ANALYTICAL MODEL OF QOS MECHANISM – PRIORITY QUEUING USING MARKOV'S CHAINS

DUŠAN NEMČEK

University of Žilina - Faculty of Informatics and management science, Univerzitná 8215/1, 010 26 Žilina
email: dusan.nemcek@fri.uniza.sk

Abstract: This paper describes the creation of an analytical model for QoS mechanism - Priority Queuing in a system with two queues. Into the queues arrive two independent streams of packets. The described model is designed to calculate a packet delay (a response time) which is one of the key QoS parameters. For this purpose, theory of Markov's chains is used.

Keywords: Quality of Service, QoS, response time, analytical model, Markov's chain, packet, transition graph, exponential distribution, generating function

1 Mechanism Priority Queuing

Priority queuing (PQ) is one of the congestion management tools. They allow controlling congestion by determining the order in which packets are sent out through an interface based on priorities assigned to those packets. The congestion management tools entail the creation of queues, assignment of packets to those queues based on the classification of the packet and the scheduling of the packets in a queue for the transmission. In PQ, one service class of packets is assigned to each queue with specific priority. The priority of particular queues is different. Within each of queues, packets are scheduled in FIFO order. First packet is taken from queue into transmitter and then it is sent bit by bit through output port. The queue with the highest priority is served first until it is empty. Then, the lower priority queues are served according to their priority. In other words, packet is taken from the queue, only if all queues with higher priority are empty. This procedure is repeated every time a packet is sent. A rate of sent packets from transmitter is given by bandwidth of output port.

2 Analytical model of Priority Queuing

Mechanism PQ can be modeled using different ways, similarly as other QoS mechanisms. This section show, how Markov's chains can be used for this purpose. For the sake of simplicity, model with only two infinite queues is assumed here. The goal is to find the *mean response time* for packets of priority queue. It is important performance measure.

2.1 Basic terms

Necessary conditions for using Markov's chains are *Poisson* distribution of packet arrivals and *exponential* distribution for service time. Thus if $A(x)$ denotes the number of packets that arrive during any time interval into the k-th queues of length x, the Poisson distribution of packet arrivals is given by

$$Pr[\,A(x)=n\,]=\frac{(\lambda_k x)^n}{n!}e^{-\lambda_k x} \qquad (1)$$

Parameter λ_k is *intensity* of packet arrivals into the k-th queue and is measured in packet per second. Compared to original, packets are not transmitted bit by bit but as a whole. Time of packet in transmitter (interval between selection packets to transmitter and his transmission on output port) is given by packet size and it is called *service time*. It has the exponential distribution for Markov's system. Thus service time for packets of k-th queue I_k with mean $1/\mu_k$ is given by

$$Pr[\,I_k \le x\,]=1-e^{-\mu_k x} \qquad (2)$$

where μ_k is *service intensity* for packets of k-th queue. The mean service time for packet of k-th queue is an inverse of service intensity and here is denoted as τ_k.

$$\tau_k=\frac{1}{\mu_k} \qquad (3)$$

Offered load ρ_k with packets for k-th queue is defined as:

$$\rho_k=\frac{\lambda_k}{\mu_k} \qquad (4)$$

Total offered load ρ is defined as:

$$\rho=\sum_{k=1}^{K}\rho_k \qquad (5)$$

where K count of queues in the system. In this model K = 2. Necessary and sufficient condition, that the Markov's chain has a unique steady-state distribution is $\rho < 1$. *Waiting time* W_k is the interval from the arrival time of packet for k-th queue to the time when its transmission is started. In other words, W_k is the total time of packet in the queue. *Response time* T_k is defined as the time interval from the arrival time of an arbitrary packet for k-th queue to the time, when the packet leaves the system after the transition is finished. The response time consists of the waiting time and the service time.

2.2 Transition graph

A state of considered system is given by three indexes (k,i,j), where $k=\{1,2\}$, $i=\{1..L_1\}$ and $j=\{1..L_2\}$. k determines queue, that packet is actually transmitted, i determines number of packets of first queue (*priority packets*) and j the number of packet of second queue (*ordinary packets*) that are actually in the system. A value of state is determined by probability that system is in this state. In further text, only the steady-state probabilities are considered. If system can pass from one to another state, a transition exists between these states. Value of transition is determined by the mean intensity of arrival (λ_k) or departure (μ_k). All possible states of system and transition between them can be represented by transition graph (Figure 1).

