International Journal of Advanced Computer Technology (IJACT)

ISSN:2319-7900

Analytical Model of the Data Transfer System with the Time Quantum

Vladimir Hottmar; Bohumil Adamec

Department of Telecommunications and Multimedia,Faculty of Electrical Engineering,University of Zilina,Slovakia

Abstract

This paper works out the problem of service system with the round robin of data after expiration of the allocated time quantum q. It brings the solved mathematical model and quantification of conditions that occur with this special type of service. It shows the dependence between the system load, distribution of the stored requests in the queue and the number of repeated returns of the in-process requests into the queue. At the end the critical limit values for p, n, and ρ parameters are considered here as well as their impact on overall system operation [1] – [4].

Introduction

For the needs of this paper we suppose there is such a mechanism of data processing by the transmission system that it will provide its capacity for every request for equally long time, and maximally for the defined time interval (time quantum – q) within which the request processing takes place [5], [6]. If the allocated time quantum does not suffice for the processing, then the in-process request is interrupted, gets into the queue, while later processing will continue where the interruption occurred. This cyclical algorithm will proceed until the expiration of n time quanta necessary for processing the entire request. We will model the mechanism of the system operation by the queuing theory [1] – [4], while we work out the problem of the time quanta allocation in detail and quantify the transmission system characteristics.

Derivation of Significant Time Parameters for the Time Quantum Service Strategy

To quantify the transmission system characteristics we will use the model shown in Fig.1. The transmission system is represented by the node U. In this node there is a queue that represents the memory for storing the requests that enter the system from external environment (or from the preceding node), as well as the requests that were serviced and their service was interrupted after expiration of the time quantum q.

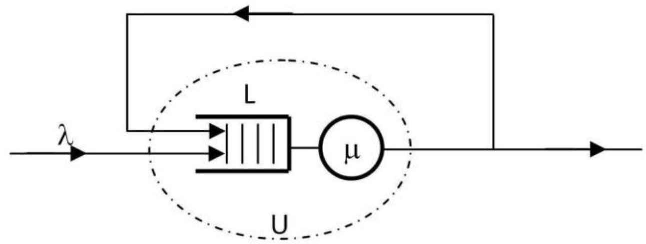

Figure 1. Model of the Service System with the Time Quantum Service Strategy

Furthermore, the node contains the execution unit of the transmission system that services the requests with the service intensity μ. For the needs of this model let us first define the following representation. The set of all processed messages will be represented to the set of requests and the execution unit of the transmission system will be represented to the set of service systems.

One of the first issues we have to deal with is how to determine the request flow entering the system. We will assume that the requests entering the system agree with the elementary request flow with the Poisson distribution of arrivals and exponential distribution of time between their arrivals, while λ is the parameter of the exponentially random variable that represents the intensity of the request arrivals from external environment [5], [6], [7], [8].

After arrival of the requests to the system they are put into queue. If the queue is empty (this occurs with the probability $p_0 = 1 - \rho$, where ρ is the system load defined by the appropriate relation $\rho = \lambda/\mu$), the arriving request is serviced immediately. Otherwise it is put into queue and waits for its servicing. Once the service of the request has finished, it leaves the system with the probability $\mu \cdot \rho$. If the allocated time quantum does not suffice for the request servicing, the servicing is interrupted, the request is put into the queue again while another request from the queue is chosen to be serviced. The requests are chosen to be serviced according to FIFO strategy. As we assumed in previous that the request entries into the system are divided by the Poisson distribution, we will again assume that each request service time agrees with the exponential distribution with the density probability distribution $\mu \cdot e^{-\mu \cdot t}$ and with the parameter μ [1]. Further on we will deal with the detailed analysis of the time

ISSN:2319-7900

conditions and subsequently with the quantification of significant parameters of the described service system.

For the needs of time parameters derivation we have to determine the general conditions of the transmission service system.

Let us first assume that the defined time quantum q will agree with the service time of at least one of the requests. If such a request exists, then after its servicing i.e. after expiration of the allocated time quantum q the request either leaves the system or enters another node.

