Efficient QoS Scheme for Voice Traffic in Converged LAN

Yoon Suk Choi*, Jong-Kwon Lee, Tag Gon Kim, and Kyou Ho Lee
Department of Electrical Engineering and Computer Science,
Korea Advanced Institute of Science and Technology (KAIST),
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea
E-mail : yschoi@smslab.kaist.ac.kr
Phone : +82-42-869-5454
Fax : +82-42-869-8054

February 28, 2003

Abstract

This paper presents a policy for guaranteeing or supporting voice service in converged LAN with Weighted Fair Queueing (WFQ). How to set the weight of voice traffic is related to how many voice users are served with the minimum delay in routers. This paper next presents two mechanisms,active rerouting and selective dropping mechanisms, to cope with temporary overload condition. Active rerouting can be implemented with active network technology. Selective dropping discards packets to avoid burst losses of packets toward a certain subnet. Various simulation results show that the proposed QoS mechanisms in a converged LAN environment can offer the voice service of good quality.

Keywords: QoS, voice traffic, selective dropping, active rerouting, converged LAN

1 Introduction

For the last decade, many researchers have tried to integrate voice and data transmission on the same networks. Voice service over the packet networks requires the same specific quality as Public Switched Telephone Networks (PSTN). However, the volume of voice and data traffic in converged networks has heavily increased, and multimedia services and data applications need more bandwidths. Because the network bandwidth is too limited to satisfy all their demands, it is difficult to guarantee quality of voice service. To fulfill voice service quality, it is very important to consider new problems caused in converged networks and to manage network resources efficiently to solve these problems.

Several challenging problems need to be solved to achieve the same level of voice quality offered by the traditional PSTNs in the context of packet networks [1]. Voice traffic is very sensitive to delay, jitter and loss. First, delay is the most important issue. End-to-end delay for voice service must be bounded under 150 ms for good voice quality, but delay in the converged networks cannot be predictable. Therefore, minimizing end-to-end delay is the primary issue to guarantee voice quality. Second, jitter must be considered to playback the original sound. Jitter is a delay variation between adjacent voice packets and caused by network conditions and multiplexing. Hence each packet which periodically departs from source host cannot arrive at destination host in the same period. This jitter needs to be corrected by the receiving side. However, if the voice packets are played as soon as they are received, the original speech can become unintelligible [2]. The third issue is related to packet losses. Packet losses occur in routers and degrade the service quality. In addition, packet reordering is needed at destination host. Because voice packets from source host can be routed to different paths, each packet can arrive out of order at destination host. Therefore, it is required to reorder arrived voice packets.

Voice service over the converged network needs new mechanisms for a certain level of quality. From the viewpoint of voice service application, the de-jittering mechanism and the reordering mechanism can be implemented on the application level. But delay issue and loss problem should be solved on the network level. As a result, new mechanisms to guarantee voice service quality–delay and loss– in the converged networks should be considered on the network level.

We expect the voice service on the converged networks will be prevalent in the LAN environment because LAN is the closest network to end-users. In this paper, we introduce the network resource management policy for voice service quality in the converged LAN environment. With this policy, converged LAN managers can support voice service users without inferior voice quality. Then, this paper proposes two mechanisms–active rerouting and selective dropping– to support worst case that the number of users exceeds the threshold of active sessions. With active rerouting and selective dropping, end-to-end delay can be bounded under 150 ms and more voice users can be served.

Active rerouting can be easily implemented by active networking. Active networking is a promising technology that allows us to control the behavior of network nodes by programming them to perform advanced operations and computations [3]. Active networks can minimize the amount of global agreement and support dynamic modification of aspects of the network. Therefore, on-the-fly experimentation on the base platform is supported by active networks [4]. We use the programmability of active networks for active rerouting.

The rest of this paper is organized as follows. In section 2, we briefly discuss related work. Section 3 describes the characteristic of voice traffic, and the delay problem in converged networks. We introduce policy for guaranteeing voice service quality in converged LAN in section 4 and two mechanisms–active rerouting and selective dropping– for supporting voice quality under overload condition in section 5. The simulation of active rerouting and selective dropping using DEVSim++ follows in section 6 to validate the functionality and evaluate the performance. Section 7 concludes this paper.

2 Related Work

Many new architectures and mechanisms have been researched for converged networks to integrate voice and data transmission. The integrated services (IntServ) approach focuses on individual packet flows. In this architecture, each flow can request specific levels of service from the network and the network grants or rejects the flow requests, based on the availability of resources and the guarantees provided to other flows. However, the acceptance of IntServ has been limited due to scalability and manageability problems. Because IntServ requires routers to maintain control and forwarding state for all flows passing through them [5].

An approach similar to the policy of this paper was studied for guaranteeing voice service quality [6]. It uses weighted fair queueing and G.726. However, it does not use accurate voice source model but the aggregated voice traffic. It also does not consider silence compression. Therefore, there is limitation to maximize the number of voice users.

