

Performance Analysis and Evaluation of Digital Connection Oriented Internet Service Systems

Shunfu Jin¹ and Wuyi Yue²

¹ College of Information Science and Engineering
Yanshan University, Qinhuangdao 066004 China
jsf@ysu.edu.cn

² Department of Information Science and Systems Engineering
Konan University, Kobe 658-8501 Japan
yue@konan-u.ac.jp

Abstract. In this paper, we propose a discrete-time connection oriented Internet service system with a release delay, and present performance analysis and evaluation of the system. We build a batch arrival Geom*/G/1 queueing model with a setup/close-delay/close-down strategy to characterize the system operation, and suppose that the batch size is a random variable following Pareto(c, α) distribution. We describe the performance measures for setup ratio and utility of connection based on numerical results.

Keywords: Connection oriented service, burst, queueing system.

1 Introduction

Advancement in Internet services is urgently needed to satisfy service-specific Quality of Service (QoS) requirements. For example, different service classes offering packet traffic considered in terms of the required packet delay, response time and system throughput need to meet the real time and high transmission rate. For design and tuning of the advanced Internet services, system performance must be mathematically analyzed and numerically evaluated [1]. Queueing theory and Markov chains are used for the performance and reliability evaluation of communication networks.

In recent years, there have been many achievements in the research and application of continuous-time vacation queueing, especially in the performance evaluation of communication [2]. However, it is indicated that it would be more accurate and efficient to use discrete-time queue models rather than continuous counterparts when analyzing and designing digital transmitting systems [3]. The generally accepted view is that discrete-time systems can be more complex to analyze than equivalent continuous-time systems. The classical discrete-time queueing analyses can be found in [3], [4]. Analyses of discrete-time queue models with server vacation or setup strategy can be found in [5], [6].

Taking into account the memoryless character of a user initiated connection oriented session in a switched virtual channel, a delayed vacation Geom/G/1

queue model with setup was built. Some performance measures were then calculated in [7], [8].

In this paper, we propose a discrete-time connection oriented Internet service system with a release delay set before a release process, and present performance analysis and evaluation of the system. We also introduce an upper limit length of a release delay called timer length T , as a system parameter to control the length of the release delay. To give performance evaluation for this system, we build a batch arrival Geom*/G/1 queue model with a setup/close-delay/close-down strategy to characterize the system operation. We derive performance measures such as the setup ratio and utility of connection. In numerical results, we show that the setting of the timer length T is significant for improving system performance.

2 System Model

We assume the time axis to be divided into slots of equal length and batch arrivals to follow a Bernoulli process with a batch size following Pareto(c, α) distribution. Packets in batches arrived in the buffer having infinite capacity are transmitted with a common channel in a First Come First Serve (FCFS) discipline. The system works as follows. And this process will be repeated.

- (1) When a batch arrives in the system, a setup period called setup period U will be started, where the setup period U corresponds to a time period for setting up a new connection using a three-handshake signaling procedure.
- (2) After the setup period U finishes, a busy period called busy period θ will begin. The busy period θ is a time period in which packets are transmitted continuously until the buffer becomes empty.
- (3) When there are not any packets in the buffer to be transmitted, the system will enter a close-delay period called close-delay period D . The close-delay period D corresponds to a release delay time with an upper limit timer length T in slots ($0 \leq D \leq T$). During the close-delay period D , the connection is reserved in anticipation of more packets being transmitted using the same connection before the system goes into close-down phase. A close-delay period will finish either T is over or a batch arrives within T .
- (4) If there is a batch arrival within the close-delay period D , a new busy period θ will be started immediately without a setup period U . Otherwise, the system will enter a close-down period called close-down period C when the timer length T is over. C corresponds to the time period required to release the connection using another three-handshake signaling procedure.
- (5) If there is a batch arrival during the close-down period C , after C finishing, a setup period U will begin. If this does not occur, the system will enter an idle period called idle period I . A batch arriving during the idle period I will make the system to enter a new setup period U .

We define a transmission period B as being the time period in slots taken to transmit a single packet with generally distributed.