Next, let us assume that the requests will enter the system one after another with different service time. Let us choose one of them and assume that n time quanta will be necessary for its entire servicing. The request service time will be $n \cdot q$. This will occur in the case of empty system and after the request arrival into the system its servicing will start immediately. At the same time we assume that during the request service time no other request will enter the system and the intelligent transmission system will choose such a utility algorithm that will recognize this state. The cyclical algorithm, (i.e. the return of the request after expiration of the time quantum q into the queue), will not be applied for this type of service. The distribution of the service time will be geometric with the parameter p and will be given by the relation [1], [2], [3]:

$$\begin{gathered} P(x = n \cdot q) = p^{n-1} \cdot (1-p) \\ 0 < p < 1 \quad n = 1; 2 \ldots k \quad 0 < q << t \end{gathered} \tag{1}$$

We will further assume that the system at the time of our request arrival contains j requests altogether. At the time of arrival of the observed request with the service time of $n \cdot q$, the $(j-1)$ requests are waiting in the queue before it and one request is being serviced. Let us define the time between the expiration of the allocated time quantum of the serviced request and its repeated return into the queue. This time will be called the transition time and will be marked as τ_i, where variable $i = 1, 2 \ldots m$. For the first transition of the observed request we have to take into account the fact that the request has to wait in the queue for $(j-1)$ requests in the queue before it to be serviced and that at the time of its arrival one request is just being serviced. The service time of j requests by the first transition is q. The state of the first transition of our observed request is given by the formula:

$$\begin{gathered} \tau_1 = a \cdot q + (j-1) \cdot q + 1 \cdot q \\ 0 \le a \le 1 \end{gathered} \tag{2}$$

The product $a \cdot q$ is the request service time at the time of arrival of the observed request in the queue. The time product $(j-1) \cdot q$ is the service time of $(j-1)$ requests that are in the queue and q is the service time of our request. The time product $(n-1) \cdot q$ equal to $(n-1)$ transitions, are left for the overall service.

By the second transition of the observed request through the service system there will be those from j requests whose service time is greater than q. These requests stay in the service system and are put into the waiting queue after the first transition. Their number is on average $j \cdot P(x > q)$ [1] – [4]. Then the probability of the service time of such requests is:

$$P(x > q) = p \tag{3}$$

For the sake of completeness we have to state that if the request service time is shorter than the time quantum q, then the appropriate probability equals $P(x \le q) = 1 - p$, while for the total probability $P(x > q) + P(x \le q) = 1$ is valid. We also have to realize that during the first transition further $\lambda \cdot E(\tau_1)$ requests on average entered the system, while λ is again the arrival intensity. The average length of the service time of the observed request during the second transition is given by the formula:

$$\begin{aligned} E(\tau_2) &= p \cdot j \cdot q + \lambda \cdot E(\tau_1) \cdot q + q = \\ &= \lambda \cdot E(\tau_1) \cdot q + q \cdot (1 + p \cdot j) = \\ &= q \cdot \left[\lambda \cdot E(\tau_1) + (1 + p \cdot j) \right] \end{aligned} \tag{4}$$

If we denote $v_2 = \lambda \cdot E(\tau_1) + (1 + p \cdot j)$, then the formula (4) can be rewritten into the form:

$$E(\tau_2) = q \cdot v_2 \tag{5}$$

In general we can rewrite for the i-th transition:

$$\begin{gathered} E(\tau_2) = q \cdot v_2 \\ i = 2; 3 \ldots n-1 \end{gathered} \tag{6}$$

As with the observed request there were $v_i - 1$ requests in the waiting queue in the i-th transition, it is valid:

$$E(\tau_{i+1}) = p \cdot (v_i - 1) \cdot q + \lambda \cdot E(\tau_1) \cdot q + q \tag{7}$$

Formula (7) can be written in the form:

$$E(\tau_{i+1}) = (p + \lambda \cdot q) \cdot E(\tau_i) + q \cdot (1-p) \tag{8}$$

If we denote $b = (p + \lambda \cdot q)$, then formula (7) can be rewritten into the form:

ISSN:2319-7900

$$E\left(\tau_{i+1}\right)= \\ = b^{i-1}\cdot E\left(\tau_2\right)\cdot q\cdot\left(1-p\right)\cdot\left[\left(1-b^{i-1}\right)/\left(1-b\right)\right] \quad (9)$$

The number of repeated returns of the considered request into the waiting queue (number of transitions) determines the average request stay time in the service system [1], [3]. If the number of requests in the waiting queue before the considered request was j altogether, then the average request stay time in the system can be marked $R_n(j)$, where n is the multiple of the time quantum q. Then, the request stay time of considered request $R_n(j)$ is given by the formula:

$$R_n\left(j\right)=E\left(\tau_1\right)+E\left(\tau_2\right)+\ldots+E\left(\tau_n\right)= \\ =\sum_{i=1}^{n}E\left(\tau_i\right) \quad (10)$$

