ELSEVIER

Computer Communications 24 (2001) 1626–1636

computer communications

www.elsevier.com/locate/comcom

Modeling and performance analysis for IPv6 traffic with multiple QoS classes

Liren Zhang*, Li Zheng

Network Technology Research Centre, School of Electrical and Electronic Engineering, Nanyang Technological University, Block S2, Singapore, Singapore 639798

Received 5 September 2000; revised 10 January 2001; accepted 10 January 2001

Abstract

This paper focuses on the modeling and performance analysis for IPv6 traffic with multi-class QoS in virtual private networks (VPN). The multi-class QoS is implemented on differentiated service basis using priority scheme of 4 bits defined in the packet header of IPv6. A VPN-enabled IP router is modeled as a tandem queuing system in which each output link consists of two parallel priority output queues. The high-priority queue is used to carry the delay sensitive traffic while the low-priority queue is used to carry the delay insensitive traffic. On the other hand, multiple thresholds are implemented in each queue, respectively, for packet loss priority control. The performance analysis is done using fluid flow techniques. The numerical results obtained from the analysis show that the differentiated service based on the priority schemes defined in IPv6 is able to effectively satisfy the multi-class QoS requirement for supporting multimedia services in VPN. The performance trade-off between the delay sensitive traffic and delay insensitive traffic in terms of traffic throughput, packet loss probability and end-to-end delay in VPN networks is presented. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Virtual private network; Quality of service; IPv6; Multimedia networking

1. Introduction

Currently, there is a significant interest in the development of virtual private network (VPN) over IP backbone. VPN is an enterprise network based on shared public network infrastructure but employing the same security, management and throughput policies as applied in a private network [13]. Comparing with the existing private networks, VPN is a more cost-effective mean of building and deploying private communication networks for multi-site communication, especially when IPv6 over broadband Internet is implemented. VPN is also able to support multi-media services such as voice, video, data and image transfer applications. For an IP-based VPN [16], the service provider connects multiple IP addresses located at geographically dispersed sites as appearing to be within a private network. As shown in Fig. 1, VPN can be implemented using VPN-enabled router [13] which plays the network layer functions in the TCP/IP protocol suite to support network security, network routing connectivity and QoS parameters.

One of the typical problems with the implementation of VPN over Internet is the difficulty of QoS guarantee. The Internet Engineering Task Force (IETF) has recommended a differentiated service mechanism for the traffic with different QoS requirement on priority basis [17,18]. IPv6 [2] is ideally designed for supporting such differentiated services [17]. However, in current IP-based networks, when the data are encrypted, it may be difficult for the network to determine the class-of-service based on packet content in the network layer, especially when a multiple class of QoS is involved. By contrast, this can be done in IP-based VPN which has the advantage that the class of service can be stated outside the VPN envelop of the IP packets [15]. Current QoS issues involved in VPN mainly focus on call admission control level [20,21] to determine the network ability for assigning network resources to mission-critical or delay-sensitive services while limiting resources committed to low-priority traffic as an essential component of VPN solution. According to service level agreements (SLA), IP traffic with multi-class of QoS from different users are classified and stored in separate buffers before they are transmitted in the network. The disadvantages of such mechanism are that (1) it complicates the implementation using separate buffers for different QoS classes, (2) the buffer utilization is inferior and (3) the re-sequencing is required at the destination which is highly undesirable for the traffic with different QoS classes but from the same user source.

* Corresponding author. Tel: +65-790-4508; fax: +65-792-0415.
E-mail address: elzhang@ntu.edu.sg (L. Zhang).

0140-3664/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0140-3664(01)00313-9

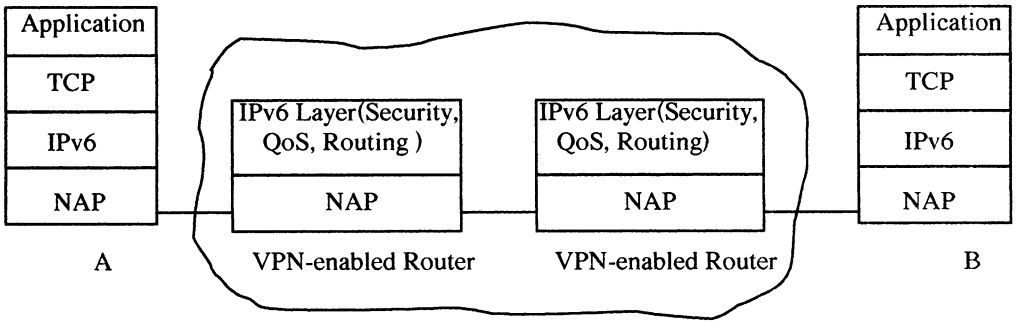

Fig. 1. VPN over IP backbone.

This paper focuses on the issues to improve QoS at IP packet level. The multi-class QoS is implemented on differentiated service basis using priority scheme of 4 bits defined in the packet header of IPv6. A VPN-enabled IP router is modeled as a tandem queuing system in which each output link consists of two space-priority output queues. The high priority queue is used to carry the delay sensitive traffic while the low priority queue for the delay insensitive traffic. On the other hand, multiple thresholds are implemented in each queue, respectively, for packet loss priority control [7]. The performance analysis is done using fluid flow technique [10,11]. The evaluation focuses on the performance trade-off between the delay sensitive traffic and delay insensitive traffic in terms of traffic throughput, packet loss probability and end-to-end delay in VPN networks are also presented.

Table 1
Recommended priority values in IPv6 for application categories

Priority	Application category
0	Uncharacterized traffic
1	'Filler' traffic (e.g. netnews)
2	Unattended data transfer (e.g. email)
3	(Reserved)
4	Attended bulk transfer (e.g. FTP, HTTP, NFS)
5	(Reserved)
6	Interactive traffic (e.g. telnet, X)
7	Internet control traffic (e.g. routing protocols, SNMP)

2. Multi-class QoS support in IP-based VPN

Since QoS control on call admission basis in VPN is enable to provide services with multi-class QoS by assigning the network resources on service application basis such as IP telephony, e-commerce, IP video, and other multimedia offerings. However, in mega-scale VPNs, when IP traffic flows increase in carrier-sized networks, QoS guarantee on flow-by-flow basis only at the call admission level is impractical [14]. One of the key alternative solutions is to implement multiple QoS classes at IP level. The IP-level QoS mechanism is to provide variable grade of service with performance guarantees to the heterogeneous mixture of the traffic flows. Currently, although IPv4 traffic has already run over Internet, however, it is still not able to fully and efficiently support the critical application traffics with different QoS requirements. As a natural increment to IPv4 [19], IPv6 is not only able to inter-operate with the current IPv4 but also able to run well in broadband networks to provide a

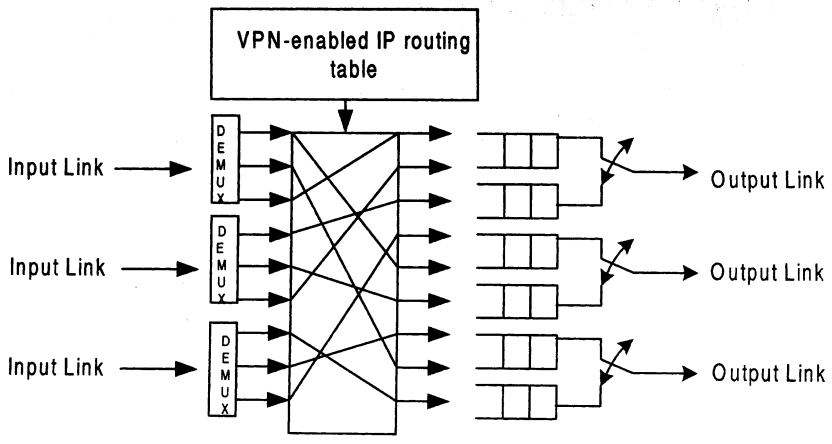

Fig. 2. Modeling of a VPN-enabled IP Router.