The setup period U , close-down period C and transmission period B are independent and identical discrete-time random variables in slots, and are assumed to be generally distributed with Probability Generation Functions (P.G.Fs.) $U(z)$, $C(z)$ and $B(z)$, respectively. Then $U(z)$, $C(z)$ and $B(z)$ are defined as follows:

$$\begin{aligned}
 u_k &= P\{U = k\}, \quad U(z) = \sum_{k=1}^{\infty} u_k z^k, & c_k &= P\{C = k\}, \quad C(z) = \sum_{k=1}^{\infty} c_k z^k, \\
 b_k &= P\{B = k\}, \quad B(z) = \sum_{k=1}^{\infty} b_k z^k. & & & & (1)
 \end{aligned}$$

Let $E[U]$, $E[C]$ and $E[B]$ be the means of U , C and B in slots, we have

$$E[U] = \sum_{k=1}^{\infty} k u_k, \quad E[C] = \sum_{k=1}^{\infty} k c_k, \quad E[B] = \sum_{k=1}^{\infty} k b_k.$$

To consider the batch arrivals in the system, we denote by Λ the number of packets in a batch called batch size Λ (packets/batch). Λ in this system is supposed to be a random variable. Let $E[\Lambda]$ be the mean of the batch size Λ . We can obtain the probability distribution, the P.G.F. $\Lambda(z)$ and $E[\Lambda]$ of Λ as

$$\lambda_k = P\{\Lambda = k\}, \quad k \geq 0, \quad \Lambda(z) = \sum_{k=0}^{\infty} \lambda_k z^k, \quad E[\Lambda] = \sum_{k=0}^{\infty} k \lambda_k \quad (2)$$

where λ_k is the probability that there are k packets in a batch per slot. Specially, $\lambda_0 = P\{\Lambda = 0\}$ is the probability that there is no batch ($\Lambda = 0$) arrival in a slot. We define the probability of no batch arrival during the timer length T to be $T(\lambda_0)$, where $T(\lambda_0) = \lambda_0^T$. From Eq. (1), we also know that the probability of no batch arrival during the close-down period C is $C(\lambda_0) = \lambda_0^C$ and the probability of no batch arrival during the transmission period B is $B(\lambda_0) = \lambda_0^B$. The ergodic condition is $\rho = E[\Lambda]E[B] < 1$.

Let A_U , A_C and A_B be random variables representing the numbers of packets arriving during U , C and B . We also define $\Lambda(B(z))$ to be the P.G.F. of the transmission time of a batch in slots. Then we can give $\Lambda(B(z))$ and the P.G.Fs. $A_U(z)$, $A_C(z)$ and $A_B(z)$ of A_U , A_C and A_B as follows:

$$\begin{aligned}
 \Lambda(B(z)) &= \sum_{k=0}^{\infty} \lambda_k (B(z))^k, & A_U(z) &= \sum_{k=1}^{\infty} u_k (\Lambda(z))^k = U(\Lambda(z)), \\
 A_C(z) &= \sum_{k=1}^{\infty} c_k (\Lambda(z))^k = C(\Lambda(z)), & A_B(z) &= \sum_{k=1}^{\infty} b_k (\Lambda(z))^k = B(\Lambda(z)). & (3)
 \end{aligned}$$

3 Performance Analysis

3.1 Queue Length

We assume that packet arrival and packet departure occur only at the boundary of a slot. Let $Q_n = Q(\tau_n^+)$ be the number of packets in the system immediately

after the n th packet departure. Then $\{Q_n, n \geq 1\}$ forms an imbedded Markov chain. We define the state of the system by the number Q of packets in the system at the imbedded Markov points as follows:

$$Q_{n+1} = \begin{cases} Q_n - 1 + A_B^{(n+1)}, & Q_n \geq 1 \\ \eta, & Q_n = 0 \end{cases} \tag{4}$$

where $A_B^{(n+1)}$ is the number of packets arriving during the transmission time of the $(n + 1)$ th packet. η is the number of packets left in the system after the departure of the first packet in a busy period Θ .

