Real-time Voice Communication over the Internet Using Packet Path Diversity

Yi J. Liang, Eckehard G. Steinbach, and Bernd Girod
Information Systems Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA 94305, USA
{yiliang, steinb, bgirod}@stanford.edu

ABSTRACT

The quality of real-time voice communication over best-effort networks is mainly determined by the delay and loss characteristics observed along the network path. Excessive playout buffering at the receiver is prohibitive and significantly delayed packets have to be discarded and considered as late loss. We propose to improve the tradeoff among delay, late loss rate, and speech quality using multi-stream transmission of real-time voice over the Internet, where multiple redundant descriptions of the voice stream are sent over independent network paths. Scheduling the playout of the received voice packets is based on a novel multi-stream adaptive playout scheduling technique that uses a Lagrangian cost function to trade delay versus loss. Experiments over the Internet suggest largely uncorrelated packet erasure and delay jitter characteristics for different network paths which leads to a noticeable path diversity gain. We observe significant reductions in mean end-to-end latency and loss rates as well as improved speech quality when compared to FEC protected single-path transmission at the same data rate. In addition to our Internet measurements, we analyze the performance of the proposed multi-path voice communication scheme using the *ns* network simulator for different network topologies, including shared network links.

Keywords

Packet path diversity, multi-stream transmission, multi-path transmission, adaptive playout scheduling, multiple description coding, forward error correction, voice over IP.

1. INTRODUCTION

High quality real-time voice communication over the Internet requires low end-to-end delay and low loss rate. Best effort networks such as today's Internet, however, are characterized by highly varying delay and loss characteristics that contradict our Quality-of-Service (QoS) requirements. One widely accepted way to reduce the effective packet loss observed by the receiver is to add redundancy to the voice stream at the sender. This is possible without imposing too much extra network load since the data rate of voice traffic is very low when compared with other types of data and multimedia traffic.

A common method to add redundancy is forward error correction (FEC), which transmits redundant information of each packet in subsequent packets [6][5][16]. In this sender-based scheme, a lost packet can be recovered from the copies piggybacked in subsequent packets should they be received successfully. In this scheme, loss recovery is performed at the cost of higher latency [5]. In many cases, however, the loss of successive packets is correlated, due to the way packets are dropped as networks get congested and router buffers are becoming full. A packet loss may usually be followed by a burst of loss, which significantly decreases the efficiency of FEC schemes [4]. In order to combat burst loss, redundant information has to be added into temporally distant packets, which introduces even higher delay. Hence, the repair capability of FEC is limited by the delay budget.

Another sender-based loss recovery technique, interleaving, which does not increase the data rate of transmission, also faces the same dilemma. The efficiency of loss recovery depends on over how many packets the source packet is interleaved and spread. Again, the wider the spread, the higher the introduced delay [19].

In this work we look at the problem of reliable voice communication over best-effort networks from a different angle. Instead of restricting our transmission to one network path, we send multiple redundant descriptions of the voice stream over different independent paths and take advantage of their largely uncorrelated loss and delay characteristics. As a result, the probability of a negative disturbance, such as packet erasure or increasing delay, impacting all channels at the same time will be small.

In previous literature, path diversity has been proposed for reliable video communication over lossy networks using multiple state encoding, where odd and even frames of a video sequence are transmitted on different network paths [2]. It has been observed in [2] that for multi-path transmission the end-to-end application sees a virtual average path which exhibits a smaller variability in quality than any of the individual paths. Multi-path transmission also alleviates the problem that the default path determined by the routing algorithm is not optimum, which might be often the case

according to [17].

In the context of delay-sensitive applications such as interactive VoIP, the novelty and key point of this work lies in the fact that we explicitly take advantage of the largely uncorrelated characteristics of the delay variation (also known as *jitter*) on multiple network paths using an adaptive multi-stream playout scheduling technique. Packet loss in such applications is a result of not only packet erasure, but also delay jitter, which greatly impairs communication quality. Due to the stringent delay budget and the need to output speech periodically and continuously, packets experiencing sudden high delay have to be discarded at the receiving end if they arrive later than the scheduled playout deadline (which results in *late loss*). With multi-stream voice transmission along different network paths we have now more freedom to trade off delay, late loss, and speech reconstruction quality. We formulate this tradeoff as a Lagrangian cost function where we can vary the relative importance of these quantities.

