Proceedings of the 8th WSEAS International Conference on APPLIED MATHEMATICS, Tenerife, Spain, December 16-18, 2005 (pp49-54)

Conditional busy and free queue systems with variable servers' number.

A. SURKOV
Sankt-Petersburg University
7/9, University emb., Sankt - Petersburg
RUSSIA

Abstract: - This article is an attempt to investigate stationary queuing systems with variable servers' number. Servers' quantity here is determined with stochastic variable. Some formulae are received for stationary state probabilities, carried load, queue mean and variance for conditional busy and free (without and with idle servers respectively) M/M/v system. The results are compared with well known ones for M/M/s systems using deterministic probability as server distribution.

Key-Words: Queue systems, variable servers' number, M/M/v system.

1 Introduction

Modern distributed systems middleware offers resource hiring from others cites which may belong to other organizations. For example, this is a good practice in GRID solutions [1], where one can build virtual organization using borrowed resources. Well known similar business technology is outsourcing. Computational resources in such environments can be offered or revoked at arbitrary time what makes these systems mutable. At another point of view, sometimes system designer is forced to apply dynamical hiring policy of some kind as a result of nonzero cost of borrowed resources in general case.

Another approach may be useful to systems if its service or reaction time is a question of importance. They use idle (awaiting job) servers, increasing in number with increasing incoming customers rate, forming conditional free systems with variable servers' quantity.

Systems with variable servers' number are also interesting in the cases when servers' quantity in a system is random at its nature. These and other reasons are supporting appearance of systems with variable servers' number and make interest of its investigation.

In the article systems with variable servers' number are named as M/M/v, where 'v' means 'variable' what denotes variable servers' quantity. Poisson was chosen as income and servicing processes due to its simplicity and a vast volume of existing investigations.

1.1 Related works.

Author hasn't found at his best knowledge queue theory any work with explicit */*/v systems' analysis. Investigations of the systems with variable parameters look like the most close to this area: Cox processes, compound and phase process, mixed processes e. t. c.

Analysis systems with variable servers' number are performed in many scientific fields. For example, the Chord [2], Tapestry [3] systems allow nodes join and leave system at random, perform node location and recovering. In application areas mutable systems were wide investigated ([4,5] for example), but its own specific questions were put and answered (content preservation, vulnerability e. t. c.) leaving queuing questions aside. Another example: compound Poisson process modeling radiation damage in [4], fault tolerant systems [1-3] congestion in network [5] and so on.

Used in the paper division the system into conditionals is similar to technique found in [6], where variable waiting room in queues was investigated.

2 Problem Formulation

The system is assumed to be within time stationary condition. We'll investigate two subsets of its states: conditional busy if there is no idle server, and conditional free if there is at least one idle server. From arriving customer's view a system is conditional busy if it is to be placed in its queue and conditional free if it is to be placed to servers' pool after arrival. If a customer has seized the last idle server then the system is not longer conditional free because next arriving customer will go to the queue. And vice versa: if in a conditional busy system with empty queue a customer was serviced then the system is not longer conditional busy because next arriving customer will go to a server pool.

In the paper independent identical distributed Poisson services X_i with mean service rate μ were used.

Poisson input has rate λ and $a = \lambda / \mu$. Servers' quantity is provided with random variable U with known probability distribution $\{u_k\}$ and mean τ.

Incoming customer in a conditional busy system is placed in a queue and waits as long as necessary for later servicing. So in the conditional busy system its queue is the object of investigations.

Incoming customer in a conditional free system is to be placed in the server pool and system is expected to be still free. It is happens if there is at least one free server after customer's departure or it is rejected otherwise if one wants stay in this model. So in the conditional busy system its server's pool is the object of investigations

The main task of the article is to obtain M/M/v models' parameters (state probability distributions, queue mean and variance e. t. c.) and show that they correspond to known ones in suitable conditions. It is done using deterministic probability distributions. Finally some discussion and generalizations were made.

