QoS Guaranteed Voice Traffic Multiplexing Scheme over VoIP Network using DiffServ

Eun-Ju Ha, Joon-Heup Kwon, and Jong-Tae Park

Department of Electronic Engineering, Kyungpook National University, 1370 KanKyug-Dong, Buk-Gu, Taegu, Korea

ejha@ain.knu.ac.kr, joonheup@hotmail.com, park@ee.knu.ac.kr

Abstract. The current VoIP transfer methods using low bit rate codes such as G.723.1 and G.729 are still very inefficient due to its small payload size in comparison with its large overhead size. Besides, the traffic load on the access routers tends to geometrically increase when the incoming traffic flows of short packets increase. These factors cause problems such as delay, jitter, and packet loss, which seriously deteriorate the voice quality. We propose a new efficient voice traffic multiplexing scheme with guaranteed end-to-end QoS between VoIP access routers using differentiated services (DiffServ). The newly defined RTP/UDP/IP packets, namely, L_packet (long packet) using DiffServ are multiplexed at ingress routers for offering real-time communication services in very high-speed backbone network. We analyze the performance of proposed scheme for various bit rate type traffics and simulate the call blocking probability. It shows that the proposed scheme quite satisfactorily guarantees the end-to-end QoS requirements.

1 Introduction

The VoIP systems are expected to be very extensively used throughout the world because of their low price, better user interfaces, and potentiality for providing integrated multimedia applications over IP networks. Internet can’t satisfy various quality of service (QoS) requirements mainly due to its transport mechanism based on best effort services. In order to transmit the voice over packet network, VoIP access routers are required. VoIP access router provides an interface between the existing circuit switched telephony system such as PSTN/PLMN and the packet switched IP networks.

To guarantee a quality of service for real-time audio and video applications, the IP based protocol stack of the real-time transport protocol (RTP) over user datagram protocol (UDP) is required. RTP is used as means of achieving interoperability among different implementations of network audio and video applications, and integrated into the H.323 protocol stack, which was standardized by the ITU-T. RTP is often used to unicast and multicast real-time traffic transmission of audio, video, and interactive simulation over Internet, but it can’t support end-to-end QoS. So, we need an appropriate RTP/UDP/IP packet encapsulation procedure for the incoming voice traffic over access networks to guarantee QoS.

In a traditional VoIP application, telephone calls between PSTN/PLMN users interconnected by a pair of IP-ARs are carried by separate IP/UDP/RTP connections. At present, the low bit rate codes such as G.723.1 [1] and G.729 [2] are used to the baseline codec of VoIP. They compress incoming speech samples, and generate packets with sizes ranging from 5 to 20 bytes per speech sample. For example, G.723.1 compresses voice traffic to 5.3 kbps/6.3 kbps and the payload size is defined to be 20 bytes. That is, only one-third of all the traffic is payload and other two-thirds (i. e., RTP (12 bytes) + UDP (8 bytes) + IP (20 bytes)) is overhead. If voice traffic is sent in a RTP packet, this means that only 33 % of the total size of the packet is used to the payload. In addition, the traffic load on the routers would increase geometrically due to the traffic flows of many short packets. This causes the delay, jitter and packet loss problems. Thus, in order to guarantee end-to-end QoS requirements for real time voice service over IP network, we must take proper actions of decreasing this large overhead of the increased IP/UDP/RTP packet headers between VoIP access routers in the first place.

It is expected that the future backbone network will be constructed by using very high-speed WDM network with more than 10Giga bps. In this kind of very high-speed backbone network, the main cause of QoS deterioration is due to poor bandwidth control. To guarantee end-to-end QoS in high-speed backbone network, integrated services (IntServ) and resource reservation protocol (RSVP) has been proposed. However, these approaches are based on the per-flow management, which may cause a scalability problem. It is also known that these approaches cost very high for the construction. As another solution, the IETF has proposed a new architecture, called differentiated services (DiffServ). This architecture allows flow aggregation, that is, a number of traffic flows being handled as one unit. Except for fixed bandwidth allocation, it is shown that the flow aggregation based on DiffServ is efficient to guarantee end-to-end QoS in a high-speed backbone network [3]. Secondly, in order to guarantee QoS in very high-speed backbone network, the bandwidth control mechanism is required to incoming RTP/UDP/IP packets of ingress routers.

There are some related works dealing with these voice streams multiplexing issues [1-3] and DiffServ [4-5]. But, previous related researches have usually focused on access network, or backbone network, not both in an integrated way. Their solutions can’t be applied for the satisfaction of the true end-to-end QoS requirements. In this paper, we propose a new end-to-end QoS guaranteeing mechanism by combining RTP/UDP/IP packet multiplexing scheme with DiffServ QoS architecture.