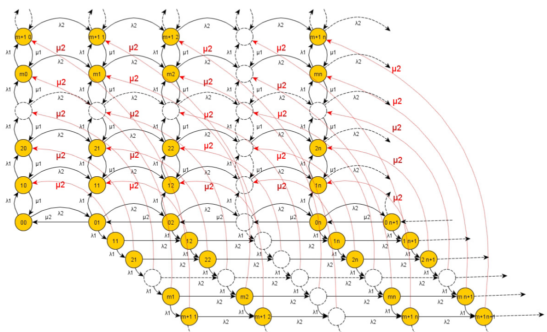

Figure 1 – Transition graph of PQ with two queues

2.3 Solution

Let

$$F_{1,i}(y)=\sum_{i=0}^{\infty}p_{1ij}y^i\ ,\ F_{2,i}(y)=\sum_{i=1}^{\infty}p_{2ij}y^i \qquad (6)$$

be the probability generating functions for states in i-th row on the vertical and horizontal layer of graph in the Figure 1.

Let

$$H_1(xy)=\sum_{i=1}^{\infty}F_{1i}x^i\ ,\ \ H_2(xy)=\sum_{i=0}^{\infty}F_{2i}x^i \qquad (7)$$

be probability generating functions for states on the vertical and horizontal layer of graph in the Figure 1.

Let $p_{0,0}$ be a probability of empty system. A probability generating probability function for the number of packets present in the system is given by

$$H(xy)=p_{00}+H_1(xy)+H_2(xy)=\sum_{k=1}^{2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}p_{kij}x^i y^j \qquad (8)$$

From *(7)*, *(8)* we have:

$$H_1(11)=\rho_1=\lambda_1/\mu_1,\ H_2(11)=\rho_2=\lambda_2/\mu_2,\ H(11)=1 \qquad (9)$$

and next for their first derivate in $x=1$, $y=1$ we have:

$$\left[\frac{\partial H(xy)}{\partial x}\right]_{x=y=1} = \sum_{i=1}^{\infty} i x^{i-1}[F_{1,i}(y)+F_{2,i}(y)] = E[L_1] \quad (10)$$

$$\left[\frac{\partial H(xy)}{\partial y}\right]_{x=y=1} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} j y^{j-1}[p_{1,i,j}+p_{2,i,j}] = E[L_2] \quad (11)$$

It is well-known fact that for Markov's systems with infinite queues and shared service holds:

$$p_0 = 1-\sum_{k=1}^{K}\rho_k \quad (12)$$

where p_0 is probability, that a system is empty.
Thus for p_{00} we have in this model:

$$p_{1,0,0} = p_{2,0,0} = p_{0,0} = 1-\rho_1-\rho_2 \quad (13)$$

Following system of equations describes probabilities of particular states:

$$(\lambda_1+\lambda_2)p_{0,0} = \mu_1 p_{1,1,0} + \mu_2 p_{2,0,1} \quad (14)$$

$$(\mu_1+\lambda_1+\lambda_2)p_{1,i,0} = \lambda_1 p_{1,i-1,0} + p_{2,i,1} + \mu_1 p_{1,i+1,0} \quad i=1...\infty \quad (15)$$

$$(\mu_2+\lambda_1+\lambda_2)p_{2,i,1} = \lambda_1 p_{2,i-1,1} \quad i=1...\infty \quad (16)$$

$$(\mu_2+\lambda_1+\lambda_2)p_{2,0,j} = \lambda_2 p_{2,0,j-1} + \mu_1 p_{1,1,j} + \mu_2 p_{2,0,j+1} \quad j=1...\infty \quad (17)$$

$$(\mu_1+\lambda_1+\lambda_2)p_{1,1,j} = \lambda_2 p_{1,1,j-1} + \mu_1 p_{1,2,j} + \mu_2 p_{2,1,j+1} \quad j=1...\infty \quad (18)$$

$$\begin{aligned}(\mu_1+\lambda_1+\lambda_2)p_{1,i,j} &= \lambda_2 p_{1,i,j-1} + \mu_1 p_{1,i+1,j} + \\ &+\mu_2 p_{2,i,j+1} + \lambda_1 p_{1,i-1,j} \quad i=2...\infty, j=1...\infty\end{aligned} \quad (19)$$

$$(\mu_2+\lambda_1+\lambda_2)p_{2,i,j} = \lambda_2 p_{2,i,j-1} + \lambda_1 p_{2,i-1,j} \quad i=1...\infty, j=2...\infty. \quad (20)$$