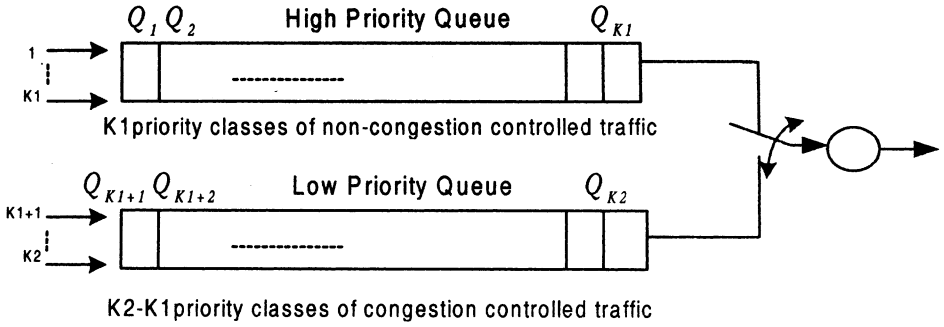

Fig. 3. The implementation of multi-QoS on priority basis.

platform for new Internet functionality, including the guarantee of QoS based on application priority levels. A 4-bit priority field is defined in the packet header of IPv6 for originating nodes and/or forwarding routers to identify and distinguish between the IPv6 packets with the different priority classes [1,3,4], so that it provides various forms of 'differentiated service'. These four priority bits representing 16 classes are divided into two categories [1]: (1) values 0–7 specify the priority of traffic for which the source is providing congestion control and (2) values 8–15 specify the priority of traffic that does not back off in response to congestion. For congestion-controlled traffic, the priority values recommended for particular application categories are shown in Table 1.

For non-congestion-controlled traffic, the lowest priority value (8) is used for those packets that the sender is most willing to be discarded under conditions of congestion, and the highest priority value (15) is used for those packets that the sender is least willing to be discarded.

The proposed QoS control mechanism in VPN is implemented at the following two levels. Firstly, VPN service providers can make appropriate pre-allocation of bandwidth to different classes of traffic such as delay sensitive applications or mission-critical applications at the call admission control level. This is based on off-line measurements of the network traffic characteristics to ensure that the SLAs of the customers are satisfied. Secondly, the ingress routers mark the IP packets to indicate the different QoS level using the priority field located in the packet header of IPv6 at the interface that packets enter the IP backbone. A VPN-enabled IP router located in the backbone network must have the capability to make high-speed traffic classification decisions, to read the content of the priority field in IPv6, to mark packets according to their QoS level and to transport packets over network with the guaranteed QoS. When congestion occurs in the router, the lowest priority traffic are most likely to be discarded first to guarantee the traffic with higher priority without loss or with the least loss on best-effort basis.

3. Modeling of VPN-enabled IP router

In VPN networks, VPN-enabled IP routers may be interconnected by several different input and output links. Each link may consist of a number of IP packet streams that may have different priority levels. However, these packet streams are multiplexed when they are transmitted over the transmission link. We assume that IP packets carried in the same IP stream all have the same priority level. It is expected that different priority levels are associated with different IP streams. Considering that a typical VPN-enabled IP router, as shown in Fig. 2, is modeled as a non-blocking tandem switching node associated with output queue at the output ports. The non-blocking switching function includes that the IP packet streams carried on the same input link are demultiplexed at the input port and then routed to the corresponding output port according to the IP routing table. At the output port, IP traffic streams with different priority levels are multiplexed before they are transmitted onto the output link. The multiplexer at the output port consists of two parallel output queues corresponding to two different delay priorities, respectively. Each output queue operates on a first-in-first-out (FIFO) non-preemptive basis while the queue is being served.

The multiple loss priority [7] is implemented using threshold control mechanism over the partitioned buffer of the two output queues, respectively. As shown in Fig. 3, the high priority queue is fed with non-congestion controlled traffic (i.e. delay sensitive traffic) which consists of $K1$ packet loss priority classes corresponding to the threshold $Q_1, Q_2, \ldots, Q_k, \ldots, Q_{K1}$, where k is the priority index and $k = 1$ (i.e. Q_1) corresponds to the highest packet loss priority. When the buffer occupancy of the high priority queue exceeds the threshold Q_k, $(1 \leq k \leq K1)$, only the non-congestion controlled IP packets with loss priority ranging

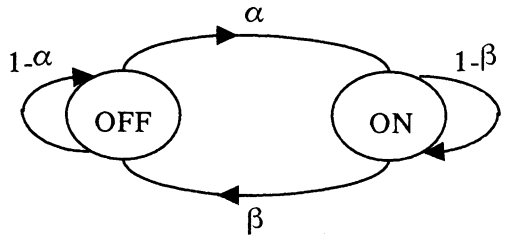

Fig. 4. ON/OFF traffic model.

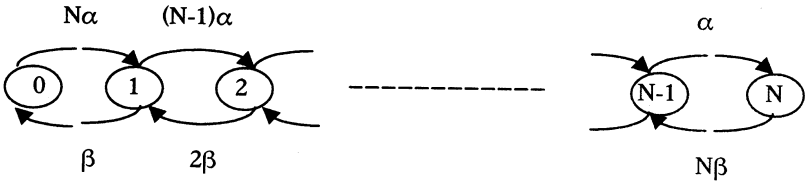

Fig. 5. Multiplexing of ON–OFF sources.

from 1 to $k-1$ are permitted to input the queue while packets with the other priority classes are discarded. When the buffer occupancy of the high priority queue exceeds the threshold Q_1, all the packets are lost. Likewise, the low priority output queue consisting of Q_{K1+1}, $Q_{K1+2},\ldots Q_{K2}$ different thresholds for packet loss control is fed with congestion controlled traffic (i.e. delay insensitive traffic) where $k = K1 + 1$ corresponds to the highest packet loss priority. Since the non-congestion controlled traffic has higher priority than those of the congestion controlled traffic, the packets contained in the high priority queue are always served first and the packets contained in the low priority queue are only served when the high priority queue is empty.

As shown in Fig. 4, IP packet stream generated by each user source is modeled as an ON–OFF process [5,6], in which the transition rate from ON state to OFF state is α and the rate of transition from OFF state to ON state is β. Then the probability that the process is in the ON state is $\alpha/(\alpha+\beta)$. When N of such ON–OFF packet streams are multiplexed, the resultant stream can be represented by an $(N+1)$-state Markov modulated process as shown in Fig. 5 [12], where the state i represents that $i(i = 0, 1, 2, \ldots, N)$ packet streams are in the ON state. The transition rate from the state i to the state $(i-1)$ is $i\beta$ and the transition rate from the state i to the state $(i+1)$ is $(N-\mathrm{i})\alpha$.