On the basis of the relation (9) we will write the expression of the average time of the n-th transition in the form:

$$E\left(\tau_n\right)=b^{n-2}\cdot E\left(\tau_2\right)+ \\ +\left(1-b\right)^{-1}\cdot q\cdot\left(1-p\right)\cdot\left(1-b^{n-2}\right) \quad (11)$$

We can modify the element $q\cdot(1-p)\cdot(1-b)^{-1}$ from the relation (11) into the form:

$$q\cdot\left(1-p\right)\cdot\left(1-b\right)^{-1}=q\cdot\left[1-\lambda\cdot q/\left(1-p\right)\right]^{-1} \quad (12)$$

Parameter b is given by the relation $b = (p + \lambda\cdot q)$. If the element $\lambda\cdot q/(1-p) = \rho$ from the equation (12) is the system load, then term $q/(1-p) = \mu^{-1}$ has to represent the average service time. We will show this is true. As p and q are the parameters of geometric probability distribution given by the relation (1), at first we will determine the moment generating function of this probability distribution [1], [2], [3]:

$$m_x\left(t\right)=\sum_{n=1}^{\infty}e^{n\cdot t}\cdot P\left(x=n\cdot q\right)= \\ =\sum_{n=1}^{\infty}e^{n\cdot t}\cdot p^{n-1}\cdot\left(1-p\right) \quad (13)$$

Formula (13) can be rewritten into the form:

$$m_x\left(t\right)=e^t\cdot\left(1-p\right)\cdot\sum_{n=1}^{\infty}\left(e^t\cdot p\right)^{n-1} \quad (14)$$

Summation in the equation (14) represents the sum of geometric progression:

$$\sum_{n=1}^{\infty}\left(e^t\cdot p\right)^{n-1}=1/\left(1-p\cdot e^t\right) \quad (15)$$

Moment generating function can be rewritten into the form:

$$m_x\left(t\right)=e^t\cdot\left(1-p\right)/\left(1-p\cdot e^t\right) \quad (16)$$

By derivation of moment generating function we will determine the k-th moment:

$$d^k\left[m_x\left(t\right)\right]_{t=0}/dt^k=E\left[x^k\cdot e^{x\cdot t}\right]=E\left[x^k\right] \quad (17)$$

Next, we determine the first $E(x)$ and second $E(x^2)$ moment of generating moment function. For $t = 0$, the term $E(x)$ represents the average value. First moment will gain the form:

$$m_1\left(0\right)=1/\left(1-p\right) \quad (18)$$

Second moment will gain the form:

$$m_2\left(0\right)=p/\left(1-p\right)^2 \quad (19)$$

Next, by applying the fundamentals statistic equations $E(q\cdot x) = q\cdot E(x)$; $D(q\cdot x) = q^2\cdot D(x)$ and $D(x) = E(x^2) - E^2(x)$ will be:

$$q\cdot m_1\left(0\right)=q/\left(1-p\right)=\mu^{-1} \quad (20)$$

Equation (20) represents the average service time. System load is given by the formula:

$$\rho=\lambda\cdot q/\left(1-p\right) \quad (21)$$

Then the element from the equation (12) given by the (21) really represents the system load. Applying the relation (12), the expression (11) will gain the form:

$$E\left(\tau_n\right)=b^{n-2}\cdot E\left(\tau_2\right)+ \\ +\left(1-\rho\right)^{-1}\cdot q\cdot\left(1-b^{n-2}\right) \quad (22)$$

To be able to calculate the average stay time of the considered request in the system we have to determine the aver-

ISSN:2319-7900

age value of the time of the first transition $E(\tau_1)$ in the relation (10). This will be determined from the relation (2) in a following way:

$$\tau_1 = a \cdot q + (j-1) \cdot q + q = q \cdot (a+j) \qquad (23)$$

Next, the average value of the time of the first transition is:

$$\begin{aligned} E(\tau_1) &= E\left[q \cdot (a+j)\right] = \\ &= q \cdot \left[E(a) + E(j)\right] = \\ &= q \cdot \left[E(a) + E(j-1) + 1\right] \end{aligned} \qquad (24)$$

