ISSN 2186-5647

2-89

Using GE/G/1 modelling to analyse performance of priority queues with bulk arrivals in LER system

China University of Technology　○Yunlung CHEN

1.Introduction

In this paper, a stable *G/G/1* queue with four priority classes in ATM-based MPLS core network under HOL scheduling disciplines is of practical interest. Under our proposal, the arrival process at each queue has been assumed bursty and is modeled by CPP with geometrically distributed bulk (batch) sizes. The mean queue lengths and the mean queueing delays can be obtained through the non-preemptive *GE/G/1* HOL priority queues, incoming arrivals have to consider bursty traffic. We endeavor to obtain the boundary conditions of the mean queue lengths and the mean queueing delays for each priority class, since this metric is a crucial factor for improving QoS and system performance of the LER system in ATM-based MPLS network.

Furthermore, we find the values of maximum allowed burst size for EF flow and AF1 flow and perform necessary policing/shaping actions on EF flow and AF 1 flow at the boundary node of the network. To analyze performance of EF flow, AF1 flow, F2 flow, and BE flow, the optimum ratios of COV (Coefficient of Variation) are carried out using HOL (Head of Line) priority rules. However, the ratios of COV value have to work within the existing constraints. This paper presents the four-class priority queues and performance analysis of the LER system are shown accurate and robust after the comparison between theoretical evaluates and computer simulation results. Note that the term “mean queueing delay” refers to the mean time spent of messages waiting in the queue.

2. Label Edge Router (LER)

In this paper, we **propose** the following service classes (Table 1): Guaranteed (hard QoS) service and EF flow map to the Gold class. AF 1 flow and AF 2 flow to Silver and Bronze classes respectively. Fig. 1 illustrates the IPS (Integrated Packet Scheduler) in an ingress MPLS LER system which is fed by a large number of input packet flows. These incoming packets are mapped with five service class flows at LER: Guaranteed flow for a hard QoS, EF flow, AF 1 flow, AF 2 flow, and BE flow. The ingress LER has an admission control and IPS function. Each flow is propagated to the IPS by admission control.

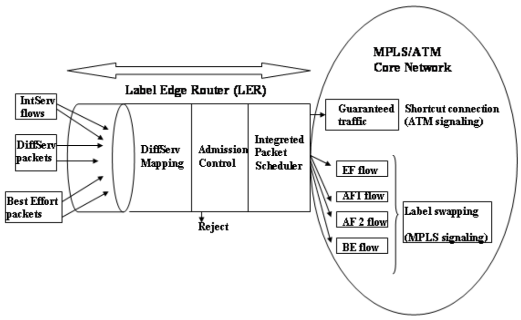

Fig. 1.　Label Edge Router (LER) system

Table 1. Service classes map from IntServ/DiffServ

Incoming service flow requirements: IntServ	DiffServ	ATM/MPLS mapping service classes
Guaranteed Service	Expedited Forwarding (EF)	Gold
—	Assured Forwarding (AF)	Silver/Bronze
—	Best Effort (BE)	Best Effort

3.Derived Results of Mean Queue Length using GE Distribution

The performance analysis of the LER system is **analyzed** by non-preemptive *GE/G/1* HOL priority queue with four-class of jobs having been established, where arrivals are assumed to be of GE distribution with mean rate λ_i and the service rate μ_i which is slightly higher than λ_i, i =1, 2, 3, 4. In this case, from Little's law, we have

$$\lambda_i = \mu_i \times \rho_i.$$

The traffic intensity is combined with ρ_1 for EF flow, ρ_2 for AF 1 flow, ρ_3 for AF 2 flow and ρ_4 for BE flow