By multiplying these equations with x^i, y^j and summing for all i,j we obtain:

$$(\lambda_1+\lambda_2)p_{0,0} = \mu_1 p_{1,1,0} + \mu_2 p_{2,0,1} \quad (21)$$

$$(\mu_1+\lambda_1+\lambda_2)\sum_{i=1}^{\infty} p_{1,i,0}x^i = \lambda_1 x\sum_{i=0}^{\infty} p_{1,i,0}x^i + \frac{\mu_2}{y}\sum_{i=1}^{\infty} p_{2,i,1}x^i y + \frac{\mu_1}{x}\sum_{i=2}^{\infty} p_{1,i,0}x^i \quad (22)$$

$$(\mu_2+\lambda_1+\lambda_2)\sum_{i=1}^{\infty} p_{2,i,1}x^i y = \lambda_1 x\sum_{i=0}^{\infty} p_{2,i,1}x^i y \quad (23)$$

$$\begin{aligned}(\mu_2+\lambda_1+\lambda_2)\sum_{j=1}^{\infty} p_{2,0,j}y^j &= \lambda_2 y[\sum_{j=1}^{\infty} p_{2,0,j}y^j + p_{0,0}] + \\ &+\frac{\mu_1}{x}[\sum_{j=0}^{\infty} p_{1,1,j}xy^j - p_{1,1,0}x] + \frac{\mu_2}{y}[\sum_{j=1}^{\infty} p_{2,0,j}y^j - p_{2,0,1}y]\end{aligned} \quad (24)$$

$$\begin{aligned}(\mu_2+\lambda_1+\lambda_2)\sum_{j=0}^{\infty} p_{1,1,j}xy^j - p_{1,1,0}x &= \lambda_2 y\sum_{j=0}^{\infty} p_{1,1,j}xy^j + \\ &+\frac{\mu_1}{x}[\sum_{j=0}^{\infty} p_{1,2,j}x^2y^j - p_{1,2,0}x^2] + \frac{\mu_2}{y}[\sum_{j=1}^{\infty} p_{2,1,j}xy^j - p_{2,1,1}xy]\end{aligned} \quad (25)$$

$$\begin{aligned}&(\mu_1+\lambda_1+\lambda_2)[H_1(xy) - \sum_{j=1}^{\infty} p_{1,1,j}xy^j - \sum_{i=1}^{\infty} p_{1,i,0}x^i] = \\ &= \lambda_2 y[H_1(xy) - \sum_{j=0}^{\infty} p_{1,1,j}xy^j] + \frac{\mu_1}{x}[H_1(xy) - \sum_{j=1}^{\infty} p_{1,2,j}x^2y^j - \\ &-\sum_{j=1}^{\infty} p_{1,1,j}xy^j - \sum_{i=1}^{\infty} p_{1,i,0}x^i] + \frac{\mu_2}{y}[H_2(xy) - \sum_{j=2}^{\infty} p_{2,1,j}xy^j - \\ &-\sum_{j=1}^{\infty} p_{2,0,j}y^j - \sum_{i=1}^{\infty} p_{2,i,1}x^i y] + \lambda_1 x[H_1(xy) - \sum_{i=1}^{\infty} p_{1,i,0}x^i]\end{aligned} \quad (26)$$

$$\begin{aligned}&(\mu_2+\lambda_1+\lambda_2)[H_2(xy) - \sum_{j=1}^{\infty} p_{2,0,j}y^j - \sum_{i=1}^{\infty} p_{2,i,1}x^i y] = \\ &= \lambda_2 y[H_2(xy) - \sum_{j=1}^{\infty} p_{2,0,j}y^j] + \lambda_1 x[H_2(xy) - \sum_{i=0}^{\infty} p_{2,i,1}x^i y]\end{aligned} \quad (27)$$