Random dropping [7] gives somewhat fairness when there are too many nodes. If too many nodes send packets to their destinations, a series of packets from a node will not be dropped when a link is congested. However, if there is small number of nodes, random dropping does not guarantee fairness to nodes. So, we focus on selective dropping. Many rules of selective dropping are investigated for multimedia services. [8] uses priority for selective dropping. Video frames are encapsulated into ATM cells with different discard priorities for different frames. Though this method is simple, there is a starvation problem all the time. To give the fairness to RED [7], Flow RED [9] is suggested. Flow RED solves fairness issues by maintaining thresholds and buffer occupancies for each active flow (per-connection information). It is obviously good method for fairness problem but it has high complexity and too much overhead to maintain per-flow information. The rule of our selective dropping is based on subnet information. This rule solves the fairness problem with the packet count to a certain subnet. Because it maintains per-subnet information not per-flow information, it has low complexity and reduced overhead.

In QoS (Quality of Service) networks, the dynamic rerouting algorithm with server-based QoS routing was proposed [10]. If a new flow request arrives at some node and its QoS requirement cannot be satisfied at this time, the sever checks whether it can be accepted by rerouting some flow or not. However in converged networks, each node is not able to request an alternate path to server due to the delay problem.

Adaptive routing [11] was also suggested for routing. It uses a random routing path searching method. It selects random destination in a period to find several paths to destination. When congestion or link failure occur, it uses the path information already gathered. However, adaptive routing has less robustness and slow convergence to find paths to destination nodes. On the other hand, our active rerouting is based on cluster-based rerouting path searching. It is sensitive to local link changes and has reduced overhead to find rerouting path. Adaptive routing uses source routing to destination, but active rerouting uses encapsulation. However, source routing makes packet size larger than original voice packet size. It will increase end-to-end delay. Encapsulation has smaller packet size than source routing but there is overhead to encapsule voice packets.

3 Voice Traffic

3.1 Factors of Voice Traffic Delay

Excessive delay of voice communication degrades the service quality. ITU-T recommends the end-to-end delay under 150 ms. ETSI TIPHON is also working on defining levels of end-to-end quality of service for IP telephony networks. TIPHON suggests the end-to-end delay under 150 ms for best quality, under 250 ms for high quality, and under 450 ms for medium quality. Therefore it is very important to reduce the end-to-end delay for voice service quality.

Voice frames generated from voice coders are injected into converged networks through RTP, UDP (User Datagram Protocol), IP, and ethernet layers. As getting down each layer, headers of each layer are attached to voice frames. Then voice packets explorer networks to arrive at the destination host. In the networks, voice packets undergo propagation delay, transmission delay, and processing delay like data packets. The factors of voice traffic delay can be classified as below.

- Delay by voice coder algorithms
- Delay by queue scheduling in routers
- Delay by network congestion

A variety of voice coders are used in IP telephony. They have different frame size and granularity each other. Table 1 shows bit-rate and granularity of some voice coders. Granularity and bit-rate decide algorithmic delay of a voice coder. Different schedulers in routers determine queue scheduling delay. Delay also increase under network congestion. To minimize delay of voice traffic, we must choose the best solution in each cases. We will discuss these problems in section 4.

Table 1: Bit-rate and Granularity of Voice Coders

Coder	Bit-rate($Kbps$)	Granularity(ms)
G.723.1	5.3/6.3	30
G.726	16/24/32/40	0.125
G.728	12.8/16	0.625
G.729	8	10

3.2 Voice Source Model

Although a voice coder generates frames at a certain rate, voice source does not generate voice at a constant rate. Human speech consists of an alternating sequence of talk spurt, typically averaging 0.4 - 1.2 sec in length, followed by silence intervals, averaging 0.6 - 1.8 sec in length [12].

Voice source model is generally regarded as ON-OFF pattern. ON is talk spurt and OFF is a silence interval. The traditional assumption of ON-OFF voice model is that ON and OFF patterns have exponential distribution.

But, a research recently report that the traditional assumption does not always fit well with the audio session [13]. They applied the G.729B Voice Activity Detection (VAD) and NeVoT silence detector to some recorded telephone conversations. They concluded that the exponential model can be used for a first-hand performance estimate, but a more precise model is needed in certain settings and where high precision is required, for example when a strict Service Level Agreement (SLA) is to be determined.

Figure 1 shows the spurt/gap distribution produced by G.729B VAD when averaged over many conversations. They concluded that the exponential model is apparently not a good fit for the gap distribution, and depending on the requirement, exponential model may be considered an inadequate fit for the spurt distribution as well. We tried to find similar gap distribution with pareto distribution, and we concluded pareto distribution with k=0.1 and a=1.25773 ($x > k$) is the most similar to the gap distribution of Figure 1.

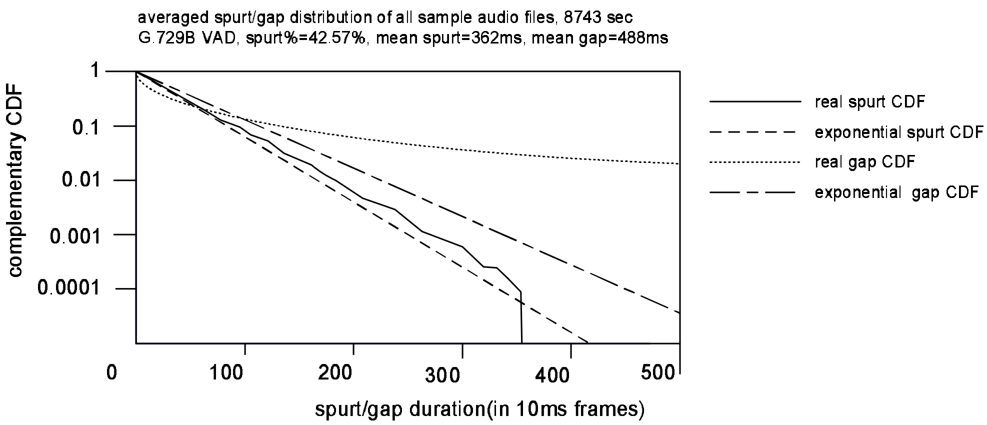


Figure 1: Spurt/Gap Distribution after Averaging over Many Conversations [13]