A busy period Θ begins with one of the following three cases:

- (1) If there is a batch arrival within the timer length T , the batch will trigger a busy period Θ immediately. The P.G.F. $\eta_1(z)$ for this case is given as follows:

$$\eta_1(z) = \frac{1}{z} \frac{\Lambda(z) - \lambda_0}{1 - \lambda_0} A_B(z) = \frac{1}{z} \frac{\Lambda(z) - \lambda_0}{1 - \lambda_0} B(\Lambda(z)).$$

- (2) If there is no batch arrival within either the timer length T or the close-down period C , the batch arriving during the idle period I will trigger a setup period U , and then a busy period Θ will begin. The P.G.F. $\eta_2(z)$ for this case is given as follows:

$$\eta_2(z) = \frac{1}{z} \frac{\Lambda(z) - \lambda_0}{1 - \lambda_0} A_U(z) A_B(z) = \frac{1}{z} \frac{\Lambda(z) - \lambda_0}{1 - \lambda_0} U(\Lambda(z)) B(\Lambda(z)).$$

- (3) If there is no batch arrival within the timer length T , but there is at least one batch arrival within the close-down period C , after the close-down period C finished, the system will directly enter a setup period U and then a busy period Θ will begin. The P.G.F. $\eta_3(z)$ for this case is given as follows:

$$\eta_3(z) = \frac{1}{z} \frac{A_C(z) - C(\lambda_0)}{1 - C(\lambda_0)} A_U(z) A_B(z) = \frac{1}{z} \frac{A_C(z) - C(\lambda_0)}{1 - C(\lambda_0)} U(\Lambda(z)) B(\Lambda(z)).$$

With these three cases, we can give the P.G.F. $\eta(z)$ of η as follows:

$$\eta(z) = \frac{1}{z} B(\Lambda(z)) \left(\frac{\Lambda(z) - \lambda_0}{1 - \lambda_0} (1 - T(\lambda_0) + C(\lambda_0) T(\lambda_0) U(\Lambda(z))) + T(\lambda_0) U(\Lambda(z)) (C(\Lambda(z)) - C(\lambda_0)) \right). \tag{5}$$

From Eq. (4), we can obtain the P.G.F. $Q(z)$ of Q as follows:

$$Q(z) = P\{Q \geq 1\} E[z^{Q-1+A_B} | Q \geq 1] + P\{Q = 0\} \eta(z). \tag{6}$$

Substituting Eq. (5) to Eq. (6), we can give that

$$Q(z) = P\{Q = 0\} \frac{B(\Lambda(z))}{B(\Lambda(z)) - z} \left(1 - \left(\frac{\Lambda(z) - \lambda_0}{1 - \lambda_0} (1 - T(\lambda_0) + C(\lambda_0) T(\lambda_0) U(\Lambda(z))) + T(\lambda_0) U(\Lambda(z)) (C(\Lambda(z)) - C(\lambda_0)) \right) \right). \tag{7}$$

Using the normalization condition and L'Hospital principle in Eq. (7), we have

$$P\{Q = 0\} = \frac{1 - \rho}{K} \tag{8}$$

where K is given as follows:

$$K = \frac{E[A]}{1 - \lambda_0} (1 - T(\lambda_0) + T(\lambda_0)C(\lambda_0) + E[A]T(\lambda_0) (E[U] + E[C])). \tag{9}$$

Substituting Eq. (8) to Eq. (7), then the P.G.F. $Q(z)$ and the mean $E[Q]$ of Q can be obtained as follows:

$$\begin{aligned} Q(z) &= \frac{(1 - \rho)(1 - \Lambda(z))B(\Lambda(z))}{E[A](B(\Lambda(z)) - z)} \left(\frac{E[A]}{(1 - \lambda_0)K} (1 - T(\lambda_0)) \frac{E[A]T(\lambda_0)C(\lambda_0)}{(1 - \lambda_0)K} \right. \\ &\quad \left. \cdot U(\Lambda(z)) + \frac{E[A](E[U] + E[C])T(\lambda_0)}{K} \cdot \frac{1 - U(\Lambda(z))C(\Lambda(z))}{(E[U] + E[C])(1 - \Lambda(z))} \right), \\ E[Q] &= \rho + \frac{A^{(2)} + (E[A])^3 B^{(2)}}{2(1 - \rho)E[A]} + \frac{E[A]T(\lambda_0)C(\lambda_0)}{(1 - \lambda_0)K} E[U] \\ &\quad + \frac{1}{2K} \left((E[A])^2 T(\lambda_0) (U^{(2)} + C^{(2)} + 2E[U]E[C]) \right). \end{aligned} \tag{10}$$