4. Performance analysis

The performance of statistically multiplexed ON–OFF traffic, as shown in Fig. 4, is generally evaluated using the following three technical approaches. The first approach is based on the work of Sriram and Whitt [9]. The key issue of their work relies on the approximation of the renewal processes as well as characterizations of the processes that are implemented by two moments. The second approach is based on Markov modulated Poisson process (MMPP) in which the packet generated by each source is a Poisson process with exponentially distributed in length. The transitions between the ON and the OFF states are controlled by an underlying continuous-time Markov chain [8]. The third approach is called fluid flow model [11,12], in which each individual source is modeled as a Markov modulated fluid source consisting of the ON state and the OFF state. In the ON–OFF process, the ON state is uniformly distributed and the transition between the ON and the OFF states is controlled by a continuous-time Markov chain which determines the rate of fluid generating. The transport of packets over a transmission link is operated in the same manner. The fluid flow technique has been applied successfully to a variety of problems in ATM networks. The fluid flow technique is used in the following performance analysis with assumptions that (1) the population of IP sources is large ($N \gg 1$), (2) the output transmission link is divided into equal time slots and each slot is equivalent to the maximum transmission time of IP packet such that the transmission time of IP packets is assumed to be uniformly distributed. This assumption is reasonable, because that when the population of IP sources is large, the bit rate allocated to each IP source comparing to the huge capacity of the broadband transmission link is negligible. Therefore, the variation of the transmission time of IP packet is also negligible.

4.1. Queuing analysis for the high priority queue

Considering that the input streams of the high priority queue consist of i packet steams which are in the ON state at time t. Now we define that$F_i^k(t,x)$ $(0 \le i \le n,\ 1 \le k < K1)$ is the cumulative probability distribution for the packet with kth loss priority in the queue at time t, where i packet streams are in the ON state. In fact, $F_i^k(t,x)$ $(0 \le i \le n,\ 1 \le k < K1)$ represents the probability that the queuing buffer occupancy is less than or equal to x $(Q_{k+1} \le x \le Q_k)$ while i packet steams in the ON state at time t. For $k = K1$, we have $0 \le x \le Q_{K1}$. As shown in Fig. 5, $F_i^k(t,x)$ can be calculated by setting up a

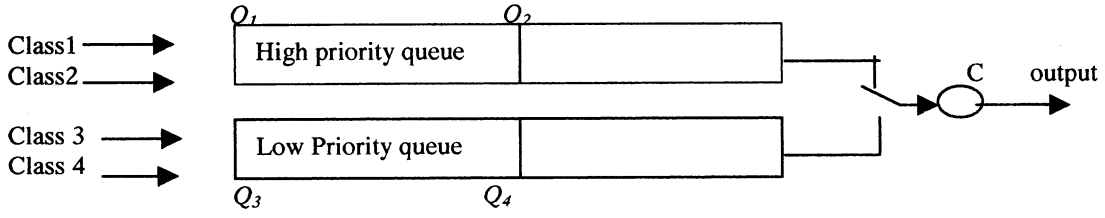

Fig. 6. Queuing Model for numerical analysis.

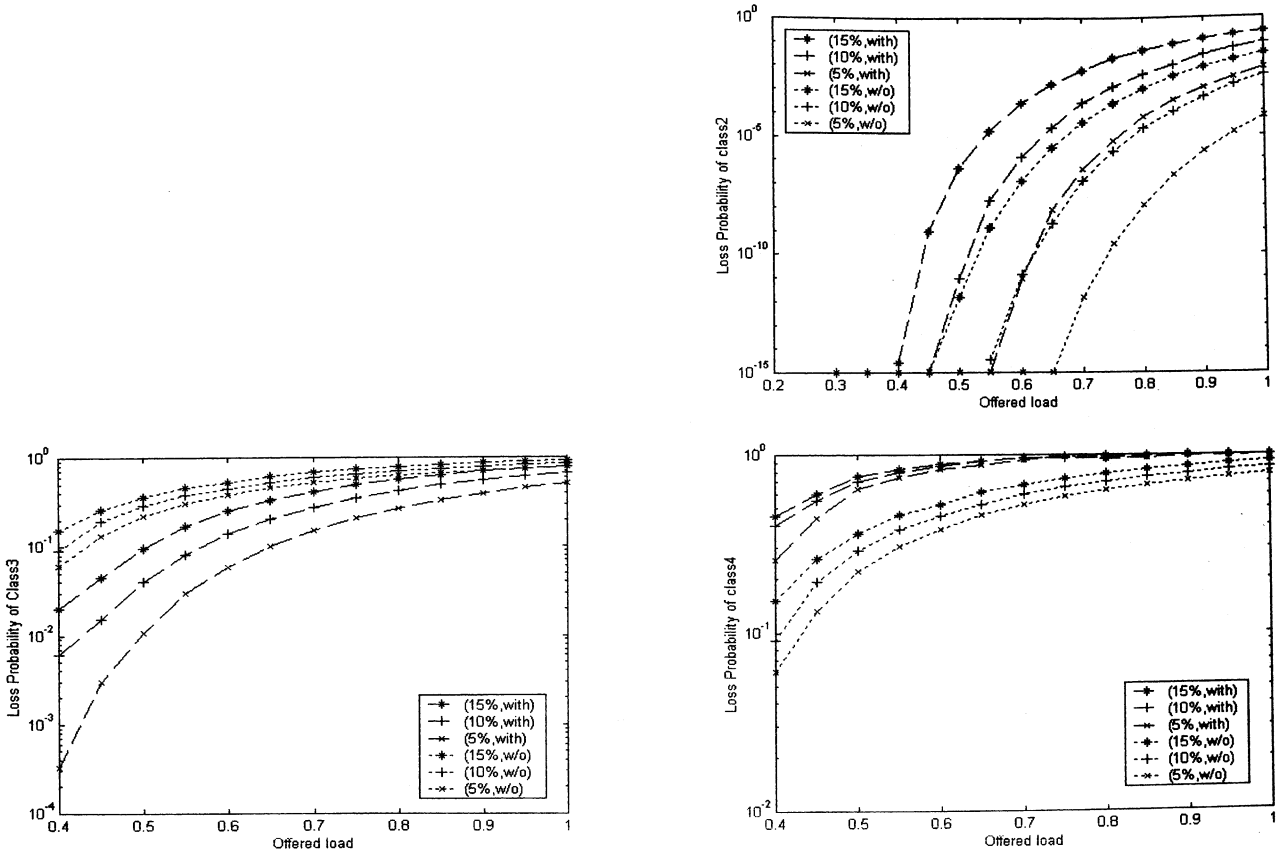

Fig. 7. The effect of Class 1 traffic load on packet loss probability.

generating equation of $F_i^k(t+\Delta t,x)$ which is the probability at an incremental time of $t+\Delta t$. Then, we have

$$F_i^k(t+\Delta t,x) = [N-(i-1)]\alpha\Delta t F_{i-1}^k(t,x) + (i+1)\beta\Delta t F_{i+1}^k(t,x) + \{1-[(N-i)\alpha + i\beta]\Delta t\}F_i^k[t,x-(i\lambda^k - C)\Delta t] + o(\Delta t), \quad (1)$$

where $\lambda^k = \sum_{j=1}^k \lambda_j$ (λ_j is the arrival rate of j class at ON state), the term $x-(i\lambda^k - C)\Delta t$ is the buffer occupancy. On the right side of Eq. (1), the first term is the probability of transition from the state $(i-1)$ to the state i at time $t+\Delta t$, the second term the probability of transition from the state $(i+1)$ to the state i, and the third term is the probability that the system state i is not changing at time t. And the term $o(\Delta t)$ represents all the higher order terms which go to zero much rapidly than Δt when Δt tends to zero. Hence, the effects of $o(\Delta t)$ is negligible when $\Delta t$is small enough. In Eq. (1), we also assume that $F_{-1}(t,x)$ and $F_{N+1}(t,x)$ are set equal to zero.