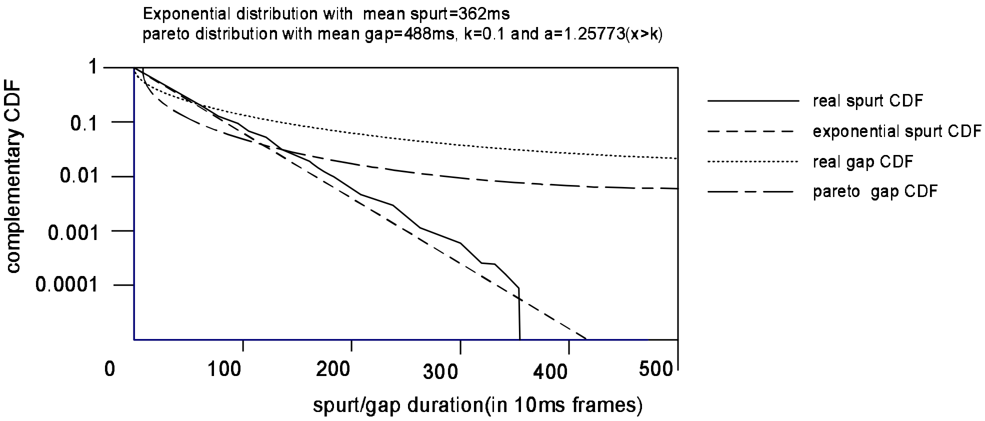


Figure 2: Spurt/Gap Distribution Used in Experiment

Figure 2 shows the gap distribution we found. However, we decided just to use exponential distribution for spurt distribution. We set that call duration has exponential distribution with mean 180 seconds and that inter-calling time has exponential distribution with mean 60 seconds. In summary, Table 2 shows the distribution of each time duration.

Table 2: Distribution for Each Time Duration

Time Duration	Distribution
Call Duration	Exponential Distribution with mean 180 s
Inter-calling Time	Exponential Distribution with mean 60 s
Spurt	Exponential Distribution with mean 362 ms
Silence Interval	Pareto Distribution with mean 448 ms

4 Policy for Supporting QoS of Voice Traffic in Converged LAN

4.1 Selection of Voice Coder

As shown in section 3, the kinds of voice coders used in VoIP applications are various. We must select the best voice coder to minimize the end-to-end delay. The comparative study of various coders had been in [14] and [15]. Those papers analyzed the speech quality of various voice coders based on bit-rate, delay, and complexity.

Speech quality of a voice coder is a function of bit rate, complexity, delay, and bandwidth. Hence, it is important to realize that there is a strong interaction between these attributes. For example, low bit-rate coders tend to have more delay than high bit-rate coders [14]. Among ITU-T recommendation, G.723 and G.729 have low bit-rate with good voice quality. However G.729 has lower frame delay than G.723, so we can conclude that G.729 is the best candidate which has low bit-rate and low delay in converged LAN.

G.729 has many improved derivatives. G.729 Annex A is a reduced complexity version of the original G.729. While G.729 requires about 20 MIPS for coding and 3 MIPS for decoding, G.729A requires about 10.5 MIPS for coding and 2 MIPS for decoding using TMS320.C54 DSP. G.729 Annex B defines a low bit-rate silence compression scheme designed and optimized to work in conjunction with the full version of G.279. Silence compression scheme can reduce the usage of bandwidth by half. Fortunately G.729 Annex A and B (G.729AB) has the characteristic of G.729 Annex A and Annex B. Then we can conclude G.729AB is the best solution for selecting voice coder in converged LAN.

4.2 End-to-End Delay of G.729AB

For minimum delay, we select G.729AB among a variety of coders. The end-to-end delay of G.729AB follows below.

$$Delay_{G.729AB} = D_{en} + D_{alg} + D_{prop} + D_{proc} + D_{tx} + D_{deji} + D_{de} \quad (1)$$

Among these delays, D_{prop} (propagation delay) and D_{proc} (processing delay) are negligible. G.729AB has D_{en} (encoding delay), D_{alg} (algorithmic delay), D_{deji} (de-jittering delay) and D_{de} (decoding delay) like below except D_{tx} (transmission delay).

- $D_{en} = 10\ ms$
- $D_{alg} = 5\ ms$
- $D_{deji} = 40\ ms$
- $D_{de} = 3\ ms$

These delays are fixed during voice communication. We can define $D_{fixed} = D_{en} + D_{alg} + D_{deji} + D_{de} = 58\ ms$. Although D_{fixed} is constant, D_{tx} is variable. Hence, how to minimize D_{tx} is the next key of the voice service quality problem in converged LAN.