3 Problem Solution

3.1 M/M/v loss system

Model of conditional free (as defined in section 2) M/M/s+1 system is general M/M/s Erlang loss system. State probabilities for stationary states of a M/M/v system are achieved using corresponding stationary states of M/M/s Erlang loss systems within which the system stays in given relative duration time (probability). We assume that server number changing is as slow as necessary to keep stationary state.

If server without customer leaves system or new idle server enters a system then nothing happens with servicing in the system. If server with customer leaves a system and there is an idle server then one may think that the servers were swapped: idle servers leaves and that to be leaved stays in the system, or one may think that the customer was passed to idle server without harm to servicing process. It may be done because servers are assumed to be indistinguishable. If server with a customer leaves a system and there is no idle server then the customer leaves the system also.

<u>Proposition 1.</u> Customers and servers numbers form full and disjoint set of parameters qualifying the system: 1) every system state is characterized with pair of customer and server numbers and 2) there are no identical (not distinguished) or overlapping system states with different parameters. This proposition is natural to queuing models and it seems that it doesn't require any additional comments.

Let Q is number of customers in the system, U is servers' number, and P^s is a norm multiplier. Then from M/M/s Erlang loss system model one gets:

$$P_{i,s} = P\{Q = i, U = s\} = \begin{cases} P^s \dfrac{a^i}{i!}, & i \le s < 0 \\ 0, & i > s \end{cases}$$

Norm equation follows from assumption that union of all possible system states with fixed server's and arbitrary customers' numbers have to produce macro state with given server's number. Proposition 1 tells that used parameters set is full then if one evaluates arbitrary chosen server's number probability from system's state probabilities aggregating customers, then the result is to be exactly the same server's probability $\{u_k\}$ as given in section 2. We use direct probability addition in the servers' probability calculation because due to proposition 1 these parameters are disjoined set against their index. So norm equation and norm multiplier are:

$$u_s = \sum_{i=0}^{s} P\{Q = i, U = s\} = \sum_{i=0}^{s} P^s \frac{a^i}{i!}$$

$$P^s = \frac{u_s}{\sum_{i=0}^{s} a^i / i!}$$

Then state probabilities are:

$$P_{i,s} = \begin{cases} \dfrac{a^i u_s}{i! \sum_{j=0}^{s} a^j / j!}, & i \le s \\ 0, & i > s \end{cases}, \tag{1}$$

Note, that here we accept zero server quantity state probability despite of the fact that in this state all incoming customers are to be rejected and lost. This is a valid stationary state of turned off customer servicing machine.

Probability generation function is found as:

$$G(x, y) = \sum_{i,s=0}^{\infty} P_{i,s} x^i y^s = \sum_{s=0}^{\infty} \sum_{i=0}^{s} \frac{a^i u_s x^i y^s}{i! \sum_{j=0}^{s} a^j / j!} \tag{2}$$

Let's see at

$$\frac{\partial G(x,y)}{\partial x} = \sum_{\substack{s=0\\ i=1,s}}^{\infty} iP_{i,s}x^{i-1}y^s = \sum_{\substack{s=0,\infty\\ i=1,s}} \frac{ia^i u_s x^{i-1}y^s}{i!\sum_{i=0}^{s} a^i / i!} = \tag{3}$$

$$= \sum_{s=0}^{\infty} a \sum_{i=0}^{s-1} \frac{a^i u_s x^i y^s}{i!\sum_{i=0}^{s}\frac{a^i}{i!}} \pm a\sum_{s=0}^{\infty}\frac{(xy)^s u_s a^s}{s!\sum_{i=0}^{s}\frac{a^i}{i!}} =$$

$$= aG(x,y) - a\sum_{s=0}^{\infty}(xy)^s u_s B(s,a)$$

Here $B(s,a)$ is Erlang loss formula. We obtain a differential equation for $G(x,y)$. Solution of equation (3) produces equation (2) back again and so has no worth.

Note: in (3) the serial contains as parameter only product of arguments. The mean of the fact is still not understood by author.