We adapt RTP/UDP/IP packet multiplexing scheme in Ref [6] to our scheme with some modification. At ingress router, the appropriate control mechanism is required to support end-to-end QoS for input traffic. We solve this problem by using DSCP (DiffServe Code Point) of DiffServ. We analyze the performance of proposed scheme using M/G/1 with HOL-NPR (head-of-line non-preemptive) queueing model. This scheme not only significantly reduces the short traffic flow management cost, but also guarantees the end-to-end QoS requirements. The remainder of this paper is organized as follows: Section 2, we present the overall architecture and multiplexing packet formats of voice traffic using DiffServ. In Section 3, we present the analysis results and Section 4 shows the numerical results. Our conclusion follows in Section 5.

2 VoIP System Architecture using DiffServ

In this section, we present overall VoIP architecture and its multiplexing packet format using DiffServ. Figure 1 shows the overall architecture. The number of access networks, which are connected to one ingress router depends on the real capacity of ingress router. We use the RTP/UDP/IP packet multiplexing concept based on Ref[5] to reduce the packet overhead to 67%. However, the major problem of Ref[5] is that it does not take into account any QoS requirements. Thus, we modify voice traffic multiplexing scheme in Ref[5] in order to support QoS requirements. We propose the new multiplexing scheme using DSCP of DiffServ [6]. The key idea is as follows; *Add voice stream multiplexing scheme* ***with identical destination IP address*** *and* ***DSCP (DiffServ Code Point)***

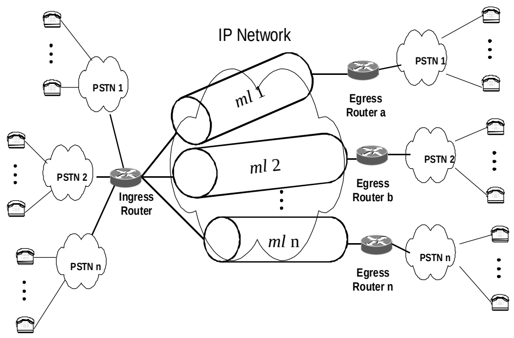


Figure 1. Overall Architecture

Figure 2 presents the transmission link between ingress router and egress router in more detail. The same model can be applied to the other links. We define the several parameters shown in Figure 2. *ml* is defined as the multiplexing link. We assume that the capacity of link is infinite. *t* is defined as a trunk. In our model, trunk is defined to be a group of voice traffic, which has identical destination and DSCP. *vt* is defined as voice traffic. We assume that the capacity of *vt* is all the same.

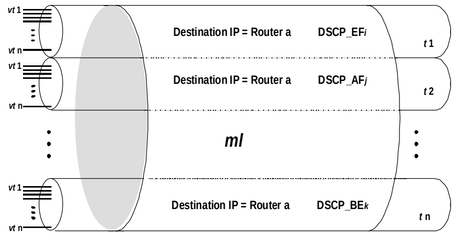
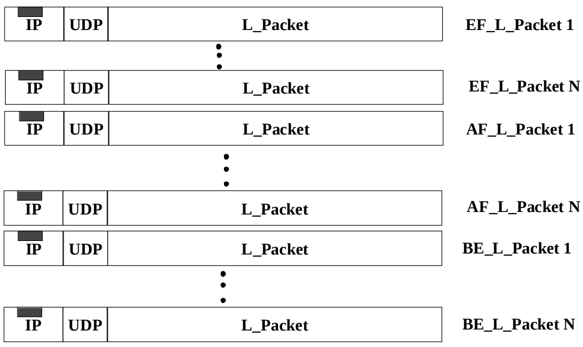


Figure 2. Multiplexing link between Ingress router and Egress router

In order to guarantee the end-to-end QoS, we propose new RTP/UDP/IP packet format at ingress router. We define L_packet (Long Packet), which adapt DiffServ DSCP to RTP/UDP/IP packet format. Figure 3 shows the L_packet format, which uses the ingress, intermediate, and egress router. At ingress router, the L_packet, which has the same destination and DSCP, is classified and destined to egress router through several intermediate routers. At intermediate router, it only transits the L_packet without any modification based on DSCP value. At egress router, the arrived L_packets are separated and processed according to the priority of DSCP value.