4.3 Policy: Weight for Voice Traffic in WFQ

The second consideration of delay factors of voice service is a scheduler in a router. Different data packets and voice packets travels to different destination hosts in converged LAN. However current routers support best-effort service in networks, so voice packet and data packets are under same processing routines. Hence, they have the same delay in converged networks according to the volume of data packets and voice packets.

However, voice packets cannot be tolerable with the same delay of data packets as already discussed. For the different service of data packets and voice packets, Weighted Fair Queueing (WFQ) is used. WFQ differently serves packets according to the class of each packet. For WFQ, the best node would ideally allocate a fair share of the available bandwidth to each packet.

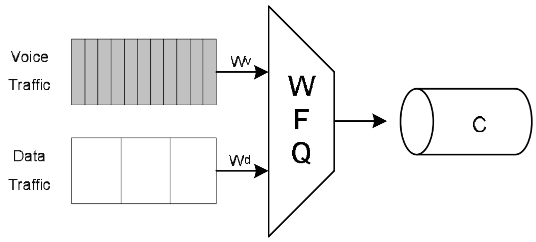


Figure 3: Output Link of Router Using WFQ

With the weight of voice traffic in WFQ, we can calculate the bandwidth that voice traffic can use. We assume the maximum number of hops in the converged LAN is 9 and all voice traffic is generated from G.729AB to minimize D_{tx} as already discussed above.

Figure 3 shows an output link of router using WFQ for voice traffic and data traffic. Voice packets are served according to the weight set for voice traffic.

To calculate the service time delay of a voice packet in WFQ, we separate one case that there is no voice packets and the other case that there are (n-1) voice packets.

Parameters used in calculation are like below.

- w_v : *weight for voice traffic*
- w_d : *weight for data traffic*
- s_v : *the volume of voice traffic*
- s_d : *the volume of data traffic relative to* s_v
- C : *the capacity of a output link*

The sum of weights, w_v and w_d is 1. ($w_v = 1 - w_d$)

1. If there is no voice packet in WFQ,

The delay of an arrived voice packet has the time to serve voice packet as minimum service time. It also has the time to serve voice packet and data packet as maximum service time.

$$\begin{aligned} Delay &= \left[\frac{s_v}{C}, \frac{s_v}{C} + \frac{s_d}{C}\right] \\ &= \left[\frac{s_v}{C}, \left(1 + \frac{w_d}{w_v}\right) \cdot \frac{s_v}{C}\right] \\ &= \left[\frac{s_v}{C}, \frac{1}{w_v} \cdot \frac{s_v}{C}\right] \end{aligned} \quad (2)$$

Therefore, the maximum delay of an arrived voice packet is $\left(\frac{1}{w_v} \cdot \frac{s_v}{C}\right)$.

2. If there are (n-1) voice packet in WFQ,

   From equation 2, we can calculate the delay of an arrived voice packet in WFQ.

$$Delay = \left[\frac{n}{w_v} \cdot \frac{s_v}{C} - \frac{s_d}{C}, \frac{n}{w_v} \cdot \frac{s_v}{C}\right] \quad (3)$$

   Therefore, the maximum delay of an arrived voice packet is $\left(\frac{n}{w_v} \cdot \frac{s_v}{C}\right)$.

With equation 2 and equation 3, a manager of a converged network can decide the weight for supporting the voice service quality of n users. At the first, the maximum delay at a router must be bound to determine the number of users, n. We already assume that the max hop of the converged LAN is 9, so the maximum delay at a router must be bound under the granularity of G.729AB, 10 ms. With this method, a manager of a converged LAN can set accurate weight for voice traffic related to the number of users. Then equation 4 can follow.

$$Delay \quad = \quad \frac{n}{w_v} \cdot \frac{s_v}{C} \; < \; T_{granularity} \tag{4}$$

We get equation 5 from modifying equation 4. However we can derive equation 6 from the idea that users do not generate voice at the same time. Equation 6 uses mean values of spurt, silence, hang on, and hang off.

$$n \quad < \quad \frac{w_v \cdot C}{s_v} \cdot T_{granularity} \tag{5}$$

$$w_v \quad > \quad \frac{n \cdot s_v}{C \cdot T_{granularity}} \cdot \frac{m\,(spurt)}{m\,(spurt) + m\,(silence)} \cdot \frac{m\,(hangon)}{m\,(hangon) + m\,(hangoff)} \tag{6}$$

We can get the number of concurrent active sessions from equation 5 and the maximum number of users from equation 6. The maximum number of users is calculated with the probability that users generates voice packets actively. Table 3 shows the number of active sessions and maximum users in 10 $Mbps$ converged LAN when voice applications use G.729AB and routers use WFQ.

Table 3: The Number of Active Sessions and Maximum Users (WFQ&G.729AB in 10 $Mbps$ Converged LAN)

w_v	Active Sessions	Maximum Users
0.05	9	28
0.10	19	59
0.15	28	87
0.20	38	118

5 Supporting Voice Service under Overload Condition

We introduced the policy for support voice service users in section 4. However a manager of a converged LAN does not know the exact number of active session in a future. Hence, if the number of users temporarily exceeds n, the weight already set for n users will not support voice service. To cope this situation, we suggest two mechanism, Active Rerouting (ARR) and Selective Dropping (SD).

5.1 Active Rerouting Mechanism

ARR uses the programmability of Active Networks (AN) [16]. In ARR, all routers need not be active nodes. Some nodes are general routers and the others are active routers. Active routers manage their cluster and gather link state information (link failure, link congestion, etc.) Before activating ARR, active routers share their link state information to make alternate routing table with capsule containing programs and payload in a period time and use encapsulation to avoid loop-back. The algorithm follows.