$$\rho = \sum_{i=1}^{4} \rho_i = \sum_{i=1}^{4} \frac{\lambda_i}{\mu_i}$$

,

$$0 < \rho = \rho_1 + \rho_2 + \rho_3 + \rho_4 < 1.$$

According to theorem 3 of [1]: For a stable *GE/G/1* queue with R (≥2) priority classes under HOL scheduling disciplines, the exact marginal mean queue lengths L_i, i =1, 2,…R, are given by the closed-form expressions. In order to expedite the analytic implementation of the ME solutions, the exact marginal mean queue lengths L_i, i =1, 2, 3, 4. A mathematical model of the *GE/G/1* HOL queue with four-class non-preemptive priority assignment is proposed. The mathematical derivation of the mean queue lengths L_i, i =1, 2, 3, 4 of a *GE/G/1* HOL priority queue may be obtained from the above-mentioned theorem 3. The mathematical closed-form expressions can be derived as equation (6), (7), (8), and (9) . Let L_1, L_2, L_3 and L_4 be the mean queue length of EF flow, AF 1 flow, AF 2 flow and BE flow respectively. Thus, the mean queue length per class in the MPLS LER system can be obtained from Eqs. (6)-(9). In the case of the non-preemptive *GE/G/1* HOL queue it can be easily verified that Kleinrock's conservation law holds [2]. However, Eqs. (6)-(9) for *GE/G/1* HOL queues are strictly exact in a stochastic sense when C_{ai}^2 and $C_{si}^2 \geq 1$ [1, 4].

$$L_1 = \rho_1 + \frac{\rho_1}{2(1-\rho_1)}\left(C_{a1}^2 - 1\right) + \frac{1}{2(1-\rho_1)}\left(\alpha_1 + \beta_1\right). \quad (6)$$

$$\alpha_1 = \rho_1^2 (C_{s1}^2 + 1)$$

$$\beta_1 = \frac{\lambda_1}{\lambda_2}\rho_2^2\left(C_{s2}^2 + 1\right) + \frac{\lambda_1}{\lambda_3}\rho_3^2\left(C_{s3}^2 + 1\right) + \frac{\lambda_1}{\lambda_4}\rho_4^2\left(C_{s4}^2 + 1\right)$$

$$L_2 = \rho_2 + \frac{\rho_2}{2(1-\rho_1-\rho_2)}\left(C_{a2}^2 - 1\right) + \frac{1}{2(1-\rho_1)(1-\rho_1-\rho_2)}\left(\alpha_2 + \beta_2\right).$$

$$\alpha_2 = \rho_2^2 (C_{s2}^2 + 1) + \frac{\lambda_2}{\lambda_1}\rho_1^2 (C_{s1}^2 + C_{a1}^2) \quad (7)$$

$$\beta_2 = \frac{\lambda_2}{\lambda_3}\rho_3^2 (C_{s3}^2 + 1) + \frac{\lambda_2}{\lambda_4}\rho_4^2 (C_{s4}^2 + 1)$$

$$L_3 = \rho_3 + \frac{\rho_3}{2(1-\rho_1-\rho_2-\rho_3)}(C_{a3}^2 - 1) + \frac{1}{2(1-\rho_1-\rho_2)(1-\rho_1-\rho_2-\rho_3)}\left(\alpha_3 + \beta_3\right).$$

$$\alpha_3 = \rho_3^2 (C_{s3}^2 + 1) + \frac{\lambda_3}{\lambda_1}\rho_1^2\left(C_{s1}^2 + C_{a1}^2\right) + \frac{\lambda_3}{\lambda_2}\rho_2^2\left(C_{s2}^2 + C_{a2}^2\right)$$

$$\beta_3 = \frac{\lambda_3}{\lambda_4}\rho_4^2\left(C_{s4}^2 + 1\right) \quad (8)$$

$$L_4 = \rho_4 + \frac{\rho_4}{2(1-\rho_1-\rho_2-\rho_3-\rho_4)}(C_{a4}^2 - 1) + \frac{1}{2(1-\rho_1-\rho_2-\rho_3)(1-\rho_1-\rho_2-\rho_3-\rho_4)}\left(\alpha_4 + \beta_4\right).$$