After summing *(22), (25), (26)* we get:

$$\begin{aligned}&(\mu_1+\lambda_1+\lambda_2-\lambda_2 y-\lambda_1 x-\frac{\mu_1}{x})H_1(xy) = \\ &= \frac{\mu_2}{y}H_2(xy) + \lambda_1 x p_{00} - [\mu_1 F_{1,1}(y) + \frac{\mu_2}{y}F_{2,0}(y)]\end{aligned} \quad (28)$$

and after summing *(21), (23), (24), (27)* we get:

$$\begin{aligned}&(\mu_2+\lambda_1+\lambda_2-\lambda_2 y-\lambda_1 x)H_2(xy) = \\ &= (\lambda_2 y-\lambda_1-\lambda_2)p_{0,0} + \mu_1 F_{1,1}(y) + \frac{\mu_2}{y}F_{2,0}(y)\end{aligned} \quad (29)$$

From *(6)*, *(21)* and *(24)* we get relation between $F_{1,1}$ and $F_{2,0}$:

$$\begin{aligned}\mu_1 F_{1,1}(y) &= (\mu_2+\lambda_1+\lambda_2-\lambda_2 y-\frac{\mu_2}{y})F_{2,0}(y) + \\ &+(\lambda_1+\lambda_2-\lambda_2 y)p_{0,0}\end{aligned} \quad (30)$$

From *(28), (29), (30)* we are able to determine $H_1(xy)$ and $H_2(xy)$. Thus now we write according to *(8)*:

$$\begin{aligned}H(xy) &= \frac{\mu_1 p_{00}(1-x)y(\lambda_2 y-\lambda_2-\mu_2-\lambda_1(1-x))}{(\lambda_1-\lambda_1 x-x(\lambda_1(1-x)+\lambda_2(1-y)))y(\lambda_2 y-\lambda_2-\mu_2-\lambda_1(x-1))} + \\ &+\frac{F_{2,0}(y)(\lambda_1+\lambda_2+\mu_2-\lambda_2 y)(\lambda_2 x(1-y)-\mu_1 y(1-x))}{(\lambda_1-\lambda_1 x-x(\lambda_1(1-x)+\lambda_2(1-y)))y(\lambda_2 y-\lambda_2-\mu_2-\lambda_1(x-1))}\end{aligned} \quad (31)$$

Hence we have depending for $H(xy)$ only on unknown $F_{2,0}(y)$. For $x \to 1$, $y \to 1$ from *(9),(29)* and *(30)* we have:

$$F_{2,0}(1) = \frac{\lambda_2}{\lambda_1+\mu_2} \quad (32)$$

After partial derivation of *(31)* by x and applying limits $x \to 1$, $y \to 1$ and substituting *(32)* into *(31)*, we obtain the mean number of priority packets in the system finally:

$$E[N_1] = \frac{\lambda_1(\lambda_2\mu_1+\mu_2^2)}{(\mu_1-\lambda_1)\mu_2^2} \quad (33)$$

Now we apply *Little's formula* to *(33)* to obtain the mean response time of priority packets:

$$E[T_1] = \frac{E[N_1]}{\lambda_1} = \frac{\lambda_2\mu_1+\mu_2^2}{(\mu_1-\lambda_1)\mu_2^2} \quad (34)$$

3 Conclusion

This paper show, how can be used Markov's chains to modelling QoS mechanism Priority Queuing. The final result of the paper is formula for the response time of priority packet. For the sake of simplicity are assumed only two queues. Moreover, the queues are infinite. If we considered model with finite queues, the number of equations analogous to (14)-(20) would be increased to 16 and the model solution would be too complicated.

Literature:

1. HIDEAKI,T.: *Queueing analysis - Volume1: Vacation and Priority System*. Amsterdam: North-Holland, 1991. ISBN 0-444-88910-8
2. WALTER, J.: *Stochastické modely v ekonomii*. Bratislava: SNTL/ALFA, 1970. ISBN 04-317-70.
3. PEŠKO, Š., SMIEŠKO, J.: *Stochastické modely operačnej analýzy*. Žilina: EDIS vydavateľstvo ŽU, 1999. ISBN 80-7100-570-3
4. VEGESNA, S.: *IP Quality of Service for the Internet and the Intranets*. Indianapolis: Cisco Press, 2000
4. CISCO Systems: *Implementing Cisco Quality of Service – Volume 1*. San Jose: Cisco Press, 2008

Primary Paper Section: I

Secondary Paper Section: N