where $U^{(2)}$, $C^{(2)}$, $B^{(2)}$ and $A^{(2)}$ are the second factorial moments of the P.G.Fs. $U(z)$, $C(z)$ and $B(z)$ of the setup period U , close-down period C and transmission period B , and $A(z)$ of the batch size Λ by differentiating the P.G.Fs. presented in Eqs. (1) and (2) with respect to z and evaluating the result at $z = 1$.

3.2 Waiting Time

We focus on an arbitrary packet in the system called tagged packet M . We note that the waiting time W of the tagged packet M can be divided into two parts as follows. One is the waiting time W_b of the batch that the tagged packet M belongs to. The other is the total transmission time J of the packets before the tagged packet M in the same batch. W_b and J are independent random variables, so we have the P.G.F. $W(z)$ of the waiting time W of the tagged packet M as $W(z) = W_b(z)J(z)$. Where $W_b(z)$ and $J(z)$ are P.G.Fs. of W_b and J .

Applying the analysis of delayed vacation Geom/G/1 queueing model with setup in [7] and referencing [3], with $\Lambda(B(z))$ given in Eq. (3), we have

$$\begin{aligned} W_b(z) &= \frac{(1 - \rho)(1 - z)}{\Lambda(B(z)) - z} \left(\frac{E[A](1 - T(\lambda_0))}{(1 - \lambda_0)K} + \frac{E[A]T(\lambda_0)C(\lambda_0)}{(1 - \lambda_0)K} U(z) \right. \\ &\quad \left. + \frac{E[A](E[U] + E[C])T(\lambda_0)}{K} \cdot \frac{1 - U(z)C(z)}{(E[U] + E[C])(1 - z)} \right), \end{aligned} \tag{11}$$

$$J(z) = \frac{1 - \Lambda(B(z))}{E[A](1 - B(z))}. \tag{12}$$

Concluding Eqs. (11) and (12), the P.G.F. $W(z)$ and the mean $E[W]$ of W can be obtained as follows:

$$\begin{aligned}
 W(z) &= \frac{1 - \Lambda(B(z))}{E[A](1 - B(z))} \cdot \frac{(1 - \rho)(1 - z)}{\Lambda(B(z)) - z} \left(\frac{E[A]}{(1 - \lambda_0)K}(1 - T(\lambda_0)) \right. \\
 &\quad \left. + \frac{E[A]T(\lambda_0)C(\lambda_0)}{(1 - \lambda_0)K}U(z) + \frac{E[A]T(\lambda_0)(1 - U(z)C(z))}{K(1 - z)} \right), \\
 E[W] &= \frac{\Lambda^{(2)}(E[B])^2 + E[A]B^{(2)}}{2(1 - \rho)} + \frac{E[A]T(\lambda_0)C(\lambda_0)}{(1 - \lambda_0)K}E[U] \\
 &\quad + \frac{E[A]T(\lambda_0)(U^{(2)} + C^{(2)} + 2E[U]E[C])}{2K} + \frac{\Lambda^{(2)}E[B]}{2E[A]}. \tag{13}
 \end{aligned}$$

3.3 Busy Period

We define a busy cycle R as a time period from the instant in which a busy period Θ completes to the instant in which the next busy period Θ completes.

If there is a batch arrival within the timer length T , R is composed of the close-delay period D and Θ . If there are no batches arriving within T and there is at least one batch arriving within the close-down period C , R is composed of T , C , U , Θ . If there are not any batch arrivals either within T or within C , R is composed of T , C , U , Θ and the idle period I .