5.1.1 Active Rerouting Algorithm

1. General routers use routing protocols (RIP or OSPF) to make routing table.

2. Active routers gather the link state information of their clusters. They can increase or decrease link cost as link state changes. When traffic is heavy, active routers will increase the link metric.

3. Active routers make alternate routing tables by using link state information (link congestion, link failure etc.). They exchange their information with capsule containing information gathering programs.

4. When a link is congested or failed, active routers use the alternate routing table to make alternate path candidates.

5. Active routers select one path among rerouting path candidates. Our path selection method follows in next section.

6. Active routers pass voice packets with encapsulation/decapsulation using the Active Network Technology.

7. If rerouting fails, routers discard a voice packet.

ARR is based on a cluster structure. Each active router collects the link information of local clusters, so ARR is sensitive to local link changes. And overhead to search rerouting paths to destinations is reduced because all active routers exchange rerouting information each other. Figure 4 shows the ARR architecture.

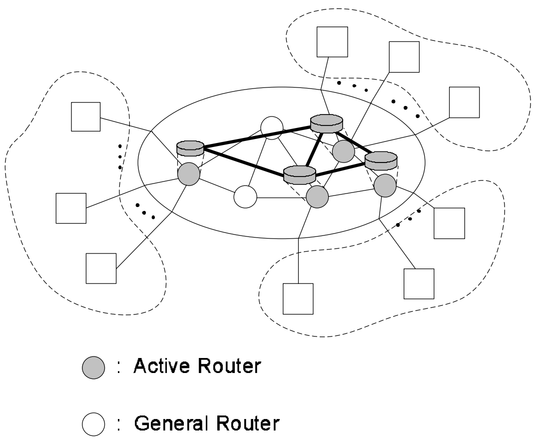


Figure 4: Active Rerouting Architecture

5.1.2 Alternate Path Selection

If a link is congested, two requisites must be checked to find alternate path.

`Check I`

Select active router(AR_i)s such that
$RT[\ DestSubnet\ ] \neq RT[\ AR_i\ ]$

`Check II`

> $\forall AR_i$ *satisfying* `Check I`,
> *select an active router(AR_i) has minimum cost such as*
> *metric[AR_i] + metric$_i$[DestSubnet]*

In Check I, RT means alternate routing table of an active router. When a link is congested, an active router tries to find other links through which voice packets can go to other active routers.

In Check II, RT_i means alternate routing table of the active router i(AR_i). Among active routers satisfying Check I, ARR selects an active router which has minimum cost such as the sum of the metric from current active router to AR_i and the metric from AR_i to a certain destination subnet.

5.2 Selective Dropping Mechanism

We derive selective dropping scheme from burst loss problem of voice packets and the compensation technique of voice application. If loss occurs in networks, TCP will try to retransmit but voice application cannot retransmit voice information due to time sensitivity. Hence, rather than retransmission, voice applications try to compensate the lost voice information from adjacent packets. It can reduce abuse of network resource. However if loss is burst, destination application is not able to regenerate original sound from too far packets. Therefore, many researches have studied to reduce or avoid network burst loss.

We make selective dropping scheme appropriate to converged LAN. It avoids burst voice packet loss of each subnet under overload condition. Its objective is fairness for each subnet, so the selective dropping probability reflects fairness.

$$P(SD) = P_a \cdot P_b \tag{7}$$

$$P_a = \frac{AS - th_{as}}{th_{max} - th_{as}} \tag{8}$$

$$P_b = \frac{n_{Dest}}{\sum_i n_i} \tag{9}$$

The probability P_a is in proportion to the number of active sessions that generate voice packets actively. And P_b is in proportion to n_{Dest}, the number of packets going to subnet i through the router. This probability gives users fairness.

Discussion about effectiveness of linearity of P_a and fairness of P_b are needed. If the shape of P_a is like a square, end-to-end delay will increase and loss rate will decrease. However, if the shape of P_a is like a square root, end-to-end delay will decrease and loss rate will increase. Figure 5 shows the expected result.

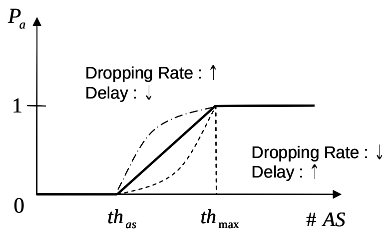


Figure 5: Linearity of P_a

6 Performance Evaluation

6.1 Validation of Policy

We run simulations 5 times of each case for validation of policy. Figure 6 is topology we used and we assume the maximum hop is 9 in 10 $Mbps$ converged LAN. There is n users in network.

(a) of Figure 6 is the simplest case and (b) of Figure 6 is the worst case of the converged LAN. Since aggregation takes place at each router, the arriving traffic will be bursty at each node. Hence, the delay due to bursts occurs in each node on the end-to-end path.

