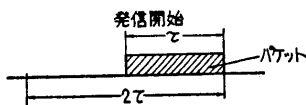


図-18 無線パケットの衝突区間 ( $2\tau$  間に他パケットが発信すれば衝突する)

#### (5) パケット長の決定

同一メッセージ分布に対し、最大パケット長を短かくすると、同一メッセージが多くのパケットに分かれて各々ヘッダーなどの overhead 分が増えると同時にノードでの処理個数が増す。一方長くすると、バッファ使用効率が悪くなり、また遅延時間も増す。したがって、最適な値が存在するがメッセージ長分布に大きく依存する<sup>75), 76)</sup>。

これらの他、(6)ルーティング(特にアダプティブな)、(7)フローコントロール、などの問題があり、待ち行列理論などによる解明が行われている。

## 10. あとがき

情報処理システムにおいてどのように待ち行列理論が使用されているか、典型的なモデルを中心に紹介した。各種システム、resource に対応して非常に多くのモデルが提起され解析されているが、基本的な考え方や解析の手法は限られている。したがって、その手法を習得することにより、システム対応に各種の定量的評価が可能となる。今後待ち行列理論は、もう一つの分析手法であるシミュレーションとの結合、待ち行列ネットワーク解析のプログラム化などにより、使いやすかつ精度が良い方向に向い、コンピュータネットワークの発展とあいまって、ますます情報処理システムの分析・評価に使われると予想される。

## 文献紹介

- 32) 橋田, 川島: バッファメモリのトラフィック設計法の比較, 電子通信学会交換研究会, SE 74-76 (1975).
- 33) F. Closs: Packet Arrival and Buffer Statistics in a Packet Switching Node, 3rd Data Communications Symp., Florida, pp. 12~17 (1973).
- 34) W. W. Chu: Buffer Behavior for Poisson Arrivals and Multiple Synchronous Constant Output, IEEE Trans., Vol. C-19, No. 6, pp. 530~534 (1970).
- 35) W. W. Chu: Buffer Behavior for Batch Poisson Arrivals and Single Constant Output, IEEE Trans., Vol. COM-18, No. 5, pp. 613~618 (1970).
- 36) G. W. Fredrikson: Buffer Behavior for Binomial Input and Constant Service, IEEE Trans., Vol. COM-22, No. 11, pp. 1862~1866 (1974).
- 37) O. Hashida: On the Busy Period in the Queueing System with Finite Capacity, J. of Operations Res. Society of Japan, Vol. 15, No. 3, pp. 115~137 (1972).