Mean customer number in the system (carried load) is got with substitution ones as arguments in (3) and it may be made once again for variance calculation:

$$E(Q) = \frac{\partial}{\partial x} G(x,y)_{\substack{x=1\\ y=1}} = \tag{4}$$

$$= a\left(1 - \sum_{s=1}^{\infty} u_s B(s,a)\right)$$

$$V(Q) = \frac{\partial^2}{\partial x^2} G(x,y) + E(Q)(1-E(Q)) =$$

$$= \frac{\partial}{\partial x} a\left(G(x,y) - \sum_{s=1}^{\infty} u_s (xy)^s B(s,a)\right) +$$

$$+ E(Q)(1-E(Q)) =$$

$$= (a+1)E(Q) - a\sum_{s=1}^{\infty} s u_s B(s,a) - E^2(Q)$$

Result for the mean may be seen as system's load in its every M/M/s slice state is proportional to its probability (relative duration) or system's loss id evaluated as weighted loss sum of corresponding systems.

Probability of incoming customers entered (not rejected) into the M/M/s system is $\upsilon_s = (1 - B(s,a))$. Because of independence server numbering and incoming Poisson processes we can gather customers entered in all time slices within its relative durations u_s and evaluate entering customers with unity incoming rate for the whole system (and use (4)):

$$\upsilon = \sum_{s=0}^{\infty} u_s \upsilon_s = \left[\sum_{s=0}^{\infty} u_s - \sum_{s=0}^{\infty} u_s B(s,a)\right] = E(Q)/a$$

If we write $L = E(Q)$, $a = \lambda\tau$ ($\tau = \mu^{-1}$ is mean service time), $\Lambda = \lambda\upsilon$ (real customer rate entering the system) and notice that in a conditional free system there is no waiting time in queue so $W = \tau$, then we have proved validity Little's theorem for conditional free M/M/v system:

$$L = E(Q) = a\upsilon = \Lambda W$$

3.2 Conditional busy delay M/M/v system with unbounded queue.

Let's again Q is number of customers in the system, queue length is L, and U is servers' number. Here a customer is to be placed into queue. In case of M/M/v system U is not a constant, so there is in not simple relation between Q and L like $Q = L + U$. This model differs from just discussed in servicing customers. A customer which was been in service with lost server is considered as suspended placed back in a queue and its service resumed in a due time. Because of Poisson type of service, loosing serving node incident doesn't hesitate analyze.

System's probabilities are made like (1) and they are also constitute full and disjoined parameterized set.

In busy M/M/v system we have Erlang delay system probability distribution for every state with fixed servers' number and exceeding it customers' number:

$$P_{i,s} = P\{Q = i, S = s\} = \begin{cases} P^s \dfrac{a^i}{s^{i-s}}, & i \ge s > 0 \\ 0, & i < s \end{cases} \tag{5}$$

Now in delay systems we will assume that zero server state has null probability ($u_0 = 0$) so it avoids null division in (5) and derived formulae. When $u_0 = 0$ there no idle server so formally this state may be considered within this framework, but calculations in 'parent' Erlang delay model requires server, so technically zero server model cannot be achieved this way. Infinity may appear in (5) with zero server state because of servers lack to service incoming customers. Hence with nonzero incoming rate the system accumulates infinite number of customer along its way to stationary state. From another point of view this state is also of no mean: how can system be busy if it has no server? What is busy there? We see that work with this state brings some troubles which are better to avoid throwing out this state (requiring zero probability of its existence). We exclude such case in later calculations assuming zero probability of its existence.

$$u_s = P\{U = s\} = \sum_{i=s}^{\infty} P\{Q = i, U = s\} =$$
$$= \sum_{i=s}^{\infty} P^s \frac{a^i}{s^{i-s}} = a^s \sum_{i=0}^{\infty} P^s \left(\frac{a}{s}\right)^i = \frac{a^s P^s}{1 - a/s}$$

So norm multipliers are:

$$P^s = u_s \frac{(1 - a/s)}{a^s}.$$

And state probabilities are

$$P_{i,s} = \begin{cases} \left(\dfrac{a}{s}\right)^{i-s} (1 - a/s) u_s, & i \geq s \\ 0 & i < s \end{cases} \quad (6)$$