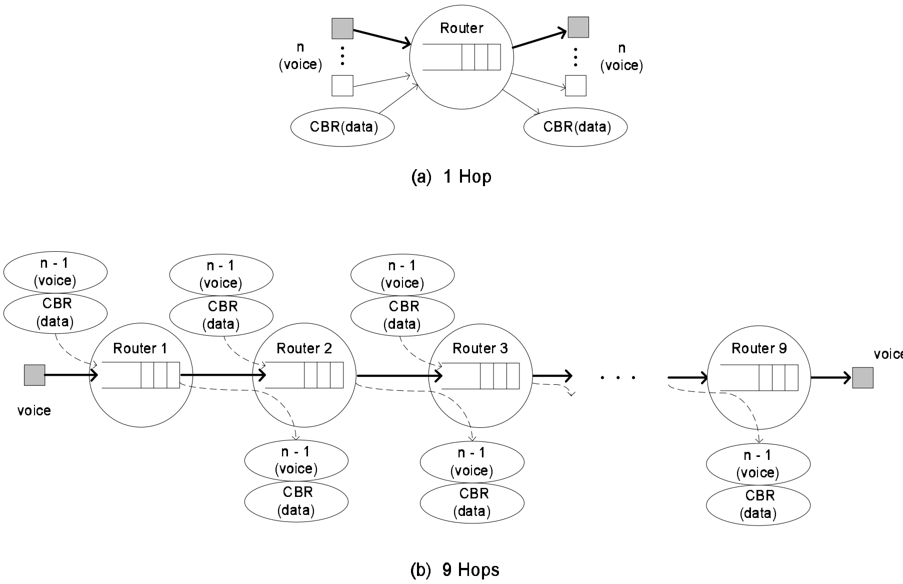


Figure 6: Queueing Scenario for Policy Validation

Table 4 is the simulation result for good quality of voice service, 150 ms end-to-end delay. In Table 4, we found that WFQ and G.729AB cannot support the maximum users which is theoretical value from equation 6. The theoretical max users do not constantly generate voice packets equal to the threshold(Active Session in Table 3) because the number of active sessions sometimes exceeds the threshold. Therefore the end-to-end delay cannot be bound under 150 ms.

Table 4: Max Users (WFQ&G.729AB): 1 Hop and 9 Hops in 10 $Mbps$ LAN

w_v	Active Session	Maximum Users	1 Hop	9 Hops
0.05	9	28	27	22
0.10	19	59	43	41
0.15	28	87	65	56
0.20	38	118	77	70

However, we can conclude the converged LAN can always support max users of 9 hops in Table 4 because 9 hop LAN is the worst case. In the next section, we apply selective dropping and active rerouting to support more users with good quality of voice service under overload condition.

6.2 Performance of Selective Dropping

Queueing scenario is same as (b) of Figure 6, the worst case in converged LAN. Table 5 is the simulation result for good quality of voice service, 150 ms end-to-end delay and 3% loss rate.

Table 5: Max Users (+Selective Dropping): 9 Hops in 10 $Mbps$ LAN

w_v	Active Session	1 Hop	9 Hops	Selective Dropping (9Hops)
0.05	9	27	22	24
0.10	19	43	41	46
0.15	28	65	56	70
0.20	38	77	70	90

There is the increase of max users when selective dropping is used rather than when just WFQ and G.729AB are used. Current routers are able to use selective dropping easily for supporting voice service. Simulation results show the availability of selective dropping to be implemented in current data networks.

6.3 Linearity and Fairness of Selective Dropping Probability

In this section, we show the simulation result of linearity of P_a and fairness of P_b. Figure 7 shows the effectiveness of linearity of P_a. As discussed in section 5, the shape of a square make long end-to-end delay, about 0.7 s, and the shape of a square root make hight loss rate, about 10 %. These values exceed the

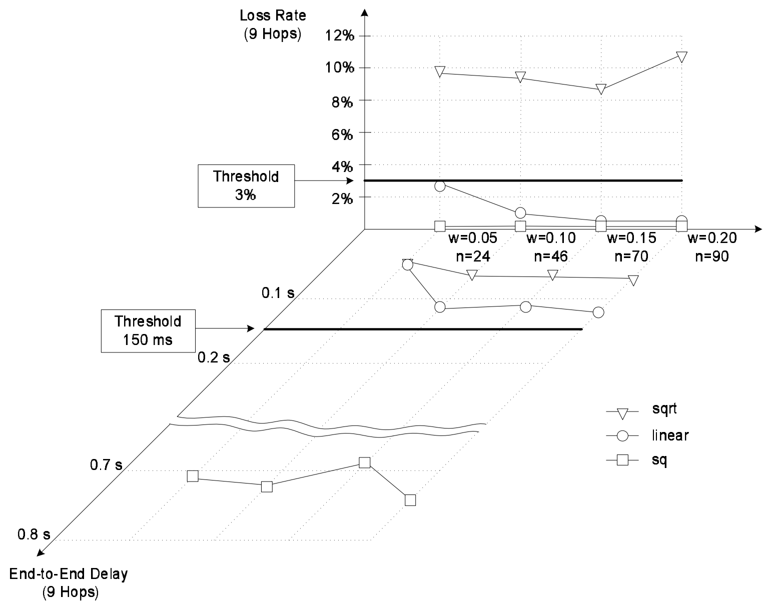
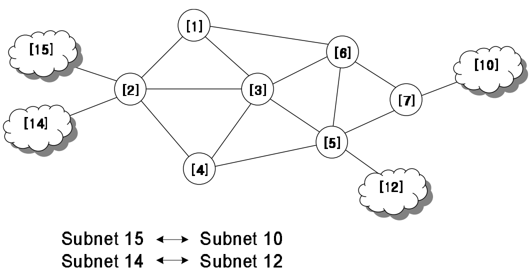