Expression $E(j-1) = Q$ represents the average value of the requests in the waiting queue. Subsequently, the relation (25) is valid:

$$\xi = \rho / 2 \qquad (25)$$

Expression (25) means, that the system by the service is loaded during $a \cdot q$ time by the load $\rho \cdot q/2$, because the average value equals $E(a)$ =1/2. The average value of the first transition time $E(\tau_1)$ will gain the form:

$$E(\tau_1) = q \cdot \left[\rho / 2 + Q + 1\right] \qquad (26)$$

After substituting into the relation (10) we will determine the average stay time of the considered request in the service system:

$$\begin{aligned} R_n(j) &= b^{n-2} \cdot E(\tau_2) + \\ &+ q \cdot (1 - b^{n-2}) / (1-\rho) \end{aligned} \qquad (27)$$

Next, it is necessary to determine the average number of requests in the system. This can be calculated from the Chinchin-Pollaczek formula [1] – [4]:

$$L = \rho + \left(\rho^2 + \lambda^2 \cdot \sigma^2\right) / 2 \cdot (1-\rho) \qquad (28)$$

By applying the fundamentals statistic equations and the formula (19) the dispersion will be:

$$\sigma^2 = p \cdot q^2 / (1-p)^2 \qquad (29)$$

The expressions (21) and (23) can be substitute into the relation (28):

$$L = \rho + \rho^2 \cdot (1+p) / 2 \cdot (1-\rho) \qquad (30)$$

The average number of requests in the waiting queue is given by the formula:

$$Q = \rho^2 \cdot (1+p) / 2 \cdot (1-\rho) \qquad (31)$$

Then the average number of requests in the waiting queue is given by the formula:

$$L = \rho + Q \qquad (32)$$

Discussion

Within the discussion we consider the real packet processing system [5] – [8]. We assume that the average size of the transported packets is $E(X)$ = 10 kbit (RTP packets transported video traffic) and the transfer capacity of the system is BR = 256 kbit/s (multilink PPP). In this case the size of the time quantum can be determined as $q = E(X)/5 \cdot BR$. We will also assume following operational parameters – system probability p = 0,85 and load coefficient ρ = 0,8. The results for the given service system are demonstrated in Fig. 2, Fig. 3 and Fig. 4.

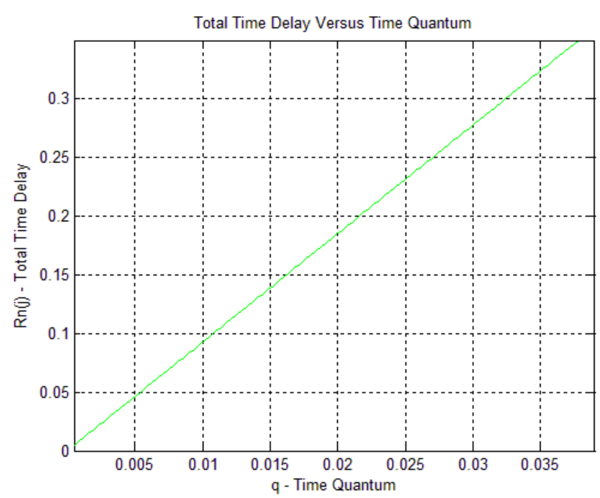

Figure 2. Total Time Delay Dependence on the Time Quantum

ISSN:2319-7900

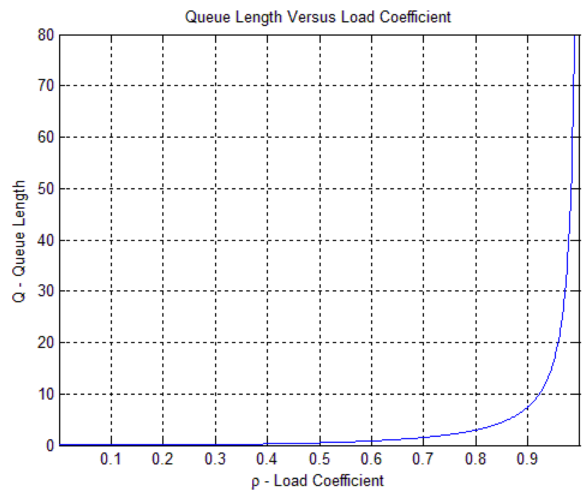

Figure 3. Queue Length Dependence on the System Load

Let us first analyze the allocated time quantum q. In this paper we deal with the time quantum with the mathematical point of view. In reality the time quantum is limited by the performance of the technical and software resources of the transported service system as well as the algorithm of the serviced requests. Serviced requests can be correctly divided into n independent time intervals. In real conditions it will always be $q > 0$.