Now we expand $F_i^k(t+\Delta t,x)$ and $F_i^k(t,x-\Delta x)$ for $\Delta x = (i\lambda^k - C)\Delta t$ in their respective Taylor series with the assumption that the appropriate continuity conditions are satisfied. Let Δt go to zero, the Eq. (1) represents the following differential function:

$$(i\lambda^k - C)\,\mathrm{d}(F_i^k(x))/\mathrm{d}x = [N-(i-1)]\alpha F_{i-1}^k(x) + (i+1)\beta F_{i+1}^k(x) - [(N-i)\alpha + i\beta]F_i^k(x), \quad (2)$$

$$1 \le k \le K1,\ 0 \le i \le N,\ F_{-1}^k(x) = 0,\ F_{N+1}^k = 0.$$

Define $F^k(x) \equiv [F_0^k(x), F_1^k(x), \ldots, F_N^k(x)]^{\mathrm{T}}$, then Eq. (2) can be expressed in the following compact matrix form:

$$\mathbf{D}^k\,\mathrm{d}F^k(x)/\mathrm{d}x = \mathbf{M}F^k(x), \quad (3)$$

where $\mathbf{D}^k = diag[-C, \lambda^k - C, \ldots, N\lambda^k - C]$, $\mathbf{M}$ is the $(N+1)\times(N+1)$ tridiagonal matrix.

Assuming $i\lambda^k - C$ is not equal to zero for any i, $(0 \le i \le N)$, the general solution of Eq. (3) is given by

$$F^k(x) = \sum_{j=0}^{N} a_j^k \mathbf{V}_j^k\, \mathrm{e}^{\mathbf{z}_j^k x}, \quad (4)$$

where the elements in vector $\mathbf{z}^k = [z_0^k, z_1^k, \ldots, z_N^k]$ are the eigen value of matrix $(\mathbf{D}^k)^{-1}\mathbf{M}$ and $\mathbf{V}_j^k$ is the eigen vector of the matrix $(\mathbf{D}^k)^{-1}\mathbf{M}$.

In Eq. (4), the coefficients $\{a_j^k\}$ can be obtained from boundary conditions by defining $E_D^k = \{i|i\lambda^k < C\}$ and

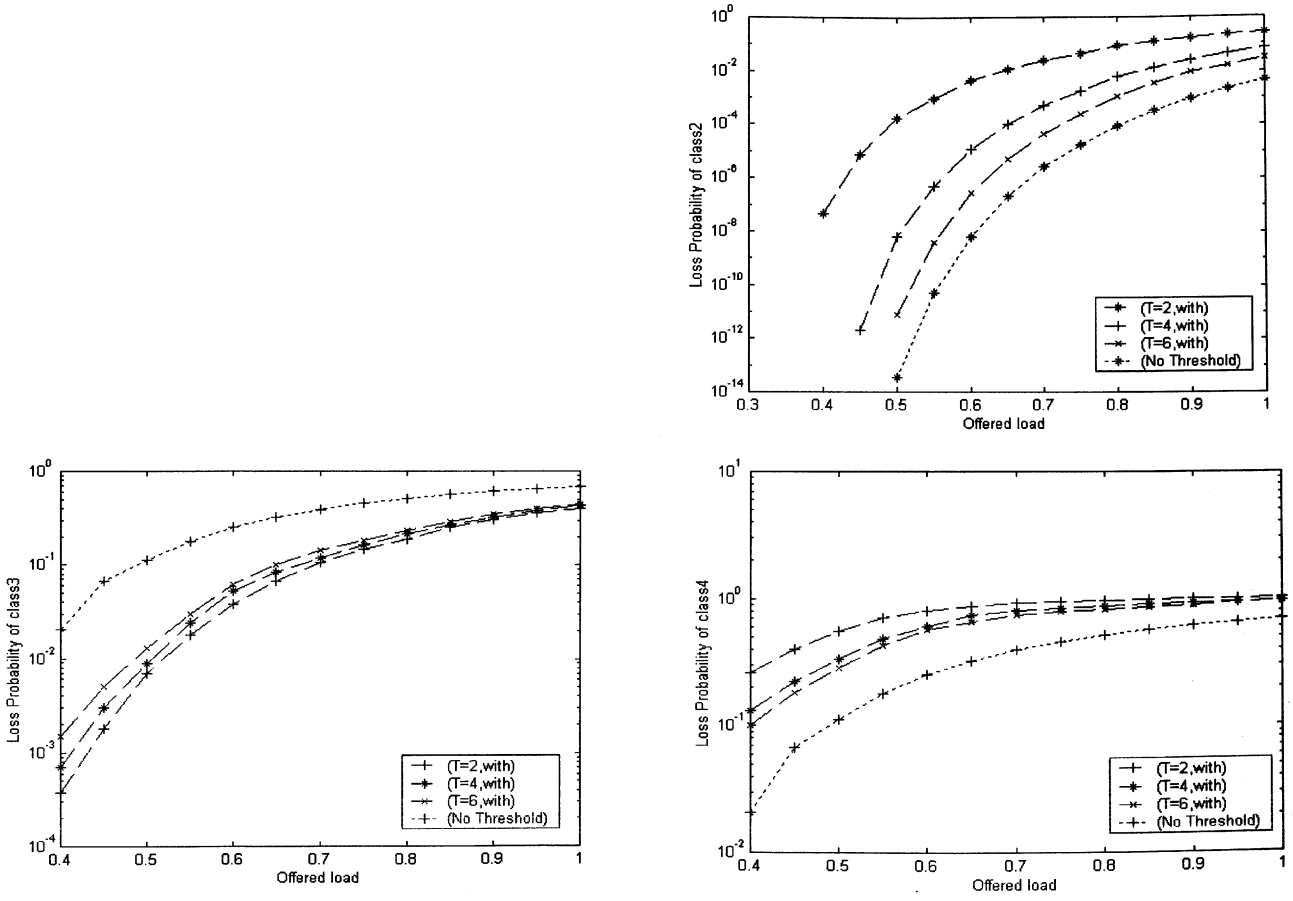

Fig. 8. The effect of threshold on packet loss probability for different traffic classes.

$E_u^k = \{i | i\lambda^k > C\}$. Then the boundary conditions in Eq. (4) can be obtained as below

$$\begin{aligned} F_i^{K1}(0) &= 0, && \text{if } i \in E_u^{K1}, \\ F_i^1(Q_1) &= P_i, && \text{if } i \in E_D^1, \\ F_i^k(Q_k) &= F_i^{k-1}(Q_k), && \text{if } i \in E_u^{k-1} \cup E_D^k,\ 2 \le k \le K1, \end{aligned} \tag{5}$$

where $P_i = {}_i^N P_{\text{on}}^i (1 - P_{\text{on}})^{N-i}$ is the probability that i sources are in the ON-state and $P_{\text{on}} = \alpha/(\alpha + \beta)$ is the probability that a source is in the ON-state. Hence, the steady state distributions under the boundary conditions of Eq. (5) can be used to calculate the throughput for the traffic with different priority classes, that is

$$T^k = \sum_{i=0}^{N} F_i^{k-1}(Q_k) i\lambda_k - \sum_{i \in E_u^k \cap E_D^{k-1}} \Big[(i\lambda^k - C)\Big(F_i^{k-1}(Q_{k-1}) - F_i^k(Q_k)\Big)\Big],$$

$k = 2, 3, \ldots, K0$.