Denote by T_U , T_D , T_C , T_Θ , T_I and T_R the actual lengths of U , D , C , Θ , I and R in slots. It is obvious that

$$T_R = T_U + T_D + T_C + T_\Theta + T_I. \tag{14}$$

The event in which T_D equals the timer length T occurs with the probability $T(\lambda_0)$, event that T_D equals a conditional interval time occurring with the probability $1 - T(\lambda_0)$. So the P.G.F. $T_D(z)$ and the mean $E[T_D]$ of T_D are

$$T_D(z) = \frac{T(\lambda_0 z)(1 - z) + (1 - \lambda_0)z}{1 - \lambda_0 z}, \quad E[T_D] = \frac{1 - T(\lambda_0)}{1 - \lambda_0}. \tag{15}$$

We can also give the means $E[T_U]$, $E[T_C]$ and $E[T_I]$ as follows:

$$E[T_U] = T(\lambda_0)E[U], \quad E[T_C] = T(\lambda_0)E[C], \quad E[T_I] = \frac{1}{1 - \lambda_0}T(\lambda_0)C(\lambda_0). \tag{16}$$

Each packet of batches at the beginning of a busy period Θ will introduce a sub-busy period θ . All the sub-busy periods brought by the packets at the beginning of a busy period combine to make a busy period Θ in the system.

The P.G.F. $T_\Theta(z)$ and the mean $E[T_\Theta]$ of T_Θ are given as follows:

$$\begin{aligned}
 T_\Theta(z) &= (1 - T(\lambda_0)) \frac{\Lambda(\theta(z)) - \lambda_0}{1 - \lambda_0} + T(\lambda_0)C(\lambda_0) \frac{\Lambda(\theta(z)) - \lambda_0}{1 - \lambda_0} U(\Lambda(\theta(z))) \\
 &\quad + T(\lambda_0)(C(\Lambda(\theta(z))) - C(\lambda_0))U(\Lambda(\theta(z))), \\
 E[T_\Theta] &= K \frac{E[B]}{1 - \rho}. \tag{17}
 \end{aligned}$$

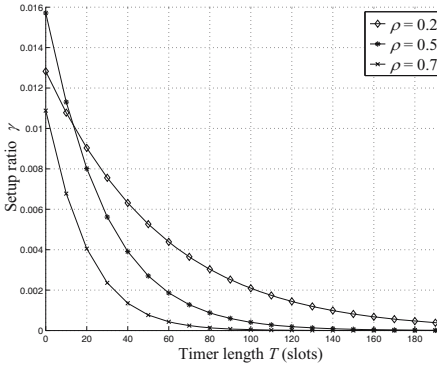


Fig. 1. Setup ratio for different offered load

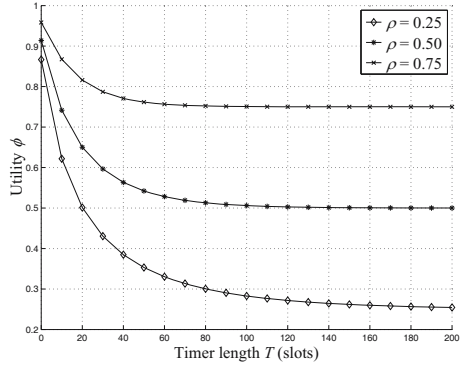


Fig. 2. Utility for different offered load

Concluding Eqs. (16) and (17), we can obtain the mean $E[T_R]$ as $E[T_R] = K/E[A](1 - \rho)$, where K is given in Eq. (9) and $E[A]$ is given in Eq. (2).