If we want to find "queue length" / "server number" parameterized state probabilities, then:

$$P\{L = J\} = \sum_{s=1}^{\infty} P\{L = J \mid S = s\} P\{S = s\} = \quad (7)$$
$$= \sum_{s=1}^{\infty} \left(\frac{a}{s}\right)^J \left(1 - \frac{a}{s}\right) u_s = a^J \sum_{s=1}^{\infty} \left(\frac{1 - a/s}{s^J}\right) u_s =$$
$$= a^J \sum_{s=1}^{\infty} \frac{u_s}{s^J} - a^{J+1} \sum_{s=1}^{\infty} \frac{u_s}{s^{J+1}} =$$
$$= a^J M^{-J}[U] - a^{J+1} M^{-(J+1)}[U]$$

Here $M^{-J}[U]$ is J-th negative moment of servers' probability distribution.

<u>Note</u>: For example if $a > 1$ and $u_1 > 0$ then $\lim_{J \to \infty} (a^J M^{-J}[U]) = \infty$. So convergence problem of serials in (7) is appeared and discussed later in section 3.1.2.

So (7) can be rewritten to:

$$P\{L \geq J\} = a^J M^{-J}[U] \quad (8)$$
$$P\{L > J\} = a^{J+1} M^{-(J+1)}[U]$$

Let's check the norm of probability distribution (7–8):

$$\sum_{J=0}^{\infty} P\{L = J\} = P\{L \geq 0\} = a^0 M^{-0}[U] = \sum_{i=0}^{\infty} u_i = 1$$

The queue's mean length calculation results in:

$$L = \sum_{J=1}^{\infty} P\{L \geq J\} = \sum_{J=1}^{\infty} a^J M^{-J}[U] =$$
$$= \sum_{i=1}^{\infty} u_i \sum_{j=1}^{\infty} \left(\frac{a}{i}\right)^j = \sum_{i=1}^{\infty} \frac{a}{i} \frac{u_i}{1 - a/i}$$

This result also is intuitively follows from the assumption that M/M/v conditional busy system's queue length is equal $\dfrac{a/s}{1 - a/s}$ in every s-server' state (M/M/s conditional busy system), gathered with account of their probabilities u_s:

$$E(L) = \sum_{s=0}^{\infty} E(L \mid S = s) P\{S = s\} = \sum_{s=0}^{\infty} \frac{u_s a/s}{1 - a/s} \quad (9)$$

To calculate variance one has to deal with the serial:

$$< i(i-1) P\{L = i\} >=$$
$$= \sum_{j=1}^{\infty} j(j-1)[a^j M^{-J} - a^{j+1} M^{-(j+1)}] =$$
$$= \sum_{j=1}^{\infty} [(j-1)^2 a^j M^{-J} + (j-1) a^j M^{-J}] -$$
$$- \sum_{j=1}^{\infty} [j^2 a^{j+1} M^{-(j+1)} + j a^{j+1} M^{-(j+1)}] =$$
$$= 2 \sum_{j=1}^{\infty} j a^{j+1} M^{-(j+1)} = 2 \sum_{s=1}^{\infty} u_s \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \frac{a^{i+1}}{s^{i+1}} =$$
$$= 2 \sum_{s=1}^{\infty} \frac{u_s}{1 - a/s} \sum_{j=1}^{\infty} \frac{a^{j+1}}{s^{j+1}} = 2 \sum_{s=1}^{\infty} u_s \frac{(a/s)^2}{(1 - a/s)^2}$$

So queue's length variance is equal:

$$V(L) = 2 \sum_{s=1}^{\infty} u_s \frac{(a/s)^2}{(1 - a/s)^2} + E(L)[1 - E(L)] \quad (10)$$

Here $E(L)$ is the queue's length mean (9).

3.1.1 Little's theorem for conditional busy M/M/v system.

This theorem is not valid in conditional busy state in its direct meaning but if we write serials with servers' number slice probabilities for general Erlang delay systems with possible idleness (here C(s,a) is Erlang delay formula) then it was correct for M/M/v systems.

$$E(Q)=\sum_{s=1}^{\infty} u_s C(s,a)\left[a+\frac{a}{(1-a/s)s}\right]$$

Mean waiting time in the “servicing place” of busy systems equal $\tau=\mu^{-1}$. The rest waiting will be in a queue.