The multiple streams to be delivered via different paths are formed by multiple description coding (MDC), which generates multiple descriptions of the source signal of equal importance. These descriptions can be decoded independently at the receiver. If all descriptions are received, the source signal can be reconstructed in full quality. If we receive only a subset of the descriptions, the quality of the reconstruction is degraded, but is still better than the quality resulting from losing all descriptions. Depending on the MDC scheme selected, the overall data rate of the payload does not necessarily increase as a result of transmitting multiple streams. The data rate only increases if we desire redundancy between the multiple streams. A small increase in data rate with the use of FEC has been widely accepted for speech communication and we therefore compare our scheme of transmitting two MDC streams with a standard scheme that uses FEC protected single-stream transmission at the same payload data rate.

In order to maximize the benefits of path diversity we have to select paths that exhibit largely uncorrelated jitter and loss characteristics. Sending streams along different routers from source to destination naturally leads to path diversity which could include streams traversing different ISPs or even streams being sent in different directions around the globe. With today's Internet protocols, the path a packet takes across the Internet is a function of its source and destination IP addresses as well as the entries of the routing tables involved. Selecting a specific path for a packet is largely unsupported in today's infrastructure. As discussed in [2], IPv4 source routing is usually turned off within the Internet for security reasons. More promising is to implement path diversity by means of an overlay network that consists of relay nodes [2],[1]. Here, packets can be sent along different routes as being encapsulated into IP packets that have the addresses of different relay nodes as their destination. At the relay nodes, packets are forwarded to other relay nodes such that the packets from different description streams travel along as few common links as possible. In the context of a peer-to-peer framework, every peer could serve as a relay node for voice traffic, potentially leading to many different paths a voice stream could take from its source to its destination. With the next-generation IP protocol IPv6, the source node has a great amount of control over each packet's route. IPv6's loose source routing (LSR) allows packets to be sent via specified intermediate nodes. This source routing feature of IPv6 will make future implementation of multi-stream transmission with path diversity even simpler.

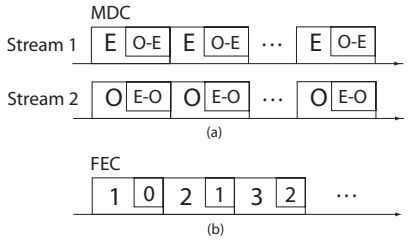

Figure 1: Source encoding: a) MDC; b) single-stream with FEC.

This paper is organized as follows. We first present the employed multiple description coding scheme used to produce two redundant voice streams that can be sent across two different paths. In Section 3, we introduce our receiver playout scheduling algorithm for multiple streams. Section 4 presents multi-path measurements performed in the Internet. Using the measured traces we compare single-path and multi-path transmissions and show that considerable improvements can be obtained for voice transmission with packet path diversity. In Section 5, we analyze the performance of the proposed multi-path voice communication scheme more systematically using a network simulator for different network topologies and varying network load.

2. MULTIPLE DESCRIPTION CODING OF VOICE STREAMS

Various MDC schemes have been proposed for speech coding [11][18][10]. For low complexity, we use the scheme described in [10] to generate the two streams with redundancy at the sender. The basic idea is to quantize the even samples in finer resolution (e.g., PCM, 8 bits/sample) and the difference between adjacent even and odd samples in coarser resolution (e.g., ADPCM, 2 bits/sample), and then packetize them into stream 1. For stream 2, we quantize even and odd samples in the opposite way (Fig. 1(a)). Using this scheme, the redundancy imposed when neglecting packet headers is 25%.

If a packet from one stream, e.g., stream 1, is dropped by the network or discarded at the receiver due to its late arrival, the chances are good that the corresponding packet from stream 2 is successfully received and can be played out if the packet erasure and delay on the two channels are largely uncorrelated. Should that take place, the odd samples of the source signal can be reconstructed in full resolution, while the even samples are reproduced at a coarser resolution. The overall speech quality is degraded with quantization noise, but is still better than the quality when losing both packet.

