Connection Oriented Internet Service for Burst Traffic With a Release Delay

Shunfu Jin[1] and Wuyi Yue[2]

[1]Yanshan University, 438 Av. Heibei, Qinhuangdao 066004 China
Phone: +86-335-8057078, Fax: +86-335-8057078, e-mail: jsf@ysu.edu.cn
[2]Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe 658-8501 Japan
Phone: +81-78-435-2530, Fax: +81-78-435-2530, e-mail: yue@konan-u.ac.jp

***Abstract*— Considering the burst traffic shown in some new Internet applications of nowadays, in this paper, we propose a slotted connection oriented Internet service system with a release delay having an upper limit timer length T. We build a batch arrival Geom*/G/1 queueing model with a setup/close-delay/close-down strategy to characterize the system operation. We suppose the batch size to follow Pareto(c, α) distributed. By using a discrete-time Markov chain approach, we derive the stationary distribution of the system and describe performance measures such as packet response time, setup ratio, and utility of connection. In numerical results, we evaluate the system performance and show that the setting of the timer length T is a significant issue for better performance of the system.**

I. Introduction

With the development of broadband, high-speed, high-capacity and high-reliability communication technologies, we are expecting new Internet applications such as voice, data, video, text, image/picture and instant messenger. Advanced Internet services are in urgent need to support these new Internet applications. At the same time, for design and tuning of the Internet services, system performance evaluation must be mathematically analyzed [1].

The generally accepted view is that vacation queueing models are much more suitable for connection oriented services. In recent years, there have been many achievements in the research and application of continuous-time vacation queueing systems, especially in the performance evaluation of communication networks [2].

However, it is indicated that it would be more accurate and efficient to use discrete-time queue models rather than continuous-time queue models when analyzing and designing digital communication systems [3]. Analyses of discrete-time queue models with server vacation or setup strategy can be found in [3]-[5].

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. Performance measures were given in [6].

However, burst data is found in many Internet applications, especially in the new Internet applications. It is well known that much additional processing overhead and time delay will be involved by setting up and releasing connections. If a connection is released immediately after a transmission finished, the connection for transmissions will be released and set up too frequently. Thus the system efficiency will be diminished.

In this paper, considering the burst shown in Internet traffic, we propose a discrete-time connection oriented Internet service system with a release delay having an upper limit timer length T. To evaluate the performance of this system, we build a batch arrival Geom*/G/1 queueing model with a setup/close-delay/close-down strategy.

We describe some key performance measures such as packet response time, setup ratio and utility of connection. Based on numerical results, we discuss the influence of the timer length T on the system performance for different offered loads and different burst degrees. We also show that the setting of the timer length T is a significant issue for better performance of the system.

II. Working Mechanism and System Model

Based on digital connection oriented Internet services, we set a release delay called close-delay period D before beginning a release process called close-down period C. And we introduce an upper limit length called timer length T as a system parameter for the close-delay period D. Batches arriving in the buffer having an infinite capacity are transmitted in a First In First Out (FIFO) discipline. Packets in each batch are transmitted at random by using a common channel. The time axis is divided into slots of fixed length.

When a batch arrives in the system, a 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. After the setup period U finishing, a busy period Θ will begin. The busy period Θ is a time period in which packets are transmitted continuously until the buffer becomes empty.

When there are not more packets in the buffer to be transmitted, the system will enter a close-delay period D. D corresponds to the release delay time with an upper limit timer length T $(0 \leq D \leq T)$. 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.

If there is a batch arrival within the close-delay period D, a new busy period Θ will be started immediately. Otherwise, the system will enter a 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.

Moreover, if there is a batch arrival during the close-down period C, after the close-down period C finishing, a setup period U will begin before the system enters the idle period I. If this does not occur, the system will enter an idle period

I. A batch arriving during the idle period I will make the system to enter a new setup period U.

The above process will be repeated.

Therefore, we can present a batch arrival Geom*/G/1 queueing model with a setup/close-delay/close-down strategy to characterize the system operation.

In this system model, we assume the batch arrival of packets to follow a Bernoulli process. The batch size is assumed to be a random variable following Pareto(c, α) distribution. We define a transmission period B to be a time period to transmit a single packet in slots. The transmission period B is assumed to be generally distributed.

III. Performance Analysis

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. The probability distributions and Probability Generation Functions (P.G.Fs.) $U(z)$, $C(z)$ and $B(z)$ of U, C and B are defined as follows:

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

Let $E[U]$, $E[C]$ and $E[B]$ be the averages of the setup period U, close-down period C and transmission period B in slots, respectively, then we can obtain them as follows:

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

We denote by Λ the number of packets in a batch, and call Λ the batch size. The probability distribution and P.G.F. $\Lambda(z)$ of the batch size Λ can be written as follows:

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

where λ_k is the probability that the number of packets in a batch is k. Specially, $\lambda_0 = P\{\Lambda = 0\}$ is the probability corresponding to that there is no batch ($\Lambda = 0$) arrival in a slot.

