関西大学理工学会誌
理 工 学 と 技 術
V o l . 21 (2 0 1 4)

解　説

待ち行列理論による情報通信ネットワーク解析

平　田　孝　志*

Analysis of communication networks with queueing theory

Kouji HIRATA

1．はじめに

本稿では、待ち行列理論について解説する。待ち行列理論は、有限の資源が複数の利用者によって共有されている状況を表現した確率モデルである待ち行列モデルを扱うもので、待ち行列モデルの身近な例には、スーパーのレジやATM等にならぶ客の列がある。また、電話回線網の動作や、インターネットにおけるルータでのパケットの振る舞いも待ち行列によってモデル化でき、情報通信ネットワークの解析や設計に欠かすことのできない理論である。本稿では、まず情報通信ネットワークの仕組みについて述べた後、待ち行列理論について解説を行う。その後、待ち行列理論を用いた情報通信ネットワークの解析について説明する。

2．情報通信ネットワーク

本稿では、情報通信ネットワークの中でも、インターネットで使用されるルータの振る舞いに着目する。インターネットは様々なネットワークをルータで相互接続したものであり（図１）、パケット交換を用いてコンピュータ間でデータの伝送を行っている。パケット交換とは、データをパケットという単位で細かく分割し、それを途中のルータにリレーしてもらいながら宛先に届けるデータ通信モデルである。中継のルータは、パケットに付加された経路情報等を保持するヘッダを読み取り、宛先へ向かう出力回線へパケットを出力する。この際、一つの出力回線へは同時に一つのパケットしか送出できないため、複数の入力ポートから複数

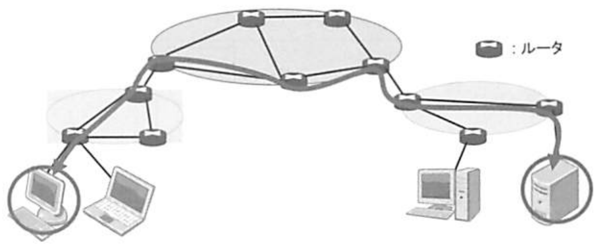

図１　インターネット

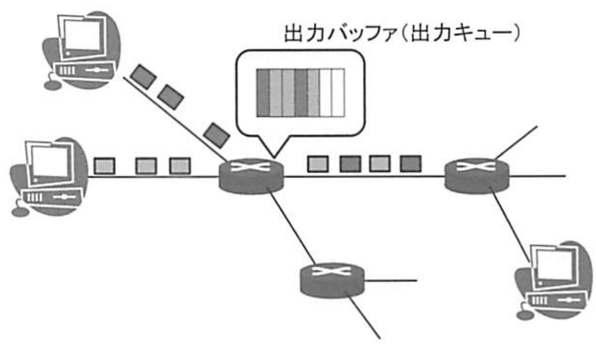

図２　ルータにおけるパケットの待ち

のパケットが到着すると、パケットの待ちが生じる（図２）。これらのパケットはバッファと呼ばれるスペースへ蓄積され、順番に処理され回線へ送出される。出力バッファの容量は有限であるため、容量以上のパケットが到着するとそれらは棄却されてしまう。このようなルータの振る舞いは待ち行列モデルによりモデル化が可能であり、ネットワークを設計する際に非常に重要なものとなる。

3．待ち行列理論

3.1　待ち行列モデル

図３に待ち行列モデルの概念図を示す。図に示され

原稿受付　平成26年９月９日
*システム理工学部 電気電子情報工学科 准教授

ているように、待ち行列モデルは、共有の資源であるサーバ（処理施設）と待合室からなるシステムに外部から客が到着し、これらの客はサービスを受けるためにシステム内で暫く滞在した後、システムを去るというモデルである。なお、システム内には一つ以上のサーバが存在し、一つのサーバは同時に一人の客に対してのみサービスを行い、客がシステム内に存在する限り常にサービスを行うものとする。また、客が到着した時点で全てのサーバが他の客ですでに埋まっている場合は、到着客はサーバが空くまで待合室で待つこととする。スーパーのレジにおいては、サーバはレジにおける会計の役割を果たし、待合室はレジに並ぶ列としてモデル化できる。また、情報通信ネットワークにおいては、客をパケットに対応させ、サーバと待合室はそれぞれ回線とルータ内のバッファに対応させることでモデル化が可能である。