For $k = 1$, the solution of T^1 is given by

$$T^1 = \sum_{i=0}^{N} F_i^1(Q_1)\ i\lambda_1 + C\left[1 - \sum_{i=0}^{N} F_i^1(Q_1)\right]. \tag{6}$$

The arrival rate for the traffic with the kth loss priority is given by $A^k = \sum_{i=0}^{N} i\lambda_k P_i$, $(1 \le k \le K1)$. Then the packet loss probability due to buffer overflow is given by

$$PL^k = 1 - T^k/A^k, \qquad 1 < k \le K1, \tag{7}$$

$$PL^1 = 1 - T^1/A^1, \qquad k = 1.$$

Likewise, the delay distribution for the traffic with different priority classes is the same as the buffer occupancy distribution. Let $W^k(t)$ be the probability that delay is less than t for the traffic with kth loss priority.

In the case of $Q_{n+1}/C < t < Q_n/C$, $n = k, \ldots, K1 - 1$, where Q_n is the nth threshold value and C is the capacity, the probability that the delay is less than t is

$$W_n^k(t) = 1/T^k \sum_{i=0}^{N} i\lambda_k F_i^n(Ct).$$

In the case of $0 \le t < Q_{K1}/C$, the probability that the

delay is less than t is

$$W_{K1}^k(t) = 1/T^k \sum_{i=0}^{N} i\lambda_k F_i^{K1}(Ct),$$

In the case of $t = Q_n/C$, the probability that the delay is less than t for the kth class of traffic is

$$Pr\{\text{delay} = Q_n/C\} = 1/T^k \sum_{i=0}^{N} i\lambda_k \Big(F_i^{n-1}(Q_n) - F_i^n(Q_n)\Big),$$

$\mathrm{n} = k, \ldots, K1.$

For traffic with the top packet loss priority, the probability that the delay is less than t is

$$Pr\{\text{delay} = Q_1/C\} = 1/T^1 \left(1 - \sum_{i=0}^{N} F_i^1(Q_1)\right) C. \tag{8}$$

The average delay for the traffic with the kth loss priority class ($k = 2, \ldots, K1$) can be obtained from the delay distribution function as shown by Eq. (8), that is

$$\begin{aligned}\mathbf{D}^k = &\int_0^{(Q_{K1}/C)^-} t\, \mathrm{d}(W_{K1}^k(t)) + \sum_{n=k}^{K1-1} \int_{(Q_{n+1}/C)^-}^{(Q_n/C)^-} t\, \mathrm{d}(W_n^k(t)) \\ &+ \sum_{n=k}^{K1} Q_n/C Pr\{\text{delay} \\ &= Q_n/C\}.\end{aligned} \tag{9}$$

For $k = 1$, the average delay is given by

$$\begin{aligned}(10)\mathbf{D}^1 = &\int_0^{(Q_{K1}/C)^-} t\, \mathrm{d}(W_{K1}^1(t)) \\ &+ \sum_{n=1}^{K1-1} \int_{(Q_{n+1}/C)^-}^{(Q_n/C)^-} t\, \mathrm{d}(W_n^1(t)) \\ &+ \sum_{n=2}^{K1} Q_n/C Pr\{\text{delay} \\ &= Q_n/C\} + Q_1/C Pr\{\text{delay} = Q_1/C\}.\end{aligned} \tag{10}$$

4.2. Queuing analysis for the low priority queue

The low priority queue carries the delay insensitive packet streams. Since the low priority queue is only served when the high priority queue is empty. The following analysis is considered into two cases as below: (1) the high priority queue has input traffic and (2) the high priority queue does not have input traffic. Let Pe be the probability that the high priority queue is empty, then Pe can be obtained from Eqs. (4) and (5) i.e., $Pe = \sum_{i=0}^{N} F_i^{K1}(0)$. In the first case, let $C = 0$. Recall Eq. (2), the differential function for the low priority queue is given by

$$\begin{aligned}(i\lambda^k - C)\, \mathrm{d}F_i^k(x)/\mathrm{d}x = &[N - (i-1)]\alpha F_{i-1}^k(x) \\ &+ (i+1)\beta F_{i+1}^k(x) - [(N-i)\alpha + i\beta]F_i^k(x),\end{aligned} \tag{11}$$

$k = K1 + 1, \ldots, K2,$

where $F_i^k(x)$ is cumulative probability distribution for the traffic with the kth loss priority in the low priority queue.

Likewise, recall Eq. (5), the boundary conditions for the low priority queue is given by

$$\begin{aligned}&F_i^{K2}(0) = 0, && 1 \le i \le N, \\ &F_i^k(Q_k) = F_i^{k-1}(Q_k), && 1 \le i \le N,\ K1 + 2 \le k \le K2.\end{aligned}$$

Hence, the steady state distributions under the above boundary conditions is given by

$$T_1^k = \sum_{i=1}^{N} i\lambda_k F_i^k(Q_k), \qquad K1 + 1 \le k \le K2. \tag{12}$$

The delay distribution function, throughput, average delay and packet loss probability for the delay insensitive traffic with different loss priority can be calculated using Eq. (12).

In the second case, the high priority queue does not have any input traffic stream. Assuming that the input streams of the high priority queue consist of j streams which are in the ON state. We define $C_j' = C - j\lambda 1$, where $\lambda 1 = \sum_{j=1}^{K1} \lambda_j$. Replacing $F_i^k(x)$ with $F_{ji}^k(x)$, in Eq. (11), we have the following differential function:

$$\begin{aligned}(i\lambda^k - C_j')\, \mathrm{d}F_{ji}^k(x)/\mathrm{d}x = &[N - (i-1)]\alpha F_{j(i-1)}^k(x) \\ &+ (i+1)\beta F_{j(i+1)}^k(x) - [(N-i)\alpha + i\beta]F_{ji}^k(x),\end{aligned} \tag{13}$$

where $F_{ji}^k(x)$ is the cumulative probability distribution for the traffic with the kth loss priority when the system in the state i. The boundary conditions for such a case is given by

$$\begin{aligned}&F_{ji}^{K2}(0) = 0, && \text{if } i \in E_{uj}^{K2}, \\ &F_{ji}^{K1+1}(Q_1) = P_i, && \text{if } i \in E_{Dj}^{K1+1}, \\ &F_{ji}^k(Q_k) = F_{ji}^{k-1}(Q_k), && \text{if } i \in E_{uj}^{k-1} \cup E_{Dj}^k,\ K1 + 2 \le k \le K2,\end{aligned} \tag{14}$$

where

$$E_{Dj}^k = \{i | i\lambda^k < C_j'\}$$

, $E_{uj}^k = \{i | i\lambda^k > C_j'\}$.