In order to make a fair comparison to previous work, we compare our proposed packet path diversity voice communi-

cation scheme with a FEC protected single-stream technique [5] at the same payload data rate. In the FEC scheme, the source packet (referred to as primary copy) is coded with the same finer quantization as before, and a secondary copy of the packet is coded with the same coarser quantization and carried by the subsequent packet (Fig. 1(b)). The additional packet header overhead resulting from transmitting multiple streams will be neglected in the following.

3. PLAYOUT SCHEDULING OF MULTIPLE STREAMS

As is mentioned in Section 1, delay jitter is a critical factor in real-time voice applications which obstructs the proper and timely reconstruction of the speech signal at the receiving end. Under the stringent delay requirements, packets could get lost due to their late arrival resulting from excessive network delay. One important functionality to be implemented at the receiver is the playout scheduling of the voice packets, or in other words, setting the time when to play out the received packets. With the existence of delay jitter, the playout scheme greatly affects the tradeoff between loss and latency. We now present our playout scheduling algorithm for multiple streams.

3.1 Playout scheduling of multiple streams

Before the arrival of each packet i, we set the playout deadline for that packet according to the most recent delays we recorded. The playout deadline of packet i is denoted by d^i_{play}, which is the time from the moment the packet is delivered to the network until it has to be played out. It is the total end-to-end delay of packet i (not including the packetization time at the sender), which characterizes the delay of transmission and playout. Table 1 summarizes the basic notation used in the following.

Table 1: Basic Notation.

Notation	Description
$d^i_{S_l}$	Network delay of packet i in stream l
$\{D^k_{S_l}\}$	Sorted order statistics of $\{d^i_{S_l}\}$
d^i_{play}	Total end-to-end delay of packet i
$\mathcal{E}(d^i_{play})$	Average total end-to-end delay of a voice stream
Δd_{S_l}	Average delay reduction for stream l
$\hat{\varepsilon}^i_{S_l}$	Estimated loss probability of packet i in stream l
ε	Total erasure rate
ε_b	Burst loss rate
w	Number of past delays recorded

When determining the playout deadlines, we have to consider the tradeoff between delay, losing both MDC descriptions (we refer to this as *packet erasure* in order to distinguish it from the next case), and losing only one description. The latter two cases result in a speech quality distortion. This tradeoff can be stated as follows. Given a certain acceptable speech distortion, minimize the average delay $\mathcal{E}(d^i_{play})$, which is a constrained problem. We convert this constrained formulation into an unconstrained one by introducing a Lagrange cost function for packet i as

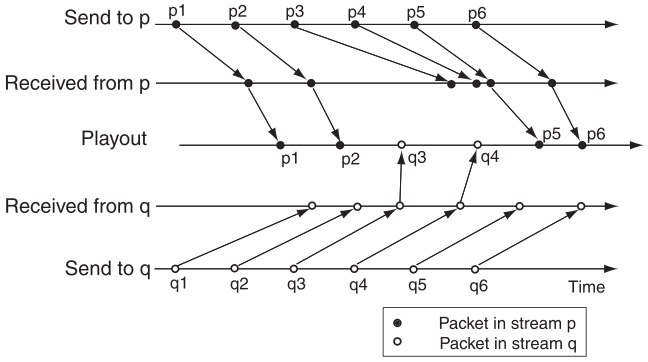

Figure 2: Playout scheduling of multiple streams.