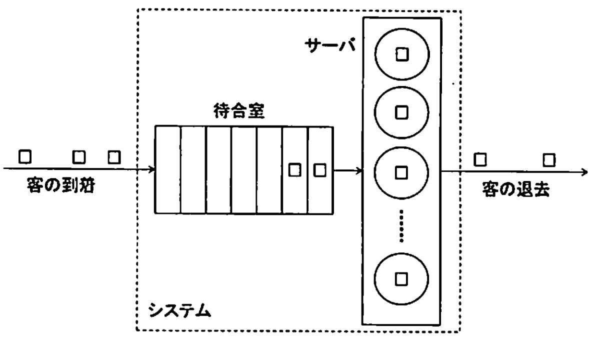

図3　待ち行列モデル

一般に待ち行列モデルは次の五つの要素からなる[1]。

1. 到着過程：客の到着時点に関する統計的情報
2. サービス時間分布：サービスにかかる時間に関する統計的情報
3. サーバ数：同時に処理できる客数の最大値
4. システム容量：システム内に滞在できる客の最大数（サーバ数と待合室容量の和）
5. サービス規律：システム内の客の処理順序を定める規則

これらの要素を定めることによって、どのような待ち行列かを決定することができる。また、対象とする待ち行列がどのような要素で構成されるかを表す方法として、ケンドールの記法が使用される。これは、通常 $A/B/c/N$ の形をしており、それぞれ、到着間隔分布 / サービス時間分布 / サーバ数 / システム容量を表す。A、B に関しては M（指数分布）、D（一定分布）、E_k（k 次のアーラン分布）、G（一般分布：特定の分布を仮定しない）などが用いられる。また、c や N には数字が入り、システム容量が無限大の場合は単に $A/B/c$ と書かれる。通常、サービス規律は先着順サービス（利用要求を到着順に処理する）が仮定されており、FCFS(First-come, First-served)と書かれる。サービス規律を明示する場合は、FCFS $A/B/c$ のように書くことが多いが、先着順の場合は基本的には書かない。

3.2　ポアソン過程と指数分布

まず、待ち行列モデル全体を考える前に、客の到着過程のみに着目する。到着過程で最も基本的なものはランダムな到着である。いま、時間区間 $(0,T]$ に K 人の客がでたらめに到着すると仮定する。すなわち、図4のように、それぞれの客は、他の客とは独立に時間区間 $(0,T]$ 内で一様分布に従って到着時点を選ぶ。この条件において、時間間隔 t の間に客が k 人到着する確率 $P_k(t)$ は

$$P_k(t) = \frac{K!}{k!(K-k)!}\left(\frac{t}{T}\right)^k\left(1-\frac{t}{T}\right)^{K-k}$$

となる。ここで、λ を単位時間あたりに到着する平均客数（平均到着率）とすると、$\lambda = K/T$ であり、$\lambda = K/T$ を一定の値に保ちながら、$T \to \infty$、$K \to \infty$ とすると、

$$P_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \qquad (k = 0, 1, \ldots) \tag{1}$$

となる。これは、ランダムな到着の場合において、時間間隔 t の間に客が k 人到着する確率を表している。この確率は、対象とする区間の位置には無関係であり、時間間隔 t と平均到着率 λ によって決定される。式 (1) の右辺で与えられる確率分布を、パラメータ λt のポアソン分布という。また、パラーメタ λt のポアソン分布の平均及び分散はどちらも λt である。

$N(t)$ を時刻 t までの客の総到着数としたときに、到着過程 $\{N(t), t \geq 0\}$ が次の二つの条件を満たすとき、$\{N(t), t \geq 0\}$ は率 λ のポアソン過程という。

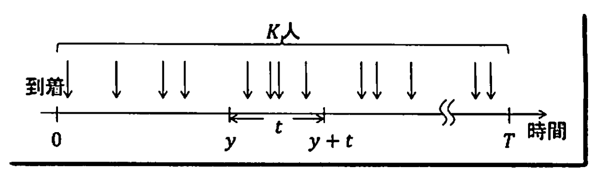

図4　ランダム到着

1. 任意の時間間隔 t の間に到着する客数がパラメータ λt のポアソン分布に従う（その確率は式 (1) で表される）。

2. 重なり合わない区間に到着する客数は互いに独立である。

ポアソン過程はランダムな到着をあらわす到着過程であり、待ち行列モデルに非常によく使用される。

次に、率λのポアソン過程に従い到着する客の到着間隔を考える。到着間隔とはある客が到着してから次の客が到着するまでの時間間隔である。いま、ある客が到着した時点を時刻0とする。この時点から、次の客が到着するまでの間隔Xがt以下である確率$P(X \le t)$は

$$P(X \le t) = 1 - P(X > t) \tag{2}$$

で表される。$P(X>t)$は、時間t以内に客の到着がないということであるので、その確率は式(1)において$k = 0$とすることで求まる。よって、

$$P(X > t) = P_0(t) = e^{-\lambda t}$$

である。これを式(2)に代入すると、

$$P(X \le t) = 1 - e^{-\lambda t} \tag{3}$$

が得られる。式(3)の右辺で表される確率分布をパラメータλをもつ指数分布という。なお、その平均は$1/\lambda$であり、分散は$1/\lambda^2$である。これらより、客の到着が率λのポアソン過程に従う場合（つまりランダム到着の場合)、客の到着時間間隔Xは独立で同一のパラメータλを持った指数分布に従うことがいえる。