From Eq. (13) with the boundary condition given by Eq. (14), the throughput and packet loss probability for the delay

insensitive traffic with different loss priority is given by

$$T_{2j}^{k} = \sum_{i=0}^{N} F_{ji}^{k-1}(Q_k) i\lambda_k - \sum_{i \in E_{uj}^{k} \cap E_{Dj}^{k-1}} \left[\left(i\lambda^k - C_j' \right) \left(F_{ji}^{k-1}(Q_{k-1}) - F_{ji}^{k}(Q_k) \right) \right],$$

$$k1 + 2 \le k \le K2,$$

$$T_{2j}^{K1+1} = \sum_{i=0}^{N} F_{ji}^{K1+1}(Q_{K1+1}) i\lambda_{K1+1} + C_j \left[1 - \sum_{i=0}^{N} F_{ji}^{K1+1}(Q_{K1+1}) \right],$$

$$k = K1 + 1.$$

Then,

$$T_2^k = \sum_{j=0}^{j<C/\lambda 1} T_{2j}^{k} F_j^{K1}(0), \qquad K1 + 1 \le k \le K2, \tag{15}$$

$$A^k = \sum_{i=1}^{N} i\lambda_k P_i, \qquad K1 + 1 \le k \le K2,$$

$$PL^k = 1 - \left\lfloor T_1^k (1 - Pe) + T_2^k \right\rfloor / A^k, \qquad K1 + 1 \le k \le K2. \tag{16}$$

5. Numerical results and discussion

The following numerical results focus on the effect of priority on the steady-state performance including the delay sensitive traffic and the delay insensitive traffic with different packet loss priority classes. For the illustrative purpose only, as shown in Fig. 6, the high priority queue of length Q_1 consists of two classes of packet loss priority, named Class 1 and Class 2, where the Class 1 traffic has the higher packet loss priority than the Class 2 traffic. The threshold Q_2 is used for the packet loss priority control in the high priority queue. Likewise, the low priority queue of length Q_3 also consists of two classes of packet loss priority, named Class 3 and Class 4, where the Class 3 traffic has the higher packet loss priority than the Class 4 traffic. The threshold Q_4 is used for the packet loss priority control in the low priority queue. The input traffic of Class h ($h =$

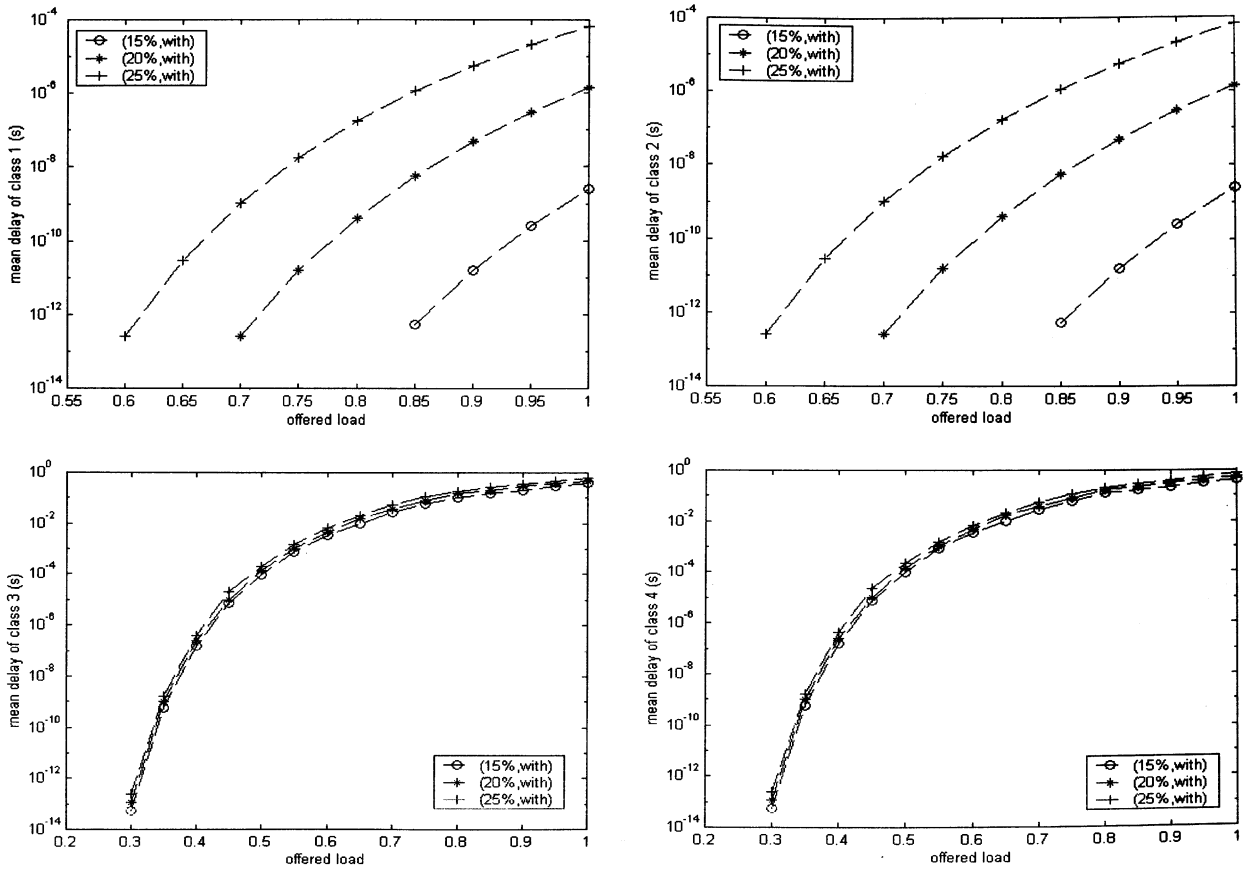

Fig. 9. The effect of Class 2 traffic load on average delay.

Table 2
Loss probability of Class 2, 3, 4 traffic with different percentage of Class 1 traffic