(a) Previous RTP/UDP/IP Packet Format in Ref [6]



(b) Proposed RTP/UDP/IP Packet Format using DiffServ

Figure 3. Multiplexing RTP/UDP/IP Packet Format

3 Performance Evaluation

3.1 Mathematical Model

In this section, we first describe the mathematical model. In the near future, it is expected that the backbone network is to be installed with very high-speed WDM network with more than 10Giga bps. In this very high-speed backbone network, the main problem for the QoS deterioration is caused by poor bandwidth control. Thus, we describe the end-to-end voice traffic blocking probability based on M/G/1 with HOL-NPR (head-of-line non-preemptive) queueing model in Figure 2. We assume that at least one trunk can be established between ingress router and corresponding egress router. The voice traffic blocking occurs by two things. One is due to the contention among voice traffics sharing the same trunk, with the same destination of egress router. The other is caused by contention among trunks sharing the link. We first derive the mathematical model of the first part. In this case, we assume that the incoming voice traffics are classified into their priorities and complete the multiplexing procedures at ingress router..

Link capacity of an ingress router is denoted by *C(ml)*. The number of voice traffics accommodated at each trunk depends upon their traffic characteristics. Each trunk has different voice traffic arriving rate and mean service time. We assume that there are P types of queues. In a queue of type k (k = 1, 2, …, P), there are m_k servers and no waiting room. When there is no trunk established between ingress router and corresponding egress router and a setup requirement of voice traffic between them newly arrives, ingress router checks whether there is still bandwidth available for a new trunk establishment on the link. The ingress router will reserve some amount of the bandwidth required by m_k voice traffic for a trunk of type k on the link if possible.

Furthermore, if the corresponding trunk is already established and it can further accommodate more flow, the flow will be accepted. Otherwise, the flow will be blocked. The trunk will be released when it has no voice traffic. Arrivals of customers at a queue of type k are characterized by an independent Poisson process with an arrival rate λ_k, and with an arbitrary or general service time distribution of form *B(x)* with a finite mean h_k.

3.2 Performance Analysis

In this section, we analyze the mathematical model described in the previous section. It follows the Ref [2,3]. Let $\rho_k = \lambda_k h_k$ denote the traffic intensity in a trunk of type *k*. We denote by $P_k(i)$ the steady state probability of *i* (*i* = 1,2, …, m_k) customers at a trunk of type *k* using the Ref [7].

$$P_k(i) = P_k(0)\frac{\rho_k^i}{i!} \quad \text{Where,} \quad P_k(0) = \left[\sum_{i=0}^{mk}\frac{\rho_k^i}{i!}\right]^{-1} \qquad (1)$$

Let $B(m,\rho)$ denotes the blocking probability with traffic intensity ρ. It is well known that $B(m,\rho)$ can be computed by the recursion [8];

$$B(m,\rho) = \frac{\rho B(m-1,\rho)}{m+\rho B(m-1,\rho)}, \qquad \text{m} = 1,2,\ldots \qquad (2)$$

with $B(0,\rho)=1$. Thus, the conditional mean number M_k is given by

$$M_k = \frac{\sum_{i=1}^{m_k} i P_k(i)}{1-P_k(0)} = \frac{\rho_k(1-B(m_k,\rho_k))}{1-P_k(0)} \qquad (3)$$

Next, we describe the mean busy period H_k of a trunk of type *k*. Since the sequence of idle and busy period is regarded as an alternating renewal process, we have

$$H_k = \frac{1-P_k(0)}{\lambda_k P_k(0)} \qquad (4)$$

Customer loss occurs in two ways. One is that a customer arrives when his trunk is full, and the other is that the trunk is empty but the reservation at the second step is failed on arrival. We denote the blocking probability of customers arriving in trunks of type *k* in the former and the latter cases by $P_{loss}^{(1)}(k)$ and $P_{loss}^{(2)}$, respectively. Apparently, the overall blocking probability $P_{loss}(k)$ of customers arriving in a trunk of type *k* is given by $P_{loss}(k) = P_{loss}^{(1)}(k) + P_{loss}^{(2)}(k)$. We first consider $P_{loss}^{(1)}(k)$. We define that the probability $r(k)$ that a trunk of type k is reserving m_k servers using the Ref [7].

$$r(k) = \frac{U_k(N, C(ml))}{N_k} \qquad (6)$$

Therefore, we have

$$P_{loss}^{(1)}(k) = r(k)\frac{B(m_k, \rho_k)}{1 - P_k(0)} \qquad (7)$$

Next we consider the overall blocking probability $P_{loss}(k)$. Since the blocking probabiity is a fraction of lost customers to customers arriving in a unit time, we have

$$P_{loss}(k) = 1 - \frac{M_k U_k(N, C(ml))}{N_k \rho_k} \qquad (8)$$

However, the above call blocking probility are effective in the case of the total waiting time for customers from type k in class N can't exceed the end-to-end delay bounds. The total waiting time (T_p) in our queueing model is defined as follows.

$$T_p = \frac{h_k(1-\sigma_k) + \sum_{i=1}^{m_k} \lambda_k h_k^2/2}{(1-\sigma_k)(1-\sigma_{k+1})} \quad \text{Where,} \quad \sigma_k = \sum_{i=1}^{m_k} \rho_k \qquad (9)$$