Figure 7: Selective Dropping: Effectiveness of Linearity of P_a



	Users	Dropping Rate
w=5		
Subnet15 ↔ Subnet10	10	**0.16%**
Subnet14 ↔ Subnet12	15	**0.18%**
w=10		
Subnet15 ↔ Subnet10	16	**0.39%**
Subnet14 ↔ Subnet12	30	**0.42%**
w=15		
Subnet15 ↔ Subnet10	22	**0.52%**
Subnet14 ↔ Subnet12	50	**0.52%**
w=20		
Subnet15 ↔ Subnet10	32	**0.47%**
Subnet14 ↔ Subnet12	60	**0.46%**

Figure 8: Selective Dropping: Fairness of P_b

threshold, each 150 ms and 3 %. As a result, Figure 7 tells that the linearity is the trade-off point between loss rate and end-to-end delay.

Figure 8 shows the fairness of P_b. In Figure 8, n voice users in subnet 15 are willing to communicate with n users in subnet 10 and m users in subnet 14 want to communicate with m users in subnet 12. As the weight of voice traffic varies, the number of users is changed. However n is not same as m. Simulation result shows that each conversation has similar loss rate.

6.4 Performance of Active Rerouting

Under overload condition, active rerouting and selective dropping will activate. If the number of active session every 10 ms exceeds th_{as}, an active router reroutes packets. If it fails rerouting, packets will be dropped according to the selective dropping probability.

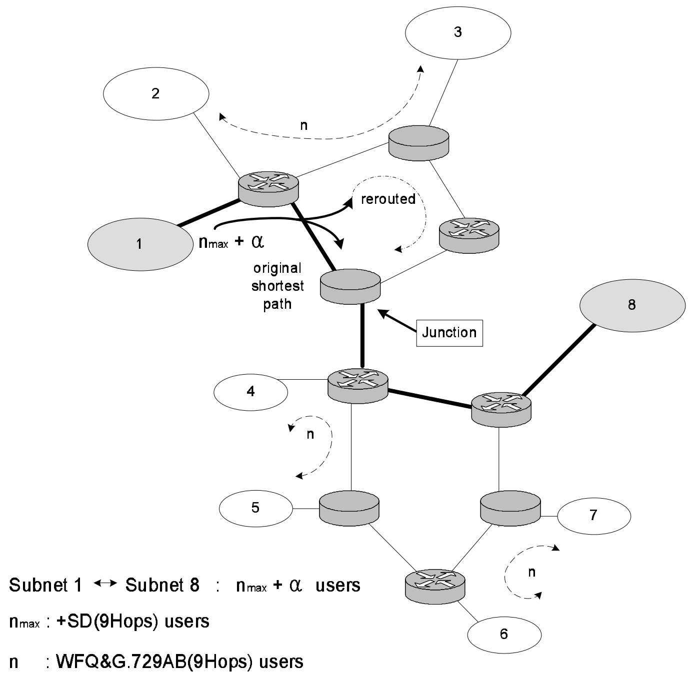


Figure 9: Active Rerouting Scenario

We used the topology of Figure 9. n_{max} users in subnet 1 want to communicate with n_{max} users in subnet 8. The hop count of rerouting path will be 6, 7, or 9 because there are 2 rerouting branches. The important point of this topology is a junction in the middle of routing path. We regard there is one junction on the rerouting path at least. Other links support users that WFQ&G.729AB can support.

Table 6: Max Users (+Active Rerouting): 9 Hops in 10 $Mbps$ LAN

w_v	WFQ & G.729AB Active Session	WFQ & G.729AB (1 Hop)	WFQ & G.729AB (9 Hops)	+ SD (9Hops)	+ SD & ARR (9Hops)
0.05	9	27	22	24	25
0.10	19	43	41	46	52
0.15	28	65	56	70	75
0.20	38	77	70	90	97

Table 6 shows the simulation result and the comparison with WFQ&G.729AB, selective dropping, and active rerouting. Figure 10 is the graphical view of Table 6 with WFQ&G.726 [6]. We can find that active rerouting does not support more voice users as well as selective dropping.

As the condition of converged networks is overloaded, networks cannot serve the voice users with good quality. However Figure 10 gives us the tolerable end-to-end delay and loss rate of good quality, if selective dropping and active rerouting are used. Hence, it is a good choice to use active rerouting and selective dropping mechanisms for supporting quality of voice service under overload condition.