指数分布には、無記憶性という重要な性質がある。いま、時刻0に客が到着したと仮定し、次の客の到着までの間隔をXとする。このとき、$(0, t_0]$の間、次の客が到着しなかったという条件の下で、時刻t_0+tまでに次の客が到着する条件付き確率$P(X \le t_0+t \mid X> t_0)$を考えると、

$$P(X \le t_0 + t \mid X > t_0) = P(X \le t) = 1 - e^{-\lambda t}$$

となる。これは、条件付き確率$P(X \le t_0+t \mid X> t_0)$が$t_0$とは無関係で、時刻$t_0$から次の到着までの間隔は、元の到着間隔$X$と同じ確率分布に従うことを意味している。つまり、将来の到着が起こるまでの時間に関する確率は、最後の到着が起こってからどのくらい時間が経過したかには無関係である。この性質を指数分布の無記憶性とよび、待ち行列モデルの解析において非常に重要な役割を果たす。

3.3 出生死滅過程

これまで、待ち行列システムに到着する客の到着過程を考えてきた。ここでは、到着に加えて、サービスを終えて客が退去するようなモデルについて考える。そのようなモデルを記述するために、人口の増減を表すモデルである出生死滅過程と呼ばれる確率過程について議論する。

時刻tにおける系内客数$L(t)$（これを状態という）がk人のときに、次の客が到着するまでの間隔がパラメータλ_k(つまり到着率)の指数分布に従い、客のサービスにかかる時間がパラメータμ_k（サービス率という）の指数分布に従う場合を考える。このとき、微小時間Δtの間に客が一人到着し、客数が一人増加する確率は、指数分布の無記憶性および式(3)から

$$\begin{aligned} &P(L(t+\Delta t) = k+1 \mid L(t) = k) \\ &\quad = P(X \le \Delta t) = 1 - e^{-\lambda_k \Delta t} \\ &\quad = 1 - \left[1 - \lambda_k \Delta t + \frac{(\lambda_k \Delta t)^2}{2} - \cdots\right] \\ &\quad = \lambda_k \Delta t + o(\Delta t) \end{aligned}$$

と表せる。なお、$o(\Delta t)$は$\lim_{\Delta t \to 0} o(\Delta t)/\Delta t = 0$となる項、すなわち$\Delta t$の高次の項を表す。同様に、微小時間$\Delta t$の間にサービスを終え客が一人退去する、つまり客数が一人減る確率は、

$$P(L(t+\Delta t) = k-1 \mid L(t) = k) = \mu_k \Delta t + o(\Delta t)$$

となる。また、微小時間の間に客が二人以上増減する確率は十分に小さいとし、無視できるものとする。客が増減する確率が上記のように与えられるとき、$L(t)$は出生死滅過程と呼ばれる。図5に出生死滅過程の状態遷移速度図（状態遷移の様子を図で表したもの）を示す。図において、円で囲まれている数字は状態を表し、ここでいう状態とは系内客数のことである。

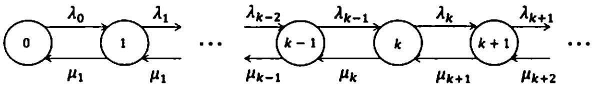

図5 出生死滅過程の状態遷移速度図

いま、システムが定常である状態において系内客数がk人である確率p_kを求めることを考える。システムが定常であるとは、時間が十分に長く経過した時に、$P(L(t)=k)$が過渡的な振る舞いをせずに一定値p_kに落ちつく状況であることをいう。この場合、ある状態に入る確率と、その状態から出ていく確率が等しくなる。つまり、状態がkである場合、

$$(\lambda_k \Delta t + \mu_k \Delta t) p_k = \lambda_{k-1} \Delta t p_{k-1} + \mu_{k+1} \Delta t p_{k+1}$$