- 38) J. H. Chang : An Analysis of Buffering Techniques in Teleprocessing Systems, IEEE Trans., Vol. COM-20, No. 3, pp. 619~630 (1972).
- 39) J. Hsu : A General Queueing Model for Buffer Storage Problems, IEEE Trans., Vol. COM-21, No. 6, pp. 744~747 (1973).
- 40) J. W. Cohen : On Memory Units with Finite Capacity, 6th Int'l Teletraffic Congress, München, No. 321 (1970).
- 41) O. Hashida and M. Fujiki : Queueing Models for Buffer Memory in Store-and-Forward Systems, 7th Int'l Teletraffic Congress, Stockholm, No. 323 (1973).
- 42) 伊藤正彦 : データ通信方式におけるチャネルバッファ容量の解析, 研究実用化報告, Vol. 20, No. 1, pp. 215~229 (1971).
- 43) 橋田, 中村 : 複数待ち行列の解析, 経営科学, Vol. 13, No. 1, pp. 30~47 (1969).
- 44) 橋田温 : ゲート式多重待ち行列, 電子通信学会論文誌, Vol. 53-A, No. 1, pp. 43~50 (1970).
- 45) 橋田, 川島 : 端末呼源 1 のポーリング方式の解析, 電子通信学会交換研究会, SE 74-5 (1974).
- 46) M. A. Leibowitz : An Approximate Method for Treating a Class of Multiqueue Problems, IBM J. of Res. and Dev., Vol. 5, pp. 204~209 (1961).
- 47) W. W. Chu and A. G. Konheim : On the Analysis and Modeling of a Class of Computer Communication Systems, IEEE Trans., Vol. COM-20, No. 3, pp. 645~660 (1972).
- 48) A. G. Konheim and B. Meister : Distributions of Queue Lengths and Waiting Times in a Loop with Two-Way Traffic, J. of Computer and System Sci., Vol. 7, pp. 506~521 (1973).
- 49) A. G. Konheim and B. Meister : Service in a Loop System, J. of ACM, Vol. 19, No. 1, pp. 92~108 (1972).
- 50) P. N. Robillard : An Analysis of a Loop Switching system with Multirank Buffers Based on the Markov Process, IEEE Trans., Vol. COM-22 No. 11, pp. 1772~1778 (1974).
- 51) W. J. Gordon and G. F. Newell : Closed Queueing Systems with Exponential Servers, Operations Res., Vol. 15, No. 2, pp. 254~265 (1967).
- 52) E. Koenigsberg : Cyclic Queue, Operations Res. Quart., Vol. 9, pp. 22~35 (1958).
- 53) R. R. P. Jackson : Random Queueing Processes with Phase Type Service, J. of Royal Statist. Soc., B, Vol. 18, pp. 129~132 (1956).
- 54) P. D. Finch : Cyclic Queues with Feedback, J. of Royal Statist. Soc., B, Vol. 21, pp. 153~157 (1959).
- 55) J. R. Jackson : Networks of Waiting Lines, Operations Res., Vol. 5, pp. 518~521 (1957).
- 56) J. P. Buzen : Computational Algorithms for Closed Queueing Networks with Exponential Servers, Comm. of ACM, Vol. 16, No. 9, pp. 527~531 (1973).
- 57) M. Reiser and H. Kobayashi : Honer's Rule for the Evaluation of General Closed Queueing Networks, Comm. of ACM, Vol. 18, 10, No. 10, pp. 592~593 (1975).
- 58) 田中穂積 : 並列循環待ち行列を用いたオンラインシステムの解析, 電子通信学会論文誌, Vol. 53-C, No. 10, pp. 758~764 (1970).
- 59) 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, pp. 248~260 (1975).
- 60) H. Kobayashi : Application of the Diffusion Approximation to Queueing Networks I : Equilibrium Queue Distribution, J. of ACM, Vol. 21, No. 2, pp. 316~328 (1974).
- 61) K. M. Chandy, U. Herzog and L. Woo : Parametric Analysis of Queueing Networks, IBM J. of Res. and Dev., Vol. 19, No. 1, pp. 36~42 (1975).
- 62) F. S. Hillier and R. W. Boling : Finite Queue in Series with Exponential or Erlang Service Times-A Numerical Approach, Operations Res., Vol. 14, No. 2, pp. 286~308 (1967).
- 63) S. S. Lam : Store-and-Forward Buffer Requirements in a Packet Switching Network, IEEE Trans., Vol. COM-24, No. 4, pp. 394~403 (1976).
- 64) V. L. Wallace and R. S. Rosenberg : Markovian Models and Numerical Analysis of Computer System Behavior, Proc. of SJCC, pp. 141~148 (1966).
- 65) F. Closs : Time Delays and Trunk Capacity Requirements in Line-Switched and Message-Switched Networks, Int'l Switching Symp. Record, pp. 428~433 (1972).
- 66) H. Rudin : Buffered Packet-Switching : A Queue with Clustered Arrivals, Int'l Switching Symp. Record, pp. 259~265 (1972).
- 67) K. Itoh, T. Kato, O. Hashida and Y. Yoshida : An Analysis of Traffic Handling Capacity of Packet Switched and Circuit Switched Networks, 3rd Data Communications Symp., Florida, pp. 29~37 (1973).
- 68) 宮原, 長谷川 : 計算機網における伝送交換方式の比較評価, 電子通信学会論文誌, Vol. 58-A, No. 7, pp. 411~418 (1975).
- 69) D. Maiwald, H. R. Mueller, H. R. Rudin and



- C. H. West : Integrated Communication System Performance, Int'l Conference on Communications, Seattle, pp. 24. 13~24. 21 (1973).
- 70) L. Kleinrock : Communication Nets, McGraw-Hill, New York (1964).
- 71) L. Kleinrock : Analytic and Simulation Methods in Computer Network Design, AFIPS Conference Proceedings, Vol. 36, SJCC, Atlantic City, pp. 569~579 (1970).
- 72) D. G. Cantor and M. Gerla : Capacity Allocation in Distributed Computer Networks, 7th HAWAII Int'l Conf. on Syst. Sci., pp. 115~117 (1974).
- 73) M. Gerla : Approximations and Bounds for the Topological Design of Distributed Computer Networks, 4th Data Communication Symp., Quebec, pp. 4. 9~4. 15 (1975).
- 74) N. Abramson and F. Kuo : Computer-Communication Networks, Prentice-Hall, New Jersey (1973).
- 75) J. C. Majithia and B. Bhar : Optimum Packet Size in a Packet-Switched Data Network, Electronic Letters, Vol. 10, No. 10, 16th (May 1974).
- 76) 宮原, 長谷川 : パケット交換方式における最大パケット長, 電子通信学会論文誌, Vol. 58-A, No. 10, pp. 670~671 (1975).
- 77) G. L. Fultz and L. Kleinrock : Adaptive Routing Techniques for Store-and-Forward Communication Networks, Proc. of Int'l Conf. on Communications, pp. 39. 1-39. 8 (1971).
- 79) R.E. Kahn and W.R. Crowther : Flow Control in a Resource Sharing Computer Network, 2nd IEEE Symp. on Problems in the Optimization of Data Communications Systems, Palo Alto, pp. 108~116 (1971).

(昭和51年10月30日受付)

(昭和51年12月8日再受付)