Let $E[\Lambda]$ be the average of the batch size Λ in packets, then we can give that

$$E[\Lambda] = \sum_{k=0}^{\infty} k \lambda_k.$$

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 know that the probability of no batch arrival during close-down period C is $C(\lambda_0)$ and the probability of no batch arrival during transmission period B is $B(\lambda_0)$.

Obviously, when $\rho = E[\Lambda]E[B] < 1$, the system will arrive at a state of equilibrium. Now, we present a performance analysis of the system in the state of equilibrium.

A. Number of Packets in the System

For a late arrival system, batch arrivals in the τth slot is assumed to occur only at the instant τ^-, while packet departure is assumed to occur at the instant τ. Let $Q_n = Q(\tau_n^+)$ be the number of packets in the system immediately after the nth 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. The formula below is tenable:

$$Q_{n+1} = \begin{cases} Q_n - 1 + A_B^{(n+1)}, & Q_n \geq 1 \\ \eta, & Q_n = 0 \end{cases} \quad (3)$$

where $A_B^{(n+1)}$ is the number of packets arriving during the transmission time B 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) There is a batch arrival within the timer length T. The batch will trigger a busy period Θ immediately. (2) There is no batch arrival within either the timer length T or the close-down period C, then the batch arriving during the idle period I will trigger a setup period U, and a busy period Θ will begin. (3) There is no batch arrival within the timer length T, but there is at least one batch arrival within the close-down period C. In this case, after the close-down period C is over, the system will directly enter a setup period U and then a busy period Θ will begin. For these, the P.G.F. $\eta(z)$ of η can be given as follows:

$$\begin{aligned} \eta(z) \ =\ & \frac{1}{z} B(\Lambda(z)) \Big(\frac{\Lambda(z) - \lambda_0}{1 - \lambda_0} (1 - T(\lambda_0) \\ & + C(\lambda_0) T(\lambda_0) U(\Lambda(z))) \\ & + T(\lambda_0) U(\Lambda(z)) (C(\Lambda(z)) - C(\lambda_0)) \Big). \end{aligned} \quad (4)$$

Substituting Eq. (4) into Eq. (3), using the normalization condition and L'Hospital principle, we can give P.G.F. $Q(z)$ of Q as follows:

$$\begin{aligned} Q(z) = & \frac{(1-\rho)(1-\Lambda(z))B(\Lambda(z))}{E[\Lambda](B(\Lambda(z)) - z)} \Big(\frac{E[\Lambda]}{(1-\lambda_0)K} \\ & \cdot (1 - T(\lambda_0)) + \frac{E[\Lambda] T(\lambda_0) C(\lambda_0)}{(1-\lambda_0)K} U(\Lambda(z)) \\ & + \frac{E[\Lambda](E[U] + E[C]) T(\lambda_0)}{K} \\ & \cdot \frac{1 - U(\Lambda(z)) C(\Lambda(z))}{(E[U] + E[C])(1 - \Lambda(z))} \Big) \end{aligned} \quad (5)$$

where K is given as follows:

$$\begin{aligned} K = & \frac{E[\Lambda]}{1 - \lambda_0} \big(1 - T(\lambda_0) + T(\lambda_0) C(\lambda_0) \\ & + E[\Lambda] T(\lambda_0) (E[U] + E[C]) \big). \end{aligned}$$

From Eq. (5), we can give the average $E[Q]$ of Q as follows:

$$E[Q] = \rho + \frac{\Lambda^{(2)} + (E[\Lambda])^3 B^{(2)}}{2(1-\rho)E[\Lambda]} + \frac{E[\Lambda]T(\lambda_0)C(\lambda_0)}{(1-\lambda_0)K}E[U] + \frac{1}{2K}\left((E[\Lambda])^2 T(\lambda_0)\left(U^{(2)} + C^{(2)} + 2E[U]E[C]\right)\right)$$

where $U^{(2)}$, $C^{(2)}$ and $B^{(2)}$ are the second factorial moments of the P.G.Fs. $U(z)$, $C(z)$ and $B(z)$ in slots, respectively, obtained from Eq. (1). $\Lambda^{(2)}$ is the second factorial moment of the P.G.F. $\Lambda(z)$ in packets obtained from Eq. (2).

B. Waiting Time of Packets

We focus on an arbitrary packet in the system called a tagged packet M. We note that the waiting time W of M can be divided into two parts as follows: One is the waiting time W_b of the batch that M belongs to. The other is the total transmission time J of the packets before 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 follows:

$$W(z) = W_b(z)J(z) \tag{6}$$

where $W_b(z)$ and $J(z)$ are P.G.Fs. of W_b and J.