Traffic load			15% with priority	10% with priority	5% with priority	15% no priority	10% no priority	5% no priority
Class 2	0.70	Simulation	$3.8\times10^{-3}\pm3.2\times10^{-4}$	$2.93\times10^{-4}\pm3.1\times10^{-5}$	–	$4.25\times10^{-5}\pm2.6\times10^{-6}$	–	–
		Analysis	5.41×10^{-3}	2.02×10^{-4}	5.43×10^{-7}	3.74×10^{-5}	5.43×10^{-7}	1.42×10^{-12}
	0.75	Simulation	$1.5\times10^{-2}\pm4.2\times10^{-3}$	$1.01\times10^{-3}\pm2.7\times10^{-4}$	$6.42\times10^{-6}\pm2.3\times10^{-7}$	$3.03\times10^{-4}\pm3.8\times10^{-5}$	$1.75\times10^{-6}\pm2.1\times10^{-7}$	–
		Analysis	2.13×10^{-2}	1.12×10^{-3}	5.81×10^{-6}	2.13×10^{-4}	2.16×10^{-6}	2.48×10^{-10}
	0.80	Simulation	$4.4\times10^{-2}\pm1.7\times10^{-3}$	$5.23\times10^{-3}\pm1.6\times10^{-4}$	$4.86\times10^{-5}\pm1.6\times10^{-6}$	$1.21\times10^{-3}\pm1.5\times10^{-4}$	$4.23\times10^{-5}\pm1.7\times10^{-6}$	–
		Analysis	4.02×10^{-2}	3.75×10^{-3}	6.03×10^{-5}	9.02×10^{-4}	9.26×10^{-5}	1.04×10^{-8}
	0.85	Simulation	$7.8\times10^{-2}\pm2.6\times10^{-3}$	$8.78\times10^{-3}\pm3.2\times10^{-4}$	$2.53\times10^{-4}\pm2.4\times10^{-5}$	$5.27\times10^{-3}\pm2.6\times10^{-4}$	$1.31\times10^{-4}\pm2.7\times10^{-5}$	–
		Analysis	0.0748	9.42×10^{-3}	3.24×10^{-4}	3.13×10^{-3}	1.22×10^{-4}	2.15×10^{-7}
	0.90	Simulation	$0.117\pm3.4\times10^{-2}$	$1.93\times10^{-2}\pm2.3\times10^{-3}$	$1.41\times10^{-3}\pm3.2\times10^{-4}$	$1.02\times10^{-2}\pm3.2\times10^{-3}$	$2.87\times10^{-4}\pm4.2\times10^{-5}$	$2.8\times10^{-6}\pm1.6\times10^{-7}$
		Analysis	0.1188	2.49×10^{-2}	1.13×10^{-3}	7.52×10^{-3}	$4.03\times10^{-}$	2.12×10^{-6}
Class 3	0.70	Simulation	$0.432\pm1.2\times10^{-2}$	$0.2546\pm1.1\times10^{-2}$	$0.1800\pm1.5\times10^{-2}$	$0.6768\pm2.2\times10^{-2}$	$0.6259\pm1.1\times10^{-2}$	$0.524\pm1.2\times10^{-2}$
		Analysis	0.4088	0.2706	0.1517	0.6522	0.6002	0.5237
	0.75	Simulation	$0.491\pm1.3\times10^{-2}$	$0.3876\pm2.7\times10^{-2}$	$0.2321\pm1.8\times10^{-2}$	$0.7332\pm1.8\times10^{-2}$	$0.6363\pm2.1\times10^{-2}$	$0.582\pm1.6\times10^{-2}$
		Analysis	0.5000	0.3501	0.2074	0.7183	0.6566	0.5799
	0.80	Simulation	$0.453\pm4.1\times10^{-2}$	$0.4713\pm1.6\times10^{-2}$	$0.2999\pm1.3\times10^{-2}$	$0.7826\pm1.5\times10^{-2}$	$0.7158\pm1.3\times10^{-2}$	$0.602\pm2.2\times10^{-2}$
		Analysis	0.5699	0.4189	0.2672	0.7454	0.7063	0.6295
	0.85	Simulation	$0.613\pm1.6\times10^{-2}$	$0.5218\pm2.2\times10^{-2}$	$0.3552\pm2.4\times10^{-2}$	$0.8256\pm2.6\times10^{-2}$	$0.726\pm1.7\times10^{-2}$	$0.692\pm1.8\times10^{-2}$
		Analysis	0.6361	0.4901	0.3301	0.8379	0.7504	0.6736
	0.90	Simulation	$0.672\pm3.4\times10^{-2}$	$0.5974\pm1.3\times10^{-2}$	$0.3711\pm1.6\times10^{-2}$	$0.8625\pm1.2\times10^{-2}$	$0.7541\pm1.2\times10^{-2}$	$0.725\pm2.1\times10^{-2}$
		Analysis	0.6945	0.5555	0.3928	0.8501	0.7895	0.7129
Class 4	0.70	Simulation	$0.958\pm1.3\times10^{-2}$	$0.9398\pm1.5\times10^{-2}$	$0.9034\pm1.6\times10^{-2}$	$0.6768\pm2.2\times10^{-2}$	$0.6259\pm1.1\times10^{-2}$	$0.524\pm1.2\times10^{-2}$
		Analysis	0.9654	0.9387	0.9206	0.6522	0.6002	0.5237
	0.75	Simulation	$0.963\pm1.9\times10^{-2}$	$0.9403\pm1.7\times10^{-2}$	$0.9178\pm1.2\times10^{-2}$	$0.7332\pm1.8\times10^{-2}$	$0.6363\pm2.1\times10^{-2}$	$0.582\pm1.6\times10^{-2}$
		Analysis	0.9700	0.9481	0.9321	0.7183	0.6566	0.5799
	0.80	Simulation	$0.971\pm3.1\times10^{-2}$	$0.9607\pm1.3\times10^{-2}$	$0.9320\pm2.3\times10^{-2}$	$0.7826\pm1.5\times10^{-2}$	$0.7158\pm1.3\times10^{-2}$	$0.602\pm2.2\times10^{-2}$
		Analysis	0.9802	0.9581	0.9433	0.7454	0.7063	0.6295
	0.85	Simulation	$0.978\pm1.4\times10^{-2}$	$0.9746\pm2.4\times10^{-2}$	$0.9830\pm1.5\times10^{-2}$	$0.8256\pm2.6\times10^{-2}$	$0.726\pm1.7\times10^{-2}$	$0.692\pm1.8\times10^{-2}$
		Analysis	0.9894	0.9857	0.9743	0.8379	0.7504	0.6736
	0.90	Simulation	$0.986\pm2.4\times10^{-2}$	$0.9923\pm1.6\times10^{-2}$	$0.9932\pm1.1\times10^{-2}$	$0.8625\pm1.2\times10^{-2}$	$0.7541\pm1.2\times10^{-2}$	$0.725\pm2.1\times10^{-2}$
		Analysis	0.9950	0.9936	0.9900	0.8501	0.7895	0.7129

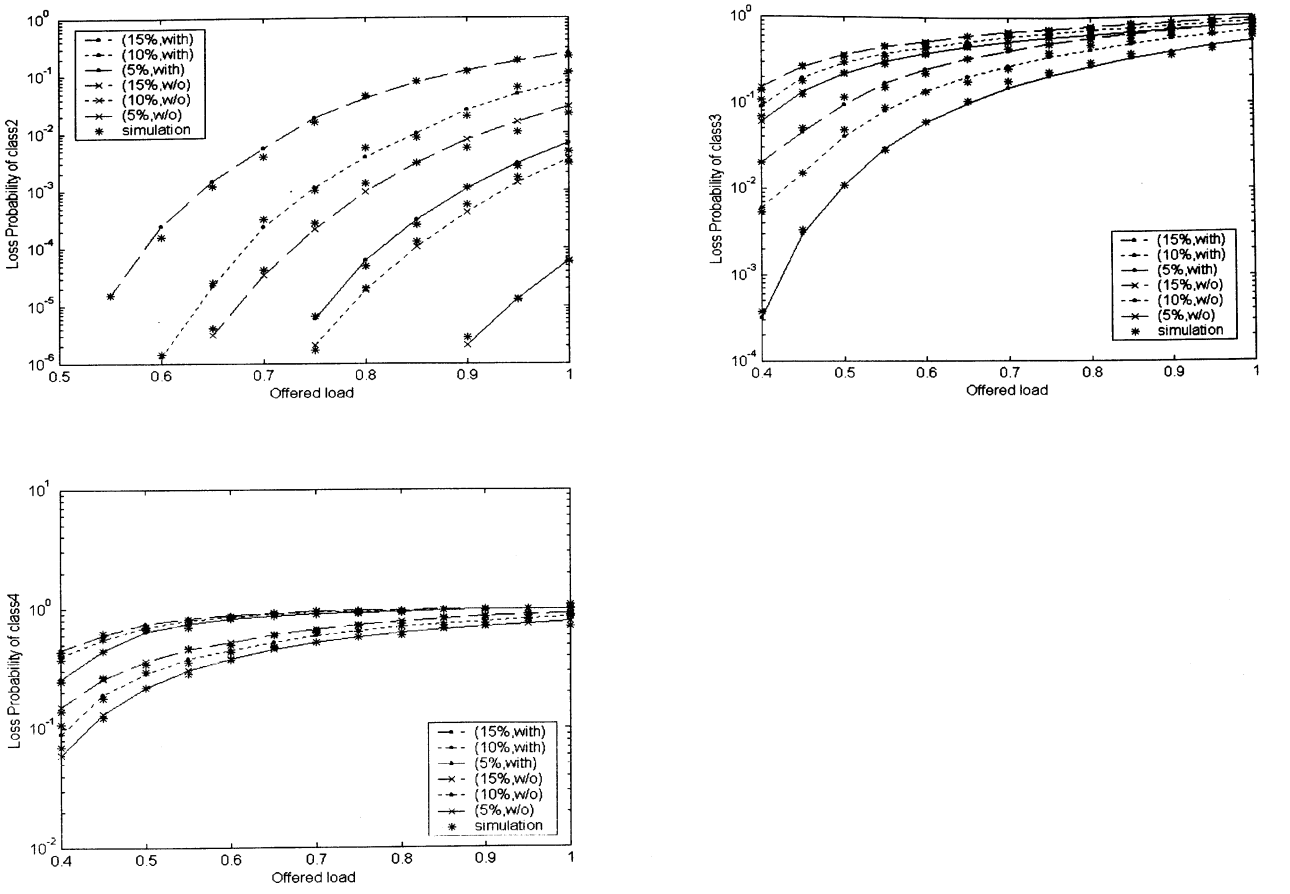

Fig. 10. Loss probability of four classes traffic at different percentage of Class 1 traffic.