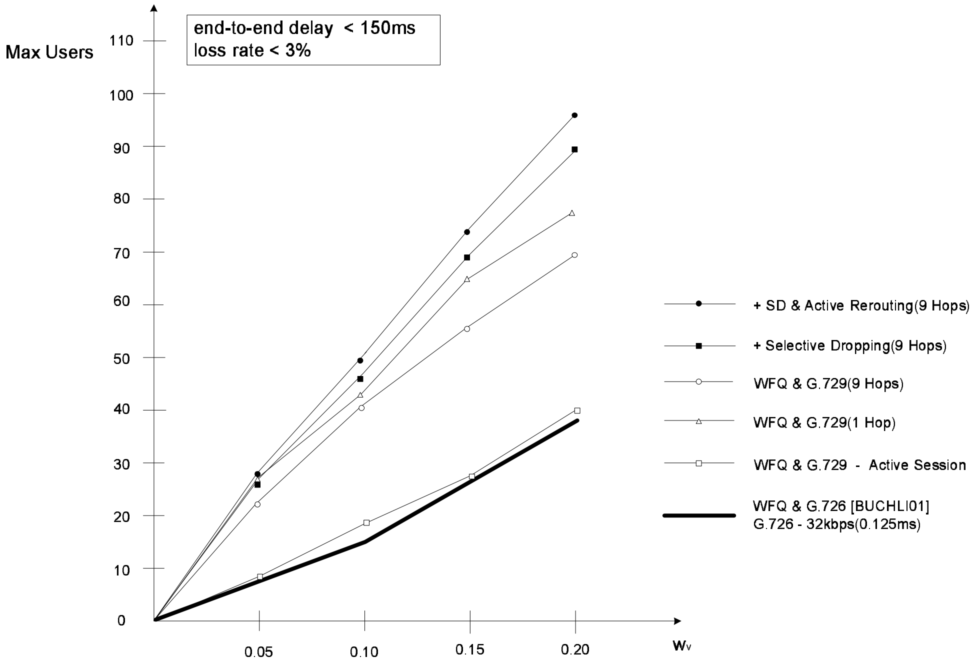


Figure 10: Simulation Result

7 Conclusion

This paper focuses on the delay and studies the solution of delay problems. For the minimum of fixed delay in a converged network, we select G.729AB voice coder providing lower delay and better quality than others. We also use WFQ for the minimum queueing delay in a router and analyze the delay time of a voice packet in WFQ. From this analysis, we presents the policy for supporting quality of voice service in converged LAN environment. A manager of a converged LAN can configure network resources dynamically with this policy and users are able to get good quality of voice service in a converged LAN. This paper next proposed two mechanisms for voice service under overload condition. Probability rule used in selective dropping was made to provide fairness in converged LAN.

The simulation results show the validation of policy and the effectiveness of active rerouting and selective dropping. Under overload condition, the simulation results show that active rerouting and selective dropping give reduced delay less than 150 ms, and converged network can support more voice service users.

Before long, voice service and data service will be converged into one network. And we expect the voice service over the network will be prevalent on the LAN environment which is the closest network to end-users. Unfortunately current network cannot support voice service efficiently. Consequently it is expected that active rerouting and selective dropping will support more users with good quality of voice service over the converged LAN.

References

[1] A. Barberis, C. Casetti, J.C. De Martin, M. Meo, *A Simulation Study of Adaptive Voice Communications for IP Networks*, Computer Communications, Elsevier, Vol. 24, No. 9, pp 757-767, May 2001,

[2] O. Hersent, D. Gurle, J.P. Petit, *IP Telephony*, Addison-Wesley, 2000.

[3] R. Boutaba, A. Polyrakis, *Projecting Advanced Enterprise Network and Service Management to Active Networks*, IEEE Network, Vol. 16, No. 1, January/February 2002.

[4] K.L. Calvert, *Architectureal Framework for Active Networks version 1.0*, 1999.

[5] C. Dovrolis, P. Ramanathan, *A Case for Relative Differentiated Services and the Proportional Differentiation Model*, IEEE Network, 13(5):26-34, September/October 1999

[6] M. Buchli, D.De Vleeschauwer, J. Janssen, A.V. Moffaert, G.H. Petit, *Resource allocation and management in DiffServ networks for IP telephony*, 11th International workshop on on Network and Operating Systems support for digital audio and video, January 2001

[7] S. Floyd, V. Jacobson, *Random Early Detection Gateway for Congestion Avoidance*, IEEE/ACM Transaction on Networking, August, 1993.

[8] A. Mehaoua and R. Boutaba, *Performace analysis of cell discarding techniques for best effort video communications over ATM*, Comp. Networks and ISDN Sys., vol. 29, no. 17-18, pp.1999-2019, Feb. 1998.

[9] D. Lim and R. Morris, *Dynamics of Random Early Detection*, Proc. of SIGCOMM, pp. 127-137, 1997.

[10] T. Ikenaga, K. Kawahara, Y. Oie, *Effectiveness and Issues of Rerouting Algorithms for QoS Networks*, IEICE Transactions on Communications, Vol.E84-B, No.6, pp.1553-1560, June 2001.

[11] G.D. Fatta, S Gaglio, G.L. Re, M. Ortolani, *Adaptive Routing in Active Networks*, IEEE Third Conference on Open Architecture and Network Programming. Tel Aviv, March, 2000.

[12] M. Schwartz, *Broadband Integrated Network*, Prentice Hall, 1996.

[13] W. Jiang, H. Schulzrinne, *Analysis of on-off patterns in VoIP and their effect*, In the 9th IEEE International Conference on Computer Comm. Networks, 2000.

[14] Richard V. Cox, Peter Kroon, *Low Bit-Rate Speech Coders for Multimedia Communication*, IEEE Communications Magazine, December. 1996.

[15] Richard V. Cox, *Three New Speech Coders from the ITU Cover a Range of Applications*, IEEE Communications Magazine, September. 1997.

[16] K.L. Calvert, M.B. Doar, E.W. Zegura, *Modeling Internet Topology*, IEEE Communications Magazine, June, 1997.