Applying the analysis formula in [6], we have that

$$\begin{aligned} W_b(z) =& \frac{(1-\rho)(1-z)}{\Lambda(B(z))-z}\Bigg(\frac{E[\Lambda](1-T(\lambda_0))}{(1-\lambda_0)K} \\ &+ \frac{E[\Lambda]T(\lambda_0)C(\lambda_0)}{(1-\lambda_0)K}U(z) \\ &+ \frac{E[\Lambda]\left(E[U]+E[C]\right)T(\lambda_0)}{K} \\ &\cdot \frac{1-U(z)C(z)}{\left(E[U]+E[C]\right)(1-z)}\Bigg). \end{aligned} \tag{7}$$

$J(z)$ can be written as follows (see [3]):

$$J(z) = \frac{1-\Lambda(B(z))}{E[\Lambda](1-B(z))}. \tag{8}$$

Substitute Eqs. (7) and (8) to Eq. (6), we obtain that

$$\begin{aligned} W(z) =& \frac{1-\Lambda(B(z))}{E[\Lambda](1-B(z))} \cdot \frac{(1-\rho)(1-z)}{\Lambda(B(z))-z} \\ &\cdot \Bigg(\frac{E[\Lambda]}{(1-\lambda_0)K}(1-T(\lambda_0)) \\ &+ \frac{E[\Lambda]T(\lambda_0)C(\lambda_0)}{(1-\lambda_0)K}U(z) \\ &+ \frac{E[\Lambda]\left(E[U]+E[C]\right)T(\lambda_0)}{K} \\ &\cdot \frac{1-U(z)C(z)}{\left(E[U]E[C]\right)(1-z)}\Bigg). \end{aligned} \tag{9}$$

From Eq. (9), we can give the average waiting time $E[W]$ of packets in slots as follows:

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

C. Busy Period of the System

We denote the averages of the actual lengths of the setup period U, close-delay period D, close-down period C, idle period I and busy period Θ by $E[T_U]$, $E[T_D]$, $E[T_C]$, $E[T_I]$ and $E[T_\Theta]$. So we have that

$$\begin{aligned} E[T_U] &= T(\lambda_0)E[U], \quad E[T_D] = \frac{1-T(\lambda_0)}{1-\lambda_0}, \\ E[T_C] &= T(\lambda_0)E[C], \quad E[T_I] = \frac{1}{1-\lambda_0}T(\lambda_0)C(\lambda_0), \\ E[T_\Theta] &= K\frac{E[B]}{1-\rho}. \end{aligned} \tag{10}$$

Next, we define a busy cycle R as a time period from the instant that the system completes a busy period to the instant that the next busy period ends. We denote by $E[T_R]$ the average of the actual lengths T_R of busy cycle R as

$$E[T_R] = E[T_U] + E[T_D] + E[T_C] + E[T_I] + E[T_\Theta]. \tag{11}$$

Substituting Eq. (10) to Eq. (11), we have that

$$E[T_R] = \frac{K}{E[\Lambda](1-\rho)}.$$

IV. Performance Measures

We define a packet response time T^* to be the time period in slots elapsed from the arrival of a packet to the end of the transmission of this packet. The average $E[T^*]$ of T^* can be given as follows:

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

Setup ratio γ is defined as how many times the system goes to setup period U per slot. Obviously, the setup ratio γ can be given by

$$\gamma = \frac{T(\lambda_0)}{E[T_R]} = \frac{E[\Lambda](1-\rho)T(\lambda_0)}{K}. \tag{12}$$

Idle ratio β is measured as the fraction of time that an open connection is not used on average as follows:

$$\begin{aligned} \beta &= \frac{E[T_D]}{E[T_\Theta] + E[T_D] + E[T_C]} \\ &= \frac{(1-\rho)(1-T(\lambda_0))}{\rho K + (1-\rho)(1-T(\lambda_0) + E[\Lambda]T(\lambda_0)E[C])}. \end{aligned} \tag{13}$$

System utility ϕ 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. Clearly, the system utility ϕ can be obtained as follows:

$$\begin{aligned} \phi &= \frac{E[T_\Theta]}{E[T_\Theta] + E[T_D] + E[T_C]} \\ &= \frac{\rho K}{\rho K + (1-\rho)(1-T(\lambda_0) + E[\Lambda]T(\lambda_0)E[C])}. \end{aligned} \tag{14}$$

V. NUMERICAL RESULTS

In numerical results, we assume the setup period U, close-down period C and transmission period B to be generally distributed with average values $E[U] = 5$ (slots), $E[C] = 2$ (slots) and $E[B] = 5$ (slots). We assume the batch size Λ to be Pareto(c, α) distributed as follows:

$$\lambda_k = ck^{-(\alpha+1)}, \quad k = 1, 2, \cdots, \quad 1 < \alpha < 2$$

where c is a normalization factor for $\sum_{k=1}^{\infty} \lambda_k = 1$, α is related to the Hurst factor H by

$$H = (3 - \alpha)/2, \quad 1 < \alpha < 2, \quad 0.5 < H < 1.$$

It is obvious that the smaller the parameter α in the Pareto distribution is, the larger the Hurst factor H is, and the more the burst will be.