$$\begin{aligned} C^i &= d^i_{play} + \lambda_1 \cdot probability(\text{both descriptions lost}) \\ &\quad + \lambda_2 \cdot probability(\text{only one description lost}) \\ &= d^i_{play} + \lambda_1 \hat{\varepsilon}^i_{S_1} \hat{\varepsilon}^i_{S_2} + \\ &\quad \lambda_2 (\hat{\varepsilon}^i_{S_1}(1-\hat{\varepsilon}^i_{S_2}) + \hat{\varepsilon}^i_{S_2}(1-\hat{\varepsilon}^i_{S_1})), \end{aligned} \quad (1)$$

where $\hat{\varepsilon}^i_{S_1}$ and $\hat{\varepsilon}^i_{S_2}$ are the estimated loss probabilities of the descriptions in stream 1 and 2, respectively, given a certain d^i_{play}. The estimate of $\hat{\varepsilon}^i_{S_1}$ and $\hat{\varepsilon}^i_{S_2}$ is based on past delays recorded for the two streams, which will be discussed in detail in Section 3.2. The higher d^i_{play} is, the lower the loss probabilities since the likelihood of playing out late packets is higher. The Lagrange multipliers λ_1 and λ_2 are predefined parameters used to trade off delay and the two loss probabilities.

The playout deadline is obtained by searching for the optimal d^i_{play} which minimizes the cost function (1). Perceptually, high latency and degraded speech quality resulting from packet loss are "orthogonal" experiences. The multiplier λ_1 is used to trade off total delay and packet erasure probability. Greater λ_1 results in lower erasure rate at the cost of higher delay.

The multiplier λ_2 is introduced to give penalty to speech distortion as a result of playing out only one description. The greater λ_2 is, the better the quality of the reconstructed speech signal at the cost of higher delay. Note that although packet erasure (the second term in (1)) and quality degradation due to the loss of one MDC description (the third term in (1)) are different perceptual experiences, they are not "orthogonal" measures. From (1) it can be deduced that increasing λ_2 also leads to lower erasure probability. However, with zero or very small λ_2 only packet erasure is given concern. In this case good reconstruction quality is not a priority but lower latency is given more emphasis, with the tradeoff between delay and erasure rate determined mainly by λ_1. In practice, this is usually desirable since the human perceptual experience is most strongly impaired by high latency, while speech distortion resulting from losing one description only increases the quantization noise in the MDC scheme we use here and is usually tolerated as a minor impairment.

Fig. 2 illustrates the scheduling process when λ_2 is small and

our emphasis is on low latency. The source stream is coded and sent in two streams p and q. The playout deadline is being kept to a minimum level and dynamically adjusted according to the varying delay jitter of the two paths. At the receiver, the first two packets played are taken from stream p, since they have lower delays. As the delay of stream p increases, the playout switches to stream q and adjusts the scheduling accordingly, so as to avoid any late loss while keeping buffering delay low. The playout switches back to stream p after the arrival of the 5th packet, when the turbulence in path p is over and its delay comes back to normal level.

When switching between streams during speech playout, the playout schedule needs to be dynamically adjusted and adapted to the delay statistics of each individual stream. The dynamic setting of each packet's playout schedule is achieved by an adaptive scheduling technique proposed in our earlier work [13]. In such a scheme, proper reconstruction of continuous output speech is achieved by scaling individual voice packets using a time-scale modification technique which modifies the playout duration of speech segments while preserving the pitch.

3.2 Estimate of loss probability

In (1) the estimates of loss probability $\hat{\varepsilon}^i_{S_1}$ and $\hat{\varepsilon}^i_{S_2}$ are based on recorded past delays of the two streams using order statistics. The network delay of w past packets in each stream l is recorded and is denoted as $d^{i-w}_{S_l}$, $d^{i-w+1}_{S_l}$, ... , $d^{i-1}_{S_l}$. The sorted version of $d^{i-w}_{S_l}$, $d^{i-w+1}_{S_l}$, ... , $d^{i-1}_{S_l}$ is denoted as $D^1_{S_l}$, $D^2_{S_l}$, ... $D^w_{S_l}$, where

$$D^1_{S_l} \leq D^2_{S_l} \leq ... \leq D^w_{S_l}. \quad (2)$$

The r-th order statistic is defined as

$$W_r = F(D^r) = P(d^i_{S_l} \leq D^r),\ r = 1,\ 2,\ ...\ ,\ w,$$

which is the probability that the future delay $d^i_{S_l}$ is no greater than D^r, or the probability that packet i can be received by time D^r. In [7], it is shown that

$$\mathcal{E}(W_r) = \frac{r}{w+1},\ r = 1,\ 2,\ ...\ ,\ w, \quad (3)$$

which is the expected probability that packet i can be received by D^r.