$$.\beta_4 = 0 \quad (9)$$

$$\alpha_4 = \rho_4^2 (C_{s4}^2 + 1) + \frac{\lambda_4}{\lambda_1}\rho_1^2\left(C_{s1}^2 + C_{a1}^2\right) + \frac{\lambda_4}{\lambda_2}\rho_2^2\left(C_{s2}^2 + C_{a2}^2\right) + \frac{\lambda_4}{\lambda_3}\rho_3^2\left(C_{s3}^2 + C_{a3}^2\right)$$

4.Derived Results of Queueing Delay using GE Distribution

When traffic from external sources **reaches** the LER system, it is queued and transmitted to the LER system at the LER interface speed, resulting in rapid queue build-up, followed by queue emptying. These queueing delays can be on the order of microseconds to milliseconds (or even hundreds) in practice. The level of acceptable dynamic congestion depends on the application. For example, queueing delay exceeding approximately 30 ms for VoIP can degrade quality, whereas data applications can typically tolerate up to 500ms of queueing delay. In this section, we derive the mean queueing delay from Little's formula. We can then infer $L_i = \lambda_i W_i$, where i = 1, 2, 3, and 4. Let W_1, W_2, W_3 and W_4 be the mean queueing delay of EF flow, AF 1 flow, AF 2 flow and BE flow respectively for four-class *GE/G/1* HOL priority queues. Thus we can derive the general result for the mean queueing delay of each i class. From Eqs (6)-(9), W_1, W_2, W_3 and W_4 can be expressed as Eqs (10)-(13) respectively.

$$W_1 = \frac{L_1}{\lambda_1} \quad (10) \qquad W_2 = \frac{L_2}{\lambda_2} \quad (11)$$

$$W_2 = \frac{L_2}{\lambda_2} \quad (12) \qquad W_4 = \frac{L_4}{\lambda_4} \quad (13)$$

5.Computer Simulations and Numerical Analysis

The traffic model of edge DiffServ is proposed under the assumption made according to the guidelines of RFC 2598 definition for EF flows._Therefore, the parameters of proposed traffic model of incoming flows with four service classes at LER system are listed in Table 2. The average message lengths are 160 byte packets for VoIP class (i.e. Gold class), 300 byte packets for Latency-Optimized class (i.e. Silver class), 500 byte packets for Throughput-Optimized class (i.e. Bronze class), and 1500 byte packets for BE class. The performance of priority queues of the LER system is analysed the non-preemptive *GE/G/1* HOL queueing bulk arrival model. In addition, packet loss occurs when one or more packets of flow travelling across the ATM-based MPLS network fail to reach their destination. In this paper, high-quality VoIP services typically target a loss rate of 0.1 percent or less [3]. Packet loss in the Silver class as well as in Gold class is also kept to a minimum, and at a loss rate of typically ≤ 0.1 percent. We assume that the loss rate is 0.25 percent for Bronze class and 0.4 percent for BE class. The error control of each incoming flow is also presented in order to obtain more realistic simulation results. We propose an SR **(**Selective Repeat) method to balance potentially loss found in data for efficient and reliable flow of incoming packets in ATM-based MPLS network. The errors of incoming flow are generated randomly, whose influence can be observed on simulation results of queue length as well as queueing delay of **e**ach incoming flow for non-preemptive priority queue. The error probability of each incoming flow is denoted by P_e. For example, P_e= 0.000 means error free, the probability of P_e= 0.001 means the probability of 1 packet error occurred for each 1000 packets transmitted, P_e= 0.0025 means the probability of 2.5 packet error occurred for each 1000 packets transmitted, P_e= 0.004 means the probability of 4 packet error occurred for each 1000 packets transmitted.

For computer simulation, we record the message numbers of EF flow, AF 1 flow, AF 2 flow, and BE flow every 50 seconds. 30 simulations are executed to obtain the numerical results in Fig. 2.