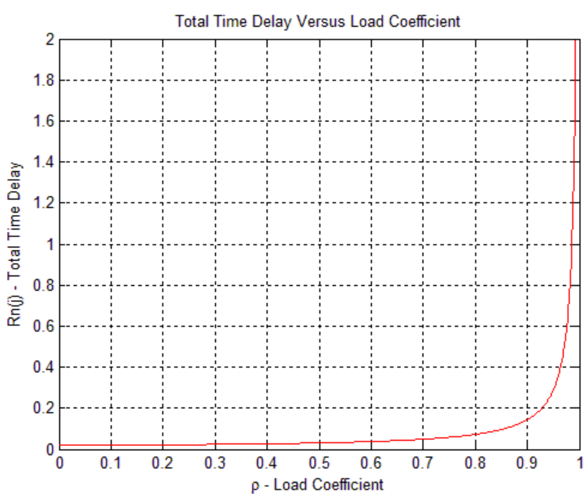

Figure 4. Total Time Delay Dependence on the System Load

The formula (1) gives relation between the probability p and the time quantum q. Parameters in formula (1) satisfy the inequalities $p \leq 1$ and $n > 1$. If with $p \rightarrow 1$ the prevailing requests in the system will be the ones with the service time longer than the time quantum q, which results in the increase of the average number of requests in the queue, the relation (31) is valid. From the point of view of the technical solution of the transmission system the queue length determines the capacity of the technical facility (most often the memory) for request recording. As resulting from the relation (27) the average number of requests in the queue Q influences the average stay time $R_n(j)$ of the requests in the system. As we can see in Fig. 2, the average time increases rapidly in the vicinity of the parameter values $p \rightarrow 1$ and $\rho \rightarrow 1$.

Acknowledgments

This work was supported by the VEGA grant no. 1/0427/15 of Ministry of Education of the Slovak Republic. The results presented in this paper are part of the solutions of this VEGA grant.

References

[1] L. Lakatos, L. Szeidl, M. Telek, “Introduction to Queueing Systems with Telecommunication Applications,” Springer Science and Business Media, 2013.

[2] G. Giambene, “Queuing Theory and Telecommunications,” Springer Science and Business Media, Inc., 2005.

[3] W. C. Chan, “Performance analysis of telecommunications and local area networks” Boston/Dordrecht/London: Kluwer Academic Publisher, 2000.

[4] F. Jeremiah, V. J. Thimma, “Modeling and analysis of telecommunications networks” John Wiley Publication, USA, 2004.

[5] V. Hottmar, B. Adamec, “Analytical model of weighted round robin service system” Hindawi publishing corporation, Journal of electrical and computer engineering, Volume 2012, ID 374961.

[6] B. Adamec, “Design of optimization model for data communication management in Canpus LAN,” Ph.D. dissertation, Dept. of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Slovakia, 2010.

[7] X Liu, K. Ravindran and D. Loguinov, “A Queuing-Theoretic Foundation of Available Bandwidth Estimation: Single-Hop Analysis,” IEEE/ACM Transactions on Networking, vol. 15, no. 4, pp. 918-931, August 2007.

[8] A. Orda and A. Sprintson, “Precomputation schemes for QoS routing,” IEEE/ACM Transactions on Networking, vol. 11, no. 4, pp. 578-591, August 2003.

Biographies

VLADIMIR HOTTMAR graduated from the University of Zilina in 1975 with the Master’s degree in Transport and Communications. At the beginning of his career he was employed at the Research Institute of Computers Techniques in Zilina, where he worked as a researcher in the Department of

ISSN:2319-7900

Microcomputers and Development Systems for 22 years, later being responsible for running the Research Department. Currently he is employed at the University of Zilina in the Department of Telecommunications and Multimedia in the position of an associate professor and is also involved in some research work. He is a project leader of the project VEGA. In 1999 he accomplished his PhD studies at the University of Zilina. Since May 2006 he has worked on the post of an associate professor at the University of Zilina.
vladimir.hottmar@fel.uniza.sk

BOHUMIL ADAMEC graduated from the University of Zilina in 2007 with the Master's degree in Telecommunications. Currently he is employed at the University of Zilina in the Department of Telecommunications and Multimedia in the position of an associate professor and is also involved in some research work. His range of interest is home networking especially simulation and mathematical modeling of modern multimedia home networks.
bohumil.adamec@fel.uniza.sk