Due to a finite first moment and an infinite second moment of a Pareto(c, α) distributed random variable, the average response time $E[T^*]$ of the packets tends to be infinite.

We present numerical examples in terms of the setup ratio γ, idle ratio β and system utility ϕ as functions of the timer length T in slots for two special cases of lower and higher offered loads with $\rho = 0.25, 0.75$, and for Pareto distribution parameters $\alpha = 1.5, 1.3, 1.1$ in Tables I and II.

TABLE I

Numerical results with $\rho = 0.25$.

α	T	γ	β	ϕ
1.5	0	0.018246	0	0.73264
	10	0.0142370	0.33909	0.51443
	40	0.0066664	0.62686	0.32924
	70	0.0030825	0.70135	0.28131
1.3	0	0.013486	0	0.78757
	10	0.0112410	0.28972	0.57991
	40	0.0064686	0.58023	0.37168
	70	0.0036999	0.66964	0.30760
1.1	0	0.0077123	0	0.86637
	10	0.0069541	0.20568	0.69734
	40	0.0050919	0.48054	0.47145
	70	0.0037231	0.59087	0.38078

In Tables I and II, the cases with the timer length $T = 0$ show the cases without considering the release delay in the systems as in previous papers [4] and [5].

By comparing the systems presented in the literature, from Tables I and II, we can know that for the same offered load ρ and the same parameter α in Pareto distribution, the longer the timer length T is, the lower the setup ratio γ will be. However, a longer timer length T will make the idle ratio β larger and the system utility ϕ lower. So, there is a trade-off between the setup ratio γ and idle ratio β or utility ϕ when the timer length T is set.

Tables I and II also show that with the same parameter α, the heavier the traffic (with larger result of ρ) is, the stronger the influence of the timer length T on the setup ratio γ will be.

TABLE II

Numerical results with $\rho = 0.75$.

α	T	γ	β	ϕ
1.5	0	0.015958	0	0.90384
	10	0.0079011	0.13784	0.81902
	40	0.0007692	0.23992	0.75620
	70	6.9564e-005	0.24910	0.75056
1.3	0	0.012291	0	0.92427
	10	0.0072946	0.11443	0.84450
	40	0.0013719	0.22687	0.76612
	70	0.0002451	0.24594	0.75283
1.1	0	0.0073486	0	0.9533
	10	0.0054351	0.077289	0.89045
	40	0.0021450	0.18878	0.79978
	70	0.0008305	0.22722	0.76852

Moreover, in the same case, the larger the offered load ρ is, the lower the setup ratio γ and idle ratio β will be, and the larger the system utility ϕ will be. The reason is because the larger the offered load ρ is, the larger the probability that batches arrived during a close-delay D period will be. So a busy period will begin again without a setup process.

From these discussions, we can conclude that when we set a timer length T, the degree of burst and the offered load must be considered.

VI. CONCLUSIONS

We proposed a slotted connection oriented Internet service system with a release delay, and built a batch arrival Geom*/G/1 queueing model with a setup/close-delay/close-down strategy to characterize the system operation. We derived the stationary distribution of the system, described some key performance measures, and showed that the choice of the timer length T is a significant issue for better performance. This paper provides theoretical basis and performance evaluation for the slotted connection oriented Internet service systems with the release delay, and has potential applications for the network design, network maintenance and network management with connection oriented service in the next generation Internet.

REFERENCES

[1] W. Yue and Y. Matsumoto, *Performance Analysis of Multi-Channel and Multi-Traffic on Wireless Communication Networks*, Kluwer Academic Publishers, 2002.

[2] Z. Niu and Y. Takahashi, "A Finite Capacity Queue with Exhaustive Vacation/Close-Down/Setup Times and Markovian Arrival Processed," *Questa*, Vol. 31, pp. 1-23, 1999.

[3] H. Takagi, *Queueing Analysis* (*Volume* 3: *Discrete-Time Systems*), North-Holland, 1993.

[4] N. Tian and G. Zhang, *Vacation Queueing Models-Theory and Applications*, Springer-Verlag, 2006.

[5] A. S. Alfa, "Vacation Models in Discrete Time," *Queueing System*, Vol. 44, pp. 5-30, 2003.

[6] S. Jin, W. Yue and M. Liu, "Performance Analysis of SVC Based on Discrete Time Factorization Principle with General Vacations," *Technical Report of IEICE*, Vol. 102, pp. 1-6, 2005.