For FIFO servicing discipline the mean in queue waiting for conditional busy M/M/v system $E(W_{Q}^{c})$ is evaluated gathering sliced waiting times for M/M/s systems with its probabilities:

$$E(W_{Q}^{c})=\sum_{s=1}^{\infty}\frac{\tau\, u_s}{s(1-a/s)}$$

Mean servicing time in a conditional busy s-server pool is $E(W_s^c)=\tau$. In a no conditional busy system case we must multiply these results with C(s,a) for every slice to break conditions and with servers’ probability also:

$$E(W)=E(W_Q)+E(W_s)=$$
$$=\sum_{s=1}^{\infty}\tau\, u_s C(s,a)\left[1+\frac{1}{s(1-a/s)}\right]$$

There is seen that $E(Q)=\lambda E(W)$.

$L\neq\lambda W$ in a conditional busy system. But it is worked if we consider as the system only queue alone without accounting for servers pool as it was suggested for investigation subject in chapter 2. Queue mean from (9) is $E(L)=\sum_{s=1}^{\infty}\frac{u_s(a/s)}{s(1-a/s)}$, waiting time is $E(W_{Q}^{c})=\sum_{s=1}^{\infty}\frac{\tau\, u_s}{s(1-a/s)}$, so $E(L)=\lambda E(W_Q^c)$ in this case.

3.1.2 Convergence of serials with negative moments.

In case $a=1$ we have to investigate convergence of negative moments sequence $\{M^{-J}[U]\}$. Obviously, every moment exists.

Theorem 1. Sequence $\{M^{-J}[U]\}$ has a limit and it is equal u_0+u_1:

$$|\,u_0+u_1-M^{-n}[U]\,|=|\,u_1(1-1/s^n)\big|_{s=1}-\sum_{s=2}^{\infty}\frac{u_s}{s^n}\,|=$$
$$=\sum_{s=2}^{\infty}\frac{u_s}{s^n}\le 2^{-n}\sum_{s=2}^{\infty}u_s\le 2^{-n}$$

So with given $\varepsilon>0$ we put $n=-\log_2\varepsilon$, after which all higher negative moments will be in ε area of u_1. Proved.

Note: If we want to receive serial with these moments, we have to claim $u_1=0$.

The case $a<1$ is obvious: $\lim_{J\to\infty}(a^J M^{-J}[U])=0$.

In case $a>1$, let $b=\lceil a\rceil>1$ is the least integer above a. Then we divide moment’s serial to two parts Σ_b,Σ_∞:

$$a^J M^{-J}[U]=\sum_{s=1}^{b-1}u_s\left(\frac{a}{s}\right)^J+\sum_{s=b}^{\infty}u_s\left(\frac{a}{s}\right)^J=$$
$$=\Sigma_b+\Sigma_\infty$$

Their limits at large numbers are:

$$0\le\lim_{J\to\infty}\Sigma_\infty=\lim_{J\to\infty}\sum_{s=b}^{\infty}u_s\left(\frac{a}{s}\right)^J\le$$
$$\le\lim_{J\to\infty}\sum_{s=b}^{\infty}u_s\left(\frac{a}{b}\right)^J\le\sum_{s=1}^{\infty}u_s\lim_{J\to\infty}\left(\frac{a}{b}\right)^J=0$$
$$\lim_{J\to\infty}\Sigma_b=\lim_{J\to\infty}\sum_{s=1}^{b-1}u_s\left(\frac{a}{s}\right)^J\ge\lim_{J\to\infty}\sum_{s=0}^{b-1}u_s\left(\frac{a}{b}\right)^J\ge$$
$$\ge\sum_{s=0}^{b-1}u_s\lim_{J\to\infty}\left(\frac{a}{b}\right)^J=\begin{cases}\infty, & \sum_{s=0}^{b-1}u_s>0\\ 0, & \sum_{s=0}^{b-1}u_s=0\end{cases}$$

If $\sum_{s=0}^{b-1}u_s=0$, then $\Sigma_b=0$ for all J.