4 Performance Measures and Numerical Results

Setup ratio γ is defined as how many times the system goes to setup period U per slot, which is a measurement of the processing overhead in a connection oriented service. Obviously, the setup ratio γ is given by

$$\gamma = \frac{T(\lambda_0)}{E[T_R]} = \frac{E[A](1 - \rho)T(\lambda_0)}{K}.$$

Utility ϕ of connections is defined as the ratio of the time during which there are packets being transmitted to the time during which the network resource is occupied. Namely, the system is at one of the states of the busy period Θ , close-delay period D , or close-down period C . The utility ϕ is useful for the optimal design of the discrete-time connection oriented service system with a release delay. Clearly, the utility ϕ can be obtained as follows:

$$\phi = \frac{E[T_\Theta]}{E[T_\Theta] + E[T_D] + E[T_C]} = \frac{\rho K}{\rho K + (1 - \rho)(1 - T(\lambda_0) + E[A]T(\lambda_0)E[C])}.$$

In numerical results, we consider that $E[U] = 5$ (slots), $E[C] = 2$ (slots) and $E[B] = 5$ (slots). Considering the burst data shown in Internet traffic, we suppose the batch size A to be Pareto(c, α) distributed with $\lambda_k = ck^{-(\alpha+1)}, k = 0, 1, \dots$, where c is a normalization factor for $\sum_{k=1}^\infty \lambda_k = 1$, the parameter α is related to the Hurst factor H by $H = (3 - \alpha)/2, 0.5 < H < 1, 1 < \alpha < 2$. The smaller the result of α is, the more the burst shows in Internet traffic. Here, we suppose that $H = 0.85$, then $\alpha = 1.3$.

We show numerical results in the Figs. 1 and 2, and discuss the influence of the timer length T on the performance measures for the setup ratio γ and utility

ϕ of connection. In Figs. 1 and 2, the cases with the timer length $T = 0$ show the cases without considering the release delay in the systems as in [4]-[6].

Fig. 1 shows γ as a function of T with three different offered loads $\rho = 0.25, 0.50, 0.75$. We can find that when T increases, γ will decrease quickly to a low level (nearly zero). We also note that the larger the offered load ρ is, the larger the decrease of the setup ratio γ will be.

Fig. 2 shows the utility ϕ as a function of T with three different offered loads $\rho = 0.25, 0.50, 0.75$. When T increases, the utility ϕ will decrease quickly to another fixed value. It can also be found that, for the same timer length T , the larger the offer load ρ is, the larger the utility ϕ will be.

5 Conclusions

One way to reduce the cost of connection oriented service is to reduce the number of connection setups. We proposed a discrete-time connection oriented service system with a release delay in this paper to reduce the cost. We built a batch arrival $\text{Geom}^*/G/1$ queue model with a setup/close-delay/close-down strategy to characterize the system operation. We supposed the batch size to be $\text{Pareto}(c, \alpha)$ distributed to describe the burst data in Internet traffic. We presented performance measures such as the setup ratio and the utility of connection. We showed that the choice of the timer length T is significant in improving performance.

This paper has potential applications in network design, network maintenance and network management for the next generation Internet.

Acknowledgments. This work was supported in part by MEXT.ORB (2004-2008), Japan and in part by NSFC (No. 10671170) and MADIS, China.

References

1. Yue W. and Matsumoto Y.: Performance Analysis of Multi-Channel and Multi-Traffic on Wireless Communication Networks. Kluwer Academic Publishers (2002)
2. Niu Z. and Takahashi Y.: A Finite Capacity Queue with Exhaustive Vacation/Close-Down/Setup Times and Markovian Arrival Process. *Questa*. **31** (1999) 1-23
3. Takagi H.: Queueing Analysis. Vol. 3: Discrete-Time Systems. North-Holland (1993)
4. Tian N. and Zhang G.: Vacation Queueing Models-Theory and Applications. Springer-Verlag (2006)
5. Tian N. and Zhang G.: The Discrete Time GI/Geo/1 Queue With Multiple Vacations. *Queueing Systems*. **40** (2002) 283-294
6. Attahiri A.: Vacation Models in Discrete Time. *Queueing System*. **44** (2003) 5-30
7. Jin S. and Tian N.: Performance Evaluation of Virtual Channel Switching System Based on Discrete Time Queue. *Journal of China Institute of Communications*. **25** (2004) 58-68 (in Chinese)
8. Jin S., Yue W. and Liu M.: Queue Model and Performance Analysis for Discrete Time Switch Virtual Channels Systems. *Lecture Notes in Operation Research and its Applications* (2005) 26-36