$1, 2, \ldots, 4$) consists of N_h homogeneous independent ON–OFF sources in which the ON state and the OFF state are exponentially distributed, respectively, with different mean values.

The effect of priority on the performance of packet loss probability for the delay sensitive traffic and the delay insensitive traffic are illustrated in Fig. 7. It can be seen that the traffic load offered by the Class 1 has the significant effects on the performance of packet loss probability for the all the other lower priority traffic classes. Therefore, in order to achieve the desired packet loss probability for different traffic classes, both the overall traffic load in the network and the traffic load offered by the Class 1 need to be properly controlled. In addition, when the traffic load of Class 1 decreases, the throughput of Class 1 traffic decreases obviously. By contrast, the throughput of the other lower priority traffic classes increases significantly.

Fig. 8 shows the packet loss probability for different threshold values, respectively, where the traffic offered load of the Class 1 is fixed at 15% of the link capacity. Fig. 8 demonstrates that the effect of the different threshold values on the performance of packet loss probability is significant for all traffic classes. For example, when the traffic loading is 80% and the threshold value of Q_2 and Q_4 increase from 2 to 6, respectively, the packet loss probability for the Class 2 traffic is reduced from 8.72×10^{-2} to 1.13×10^{-3} and the packet loss probability for the Class 4 is also improved. Likewise, the different threshold values make the same significant effect on performance of the throughput for Class 3 and 4 traffic.

Fig. 9 presents the average delay versus the offered load for the traffic with highest priority class. As shown in Fig. 9, when traffic load of Class 1 decreases and the traffic load offered by other traffic classes are fixed, the improvement of delay for Class 1 and 2 is significant, by contrast, the improvement for Class 3 and 4 is insignificant. Likewise, when the traffic load of Class 2 increases, the results are same as Fig. 9. However, when the traffic load of Class 4 increases, the improvement of delay for the traffic of Class 3 and 4 is insignificant. This is because that the effect of Class 4 traffic on the performance of delay is limited.

Simulation has also been performed to verify the analytic results. The simulation model and the related parameters are the same as those are used in the numerical analysis. Fig. 10 and Table 2 illustrate the comparisons of the performance of packet loss probability with different percentage of the Class 1 traffic obtained from the simulation and the numerical analysis. In the simulation, a number of independent simulation runs have been performed. Each simulation run consists of 10^8 time slots plus an additional start-up transient

period of 1000 time slots. The error bars of the different simulation runs are also illustrated in the tables. From the above figure and table, it can be seen that the results obtained from the simulation match the numerical results well.

6. Conclusions

The introduction of multi-class priority defined in IPv6 traffic on priority basis makes the QoS control in VPN more flexible. The numerical results show that incorporating priority scheme defined by IPv6 in VPN is able to improve the performance for the high priority traffic classes. However, the priority schemes do not reduce the total packet loss but do protect the high priority traffic from packet loss while allowing the performance of the low priority traffic to degrade as little as possible, especially when the traffic loading and the threshold value are properly controlled. The behavior of multi-class priority scheme is studied with a variety of traffic conditions. The obtained results show that the high priority traffic improve vastly with the use of multi-class priority scheme under the condition that the proportion of high priority traffic (for both the high priority offered load and the ratio of high priority sources) must be kept to a small percentage. On the other hand, the burstiness of the traffic must also be carefully controlled.

References

[1] S. Deering, R. Hinden, Internet Protocol, Version 6 (IPv6) Specification, RFC1883 December (1993).
[2] S. Deering, R. Hinden, Internet Protocol, Version 6 (IPv6) Specification, RFC2460 December (1998).
[3] S.A. Thomas, IPng and the TCP/IP Protocols: Implementing the Next Generation Internet, Wiley, New York, 1996.
[4] S.O. Bradner, IPng, Internet Protocol Next Generation, Addison Wesley, Reading, MA, 1996.
[5] M. Hirano, N. Watanabe, Characteristics of a Cell Multiplexer for Bursty ATM Traffic, ICC'89 (1989) 399–403.
[6] D.E. McDysan, D.L. Spohn, ATM: Theory and Application, McGraw Hill, New York, 1994.
[7] A. Elwalid, D. Mitra, Fluid models for the analysis and design of statistical multiplexing with loss priorities on multiple classes of bursty traffic, IEEE INFOCOM' 92 May (1992) 0415–0425.
[8] H. Heffes, D. Lucantoni, A. Markov, Modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, IEEE JSAC 4 (6) (1986) 856–868.
[9] K. Sriram, W. Whitt, Characterizing superposition arrival processes in packet multiplexers for voice and data, IEEE JSAC 4 (6) (1986) 833–846.
[10] B. Maglaris, P. Anastassiou, P. Sen, G. Karlsson, J.D. Robbins, Performance models of statistical multiplexing in packet video communications, IEEE Trans. Commun. 36 (7) (1988) 834–843.
[11] D. Anick, D. Mitra, M.M. Sondhi, Stochastic theory of a data-handling system with multiple sources, Bell Syst. Technol. J. 61 (1982) 1871–1894.
[12] M. Schwartz, Broadband Integrated Networks, Prentice-Hall, Englewoodcliffs, NJ, 1996.
[13] Cisco Systems, Intranet and Extranet Virtual Private Networking, http://www.cisco.com/warp/public/cc/cisco/mkt/servprod/dial/tech/ievpn_rg.htm.
[14] Cisco Systems, Quality of Service for Virtual Private Networks,http://www.cisco.com/warp/public/cc/sol/mkt/ent/vpne/qsvpn_wp.htm.
[15] R. Moskowitz, IPv6 For VPNs: It's Looking Better All The Time,http://www.nwc.com/901/901colmoskowitz.html.
[16] B. Gleeson, A. Lin, J. Heinanen, G. Armitage, A Framework for IP Based virtual Private Networks, Internet Draft, draft-gleeson-vpn-framework-00.txt, September 1998.
[17] P.B. Busschbach, Toward QoS-capable virtual private networks, Bell Labs Technol. J. October–December (1998) 161–175.
[18] S. Blake, et al., An architecture for differentiated services, IETF RFC 2475 December (1998).
[19] J. Postel, Internet Protocol, RFC-791 September (1981).
[20] D. Mitra, J.A. Morrison, K.G. Ramakrishnan, Virtual private networks: joint resource allocation and routing design, Proc. IEEE INFOCOM'99 April (1999) 480–490.
[21] D. Mitra, I. Ziedins, Hierachical virtual partitioning: algorithms for virtual private networking, Proc. IEEE GLOBECOM (1997) 1784–1791.