4 Numerical Results

In this section, we evaluate the blocking probability using our proposed scheme. We treat the heterogeneous trunk case. We defined related parameters as follows;

Defined Parameters

- Link capacity: $C(ml)$ = 150 Mbps
- Allocated bandwidth for each voice traffic: $N(vt)$ = 10 kbps
- Number of traffic types: P = 2 (high and low)
- Number of low trunks (i. e. DSCP_AF and DSCP_BE): $N(L)$ = 100
- Number of high trunks (i. e. DSCP_EF): $N(H)$ = 10
- Mean traffic intensity for each low trunk: ρ_L = 2 voice traffic
- Mean holding time: h_L = 1 [min]
- Mean interarrival time: $1/\lambda_L$ = 30 [sec]
- Mean traffic intensity for each high trunk: ρ_H = 1250 voice traffic
- Mean holding time: h_H = 1 [min]

- Mean interarrival time : $1/\lambda_H$ = 48 [msec]
- Capacity of each of DSCP_AF and DSCP_BE trunks : m_L flows
- Capacity of each of DSCP_EF trunks : m_H flows

We assume two types of traffic such as high and low. The high traffic flows are preferentially assigned to the DSCP_EF trunks. The high trunks have larger capacity to meet their higher demand than the low trunks. The low traffic flows are assigned to the DSCP_AF and DSCP_BE trunks according to the trunk conditions. In our evaluation, the total amount of traffic is fixed to 12700 flows. So, the mean utilization becomes 84.7% if there is no blocking. Table 1 shows the voice traffic blocking probability according to the varying capacity of each of low class trunks. In this case, the amount of DSCP_EF is fixed to 1400 flows. From Table 1, the optimal value to minimize the voice traffic blocking probability for DSCP_AF and DSCP_BE trunks exists in which both DSCP_EF, DSCP_AF, and DSCP_BE trunk blocking probability becomes less then 10^{-4}. It demonstrates that the QoS requirements are well satisfied by using our proposed scheme. Also, the same results are obtained in case where the capacity of each of high-class trunks is varied as time goes by.

Table 1. Voice Traffic Blocking Probability

M_L	$P_{loss}(H)$	$P_{loss}(L)$
2	1e-06	0.798
4	1e-06	0.1
6	1e-06	0.098
8	1e-06	0.001
10	1e-06	0.0000429
12	1e-06	0.096
14	1e-06	0.1098
16	1e-06	0.689

5 Conclusion

We present the end-to-end QoS guaranteed voice traffic multiplexing scheme between VoIP access routers using differentiated services (DiffServ). At ingress router, the newly defined RTP/UDP/IP packets, namely, L_packet (long packet) are multiplexed according to same destination egress router and same DSCP. To prove the validity of our scheme, we have described the network model developed here to analyze the blocking performance. Also, we analyzed the model using M/G/1 with HOL-NPR queueing system. Finally, we have presented numerical results using the analysis. Through the results, the proposed scheme is shown to be very efficient to guarantee the QoS requirements over VoIP.

References

1. Floyd S, "Notes on CBQ and Guaranteed Service", ftp://ftp.ee.lbl.gov/papers/guarantedd.ps, July 1995.
2. Katsuyoshi IIDA, Tetsuya TAKINE, Hideki SUNAHARA, and Yuji OIE, "Delay Analysis for CBR traffic in static-priority scheduling: single-node and homogeneous CBR traffic case", IEEE SPIE'97.
3. Katsuyoshi IIDA, Kenji KAWAHARA, Tetsuya TAKINE, and Yuji OIE, "Performance Evaluation of the Architecture for End-to-End Quality-of-Service Provisioning", IEEE Communication Magazine, April 2000.
4. Giuseppe Bianchi, Antonio Capone, and Chiara Petrioli, "Throughput Analysis of End-to-End Measurement-Based Admission Control in IP", IEEE INFOCOM'2000.
5. Tohru HOSHI, Keiko TANIWAQA, ad Koji TSUKADA, "Voice Stream Multiplexing between IP Telephony Gateways", IEICE Trans. INF. & SYST., Vol, E82-D, No. 4, April 1999.
6. Hassan Naser, and Alberto Leon-Garcia, "Voice over Differentiated Services", IEEE draft, draft-naser-voice-diffserv-eval-00.txt, December 1998.
7. Ross, K.W., "Multiservice Loss Models for Broadband Telecommunication Networks", Sprnger, Berlin, 1995.
8. Cooper, *Introduction to Queueing Theory*, CEEPress, Washington, D.C., 3rd edition, 1990.