となる（図6）。これより、$k = 0$のとき

$$\lambda_0 p_0 = \mu_1 p_1 \tag{4}$$

となり、$k \geq 1$ のとき

$$(\lambda_k + \mu_k)p_k = \lambda_{k-1}p_{k-1} + \mu_{k+1}p_{k+1} \tag{5}$$

となる。式(4)、(5)を（状態 k に対する）平衡状態方程式という。

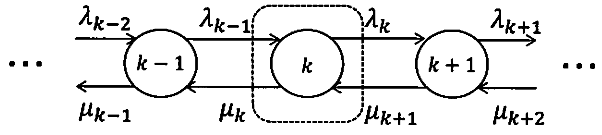

図6　ある状態に対する確率フローの総和

ここで、式(4)及び(5)を $k = 0, 1, 2, \ldots$ と順々に足しあわせると

$$\lambda_{k-1}p_{k-1} = \mu_k p_k \tag{6}$$

を得る。式(6)より

$$p_k = \frac{\lambda_{k-1}}{\mu_k}p_{k-1} = \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\mu_k\mu_{k-1}\cdots\mu_1}p_0, \quad k = 1, 2, \ldots \tag{7}$$

と計算できる。ここで、確率の和が 1 ということを利用すると

$$\sum_{k=0}^{\infty} p_k = p_0 + \sum_{k=1}^{\infty} \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\mu_k\mu_{k-1}\cdots\mu_1}p_0 = 1$$

となり、未知の確率 p_0 は

$$p_0 = \left[1 + \sum_{k=1}^{\infty} \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\mu_k\mu_{k-1}\cdots\mu_1}\right]^{-1} \tag{8}$$

で与えられる。定常状態が存在するための条件は式(8)の右辺に現れる無限和が有限に収束することである。以上のように、出生死滅過程の定常状態確率は式(7)及び(8)によって表される。

4．様々な待ち行列モデル

出生死滅過程は客の到着と退去が発生する場合の状態（ここでは系内客数 $L(t)$）を表すものであり、これはまさに待ち行列モデルの振る舞いそのものである。以下では、この出生死滅過程をもとにいくつかの待ち行列モデルの例をあげる。なお、サーバは系内に客がいる限り、常にサービスを行うものとする。

4.1　M/M/1

まずはじめに、最も単純であるが重要である *M/M/*1 待ち行列について考える。*M/M/*1 とは次のような待ち行列モデルである。

1．到着過程が率 λ のポアソン過程に従う（つまり、到着時間間隔が独立で同一なパラメータ λ を持つ指数分布に従う）。

2．サービス時間が独立で同一なパラメータ μ の指数分布に従う（サービスは客がいるときのみであるから、ポアソン過程のように次から次へとサービスをするとは限らない）。

3．到着（到着過程）と退去（サービス時間）は互いに独立である。

4．サーバは一つであり、客が到着したときサービス中の客がいれば、到着した客は待合室に入る。

5．無限のシステム容量（サーバ数と待合室の容量の和）を持つ。

なお、この場合、到着率及びサービス率はそれぞれ λ 及び μ であり、平均到着間隔及び平均サービス時間はそれぞれ $1/\lambda$ 及び $1/\mu$ となる。

*M/M/*1 の定常状態確率は、式(7)と(8)において $\lambda_k = \lambda$ 及び $\mu_k = \mu$ とすると求まる。つまり、

$$p_k = \left(\frac{\lambda}{\mu}\right)^k p_0, \quad k = 1, 2, \ldots \tag{9}$$

及び、

$$p_0 = \left[1 + \sum_{k=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^k\right]^{-1}$$

となる。定常状態確率が存在するためには上式の無限和が収束する必要があるため、$\lambda/\mu < 1$ である必要がある。その場合、無限等比級数の和を計算すると

$$p_0 = \left[1 + \frac{\lambda/\mu}{1 - \lambda/\mu}\right]^{-1} = 1 - \frac{\lambda}{\mu}$$

が得られる。これを式(9)に代入すると、定常状態確率は

$$p_k = \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right), \quad k = 0, 1, 2, \ldots$$