Hence to have convergence for serials in (7) we have to claim (it is equal stability requirement $1>\lambda/\mu s$ for all appeared there M/M/s systems):

$$P\{U\in[1,a)\}=0. \qquad (11)$$

3.4 Comparison with existing results.

The M/M/s system can be achieved from M/M/v system's model using deterministic probability distribution p^s:

$$p_j^s = \begin{cases} 1, \ j = s > 0 \\ 0, \ j \neq s \end{cases} . \tag{12}$$

Queue length mean (9) and variance (10) with probability distribution (12) for conditional busy systems are calculated as follows and coincide with corresponding conditional busy system M/M/s values:

$$L = \frac{a/s}{1 - a/s},$$

$$V = 2\sum_{j=1}^{\infty} p_j \left(\frac{a}{j}\right)^2 \frac{1}{(1 - a/j)^2} + L(1 - L) =$$

$$= \frac{(a/s)}{(1 - a/s)}\left[\frac{a/s}{(1 - a/s)} + 1\right] = \frac{a/s}{(1 - a/s)^2}$$

For a conditional free M/M/v system the queue length mean coincides to corresponding Erlang loss system with deterministic servers' probability distribution as told after equation (4).

If we look at (4) we see equation for evaluating carried load L coinciding with well known results for M/M/s system with deterministic probability (12).

The same actions with deterministic probability can be done to customize the Little's theorem.

Author has checked probability distributions of loss and busy M/M/v system models with GPSS simulation software. Tests were quite obvious and received results support offered equations. Simulation results are not included here due to article's space limitation and its simplicity.

4 Conclusion.

Perhaps it may be more convenient in some cases to place more information (such as distribution type) in queue's server's number classification place instead of simple letter 'v'. For example, M/M/M queue has Poisson server's distribution with Poisson servicing.

Comparison with M/M/s systems gives the heuristic that M/M/v system may be considered as generalization of some kind of known M/M/s systems. It seems that this fact has its own science significance.

In this paper server number was controlled with a random process. One can use functional control instead and investigate system's behavior in a framework of control, decision or another theory.

Stationary probability equations were used in calculations. To be this action correct, either server changing is to occur as rare as enough for the system to come back into stationary states and weight of intermediate states is too small to take it into account, or we have to treat these results as an approximation.

Equation (11) seems to be very strong restriction. Another approach is needed to correct description of these cases. GPSS simulations gave stable distributions in some cases when load belongs to 'forbidden area'. 'Stable' means that a system comes into some fixed probability distribution with varying involved customers number and initial state of the system. So claim of (11) is better to treat as application boundary of used methods.

References:

[1] I. Foster, C. Kesselman, S. Tuecke, *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*, http://www.globus.org/alliance/publications/papers/anatomy.pdf.

[2] D. Liben-Nowell, H. Balakrishnan, D. Karger. Analyze of the evolution of Peer-to-Peer systems; *Proc. of the 21-th symposium on distributed computing*, 2002.— pp. 233–242

[3] B. Zhao, J. Kubiatovicz, D. Joseph, Tapestry: An Infrastructure for Fault tolerant wide area Location and routing, Tech report UCB/CSD-01-1141 EECS uni. of California, 2001, 27p.

[4] E. Gudzovska–Novak, S. Ritter, G. Taucher–Scholz, G. Kraft, Compound Poisson processes and clustered damage of radiation induced DNA double strand breaks, *ACTA physica polonica*, Vol. 31, No 5, 2000, pp. 1109–1123

[5] E. Altman, K. Avrachenkov, C. Barakat, R. Nunez-Queija, *State-dependent M/G/1 type queuing analysis for congestion control in data networks*, www-sop.inria.fr/maestro/personnel/K.Avrachenkov/pubs/Infocom01.ps, 2001, 10p.

[6] J. Rueda, *The $Ph_t/Ph_t/s/c$ Queuing Model and Approximation*, master's thesis, Virginia univ., 2003, 154p.