Table 2. Parameters of proposed traffic model of each incoming flow

Priority type of message	EF flow	AF 1 flow	AF 2 flow	BE flow
Average packet length (bytes)	160	300	500	1500
Loss rate (%)	0.1	0.1	0.25	0.4

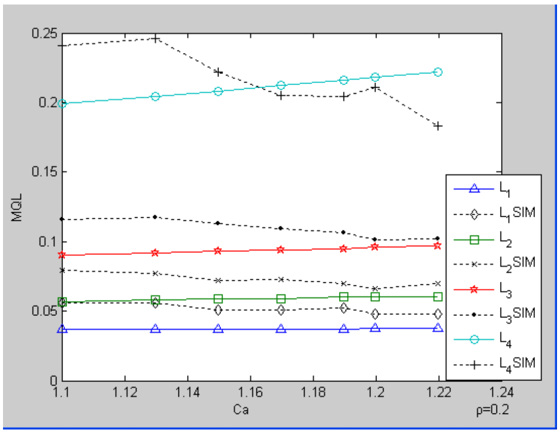

Fig. 2 Comparisons of EF flow, AF 1 flow, AF 2 flow and BE flow for MQL and MQD respectively (C_{si}=1.8, ρ = 0.2 and P_e= 0.000)

6.Conclusions

The analytic and simulation results show that the non-preemptive priority model provides reliable estimates for individual mean queue lengths of each priority class in the LER system. The analysis **discussed above** is true if the condition $C_{ai}^2 > C_{si}^2 > 1$ or $C_{si}^2 > C_{ai}^2 > 1$, i=1, 2, 3, 4 is satisfied by the selected network parameters. **It** is also possible to simulate analytical expressions for the mean queue length for cases such as $C_{a1}^2 > C_{a2}^2 > C_{a3}^2 > C_{a4}^2$ and $C_{s1}^2 > C_{s2}^2 > C_{s3}^2 > C_{s4}^2$; or $C_{a1}^2 > C_{a2}^2 < C_{a3}^2 > C_{a4}^2$ and $C_{s1}^2 < C_{s2}^2 > C_{s3}^2 > C_{s4}^2$. As memtioned above, the simulation results show that the estimates are demonstrated to be reliable even if the entire network load up has reached saturation. In addition, the higher priority class traffic is coherently served at the low mean queue length, regardless of the level of traffic intensity, thus the effectiveness of differential services is assisted for the high priority users. In contrast, the lower priority traffic classes are easily influenced by traffic intensity. We have found that when we calculate the metric of a certain priority class, all the higher classes can be massed together and all the lower classes can be overlooked. Furthermore, when traffic intensity ρ increases, the priority class assignment of the traffic packets have a significant_effect on the mean queue length per class, whereas when ρ is equal to or smaller than 0.4, the priority of the traffic packets does not have much of a difference in the mean queue length. For instance, when ρ slowly increases from 0.4 to 0.95, the higher priority packets experience a greater decrease than the lower priority packets do in the mean queue length, which support for the advantage of differential services in MPLS network. It is clear that as the traffic intensity increases the benefit in deploying priority service discipline increases as well. Using Little's theorem to these results will allow the mean queueing delay to be achieved as well.

References

[1] Demetres Kouvatsos and Nasreddine Tabet-Aouvel. A Maximum Entropy Priority Approximation for a Stable *G/G/1* Queue. Acta Informatica 27, 1989, pp. 247-286.

[2] Demetres Kouvatsos and Irfan Awan. Entropy maximization and open queueing network with priorities and blocking. Performance Evaluation 2003, pp. 191-227.

[3] B. Davie, et al. An Expedited Forwarding PHB (Per-Hop Behavior). RFC 3246, March 2002.

[4] Demetres Kouvatsos. Entropy maximization and queueing network models. Annals of Operation Research 1994, pp. 63-126