となることがわかる。

4.2　M/M/c/c

次に、待合室を持たず、サーバが c 個の待ち行列

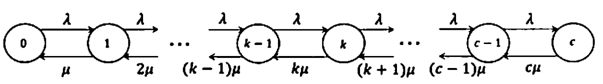

図7　$M/M/c/c$ の状態遷移速度図

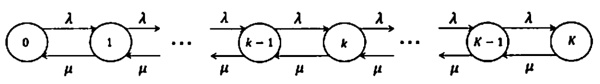

図8　$M/M/1/K$ の状態遷移速度図

$M/M/c/c$ を考える。この場合、c 人までは同時にサービス可能であるが、客が到着した時点でc個全てのサーバが稼働中である場合、システムに入らず退去してしまう（これを呼損という）。このように、待合室を持たない有限システム容量の待ち行列を即時式システムという。このシステムは、図7のように状態が c までしか無いと考え、出生死滅過程において $\lambda_k=\lambda\,(k=0,1,\ldots,c-1)$、$\mu_k=k\mu\,(k=1,2,\ldots,c)$ と設定することでモデル化できる。式(7)及び(8)に上記を代入すると

$$p_k=\frac{(\lambda/\mu)^k}{k!}p_0, \qquad k=1,2,\ldots,c$$

及び

$$p_0=\left[\sum_{k=0}^{c}\frac{(\lambda/\mu)^k}{k!}\right]^{-1}$$

となり、常に安定で定常状態が存在する。よって、系内客数の定常状態確率 p_k は

$$p_k=\frac{(\lambda/\mu)^k/k!}{\sum_{i=0}^{c}(\lambda/\mu)^i/i!}, \qquad k=0,1,\ldots,c \tag{10}$$

で与えられる。

ここで、全てのサーバが稼働中であるために客がシステムに入らず退去するという確率を表す呼損率について考える。全到着客数を A 人、退去した客数を B 人としたとき呼損率 P は $P=B/A$ で与えられる（事象平均）。一方、上記で求めてきた p_k は長い時間の中で過程が状態 k にいる割合を表す（時間平均）。一般に事象平均と時間平均は一致しないが、客の到着がポアソン過程に従う場合、これらは等しくなる。よって呼損率は p_c で与えられ、式(10)において、$k=c$ とすると

$$p_c=\frac{(\lambda/\mu)^c/c!}{\sum_{i=0}^{c}(\lambda/\mu)^i/i!}$$

と得ることができる。上式はアーラン呼損式とよばれ、電話網の設計に利用される。

4.3　M/M/1/K

最後に、システムが収容できる客の合計が最大値 K 人（サービス中の客も含む）である単一サーバ待ち行列 $M/M/1/K$ を考える。このモデルでは客が到着した時点で系内客数が K 人であれば（サービス中の客が1人と待合室にいる客が $K-1$ 人)、その到着客はシステムに入ることができず退去する。このシステムは系内客数が $K+1$ 以上となることはなく、K までしか存在しないと考えて、出生死滅過程において $\lambda_k=\lambda\,(k=0,1,\ldots,K-1)$、$\mu_k=k\mu\,(k=1,2,\ldots,K)$ と設定することでモデル化できる。図8にその状態遷移速度図を示す。これを式(7)及び(8)に適用すれば定常状態確率 p_k は

$$p_k=\left(\frac{\lambda}{\mu}\right)^k p_0=\rho^k p_0, \quad k=1,2,\ldots K$$

となり、p_0は

$$p_0=\begin{cases}\dfrac{1-\rho}{1-\rho^{K+1}}, & \rho\neq 1\\[2mm] \dfrac{1}{K+1}, & \rho=1\end{cases}$$

で表される。ただし、$\rho=\lambda/\mu$である。これらをまとめると、

$$p_k=\begin{cases}\dfrac{(1-\rho)\rho^k}{1-\rho^{K+1}}, & \rho\neq 1\\[2mm] \dfrac{1}{K+1}, & \rho=1\end{cases}, \quad k=0,1,\ldots,K \tag{11}$$

となる。このモデルでは常に定常状態が存在する。

$M/M/1/K$ においても、$M/M/c/c$ と同様に、到着した客がシステムに入れず呼損してしまう確率である呼損率を求めることができる。呼損はシステム内に K 人いるときに新たな客が到着した場合発生するので、$M/M/1/K$ において、呼損率 P は客が K 人いた時間の割合、つまり p_K で与えられる。$M/M/1/K$ において、客をパケット、システムを回線及びルータ内のバッファとして考えれば、情報通信ネットワークにおけるルータの振る舞いを表すことができる。呼損率はパケットの棄却率であり、このようなシステムを考えることで、どの程度の処理速度があれば、どの程度のパケットをほぼ棄却なしで扱えるのかといった設計が可能となる。

5．まとめ

本稿では、待ち行列理論についての解説を行った。待ち行列理論は、情報通信ネットワークの設計や解析に欠かせない理論であり、ルータにおけるパケットの

振る舞いを表現できる。本稿では、待ち行列理論を勉強するための第一歩として、初歩的な待ち行列システムのみを解説したが、今後機会があれば、より詳しく解説を行いたい。

参考文献

[1] 滝根，伊藤，西尾，ネットワーク設計理論，岩波書店，2001.