In our application, we extend (2) by additionally defining

$$D^0_{S_l} = D^1_{S_l} - 2 \cdot std(d^{i-w}_{S_l}, d^{i-w+1}_{S_l}, ..., d^{i-1}_{S_l}), \text{ and}$$

$$D^{w+1}_{S_l} = D^w_{S_l} + 2 \cdot std(d^{i-w}_{S_l}, d^{i-w+1}_{S_l}, ..., d^{i-1}_{S_l}),$$

such that we obtain the extended order statistics

$$D^0_{S_l} \leq D^1_{S_l} \leq ... \leq D^w_{S_l} \leq D^{w+1}_{S_l}. \quad (4)$$

The definitions of $D^0_{S_l}$ and $D^{w+1}_{S_l}$ are empirically based on the standard deviation of past delays $std(d^{i-w}_{S_l}, d^{i-w+1}_{S_l}, ..., d^{i-1}_{S_l})$. This solves the problem that the expected playout probability in (3) cannot go beyond $\frac{w}{w+1}$ or go below $\frac{1}{w+1}$. (3) is hence revised as

$$\mathcal{E}(W_r) = \frac{r}{w+1},\ r = 0,\ 1,\ ...\ ,\ w+1. \quad (5)$$

The expected probability corresponding to any d^i_{play} equal to any $D^k_{S_l}$, $k = 0,\ 1,\ ...\ ,\ w+1$, can be determined directly from (5). The expected probability associated with any d^i_{play} in between these discrete values of $D^k_{S_l}$ is found by interpolation.

For stream l,

$$r_l = \max\{k | D^k_{S_l} \leq d^i_{play}\},$$

is the index of the greatest D_{S_l} that is no greater than d^i_{play}. The expected probability that packet i can be received by the deadline d^i_{play} is

$$\mathcal{E}(F(d^i_{play})) = \begin{cases} 0 & \text{for } d^i_{play} \leq D^0_{S_l}; \\ \frac{r_l}{w+1} + \frac{d^i_{play} - D^{r_l}_{S_l}}{(w+1)(D^{r_l+1}_{S_l} - D^{r_l}_{S_l})} & \text{for } D^0_{S_l} < d^i_{play} < D^{w+1}_{S_l}; \\ 1 & \text{for } d^i_{play} \geq D^{w+1}_{S_l}. \end{cases}$$

The expected loss probability of packet i in stream l is then

$$\hat{\varepsilon}^i_{S_l} = 1 - \mathcal{E}(F(d^i_{play})),$$

which is used in the cost function in (1).

The estimation of the loss probability of a future packet using past values is based on the assumption that the past w delays have a similar probability density function (p.d.f.) although the delay distribution varies in the long term. The effectiveness of this estimation depends on how close the accumulated history represents the present delay statistics.

4. INTERNET EXPERIMENTS

Two experiments over the Internet are performed where we transmit MDC streams between hosts in different geographic locations and monitor the end-to-end quality, including delay and packet loss rate. The performance of our proposed multi-stream voice transmission scheme is compared with FEC protected single-stream transmission[1].

4.1 Experimental setup

The first network path we use is the path determined by the default routing algorithm. For the second path, we send the stream to a designated relay server in a different location and let the relay server forward the packets to the destination. For each stream, we send 30ms UDP packets with a payload size of 150 bytes, reflecting 8-bit PCM for finer quantization and 2-bit ADPCM [8] for coarser quantization at 8KHz sampling rate. Packet sequence numbers and delays are collected at the receiving host for both streams. Each experiment runs for 180 seconds. The clocks of the source and destination hosts are synchronized using the Network Time Protocol [15].

In Experiment 1, the source host is located at Netergy Networks in Santa Clara, California, and the destination host is at MIT. The first stream follows the direct path. For the alternative path, we explicitly direct the flow to a designated relay server at Harvard and let the relay server forward the packets to the destination. The route from the

[1]Parts of this experiment have been published in [14] and are included here for completeness.

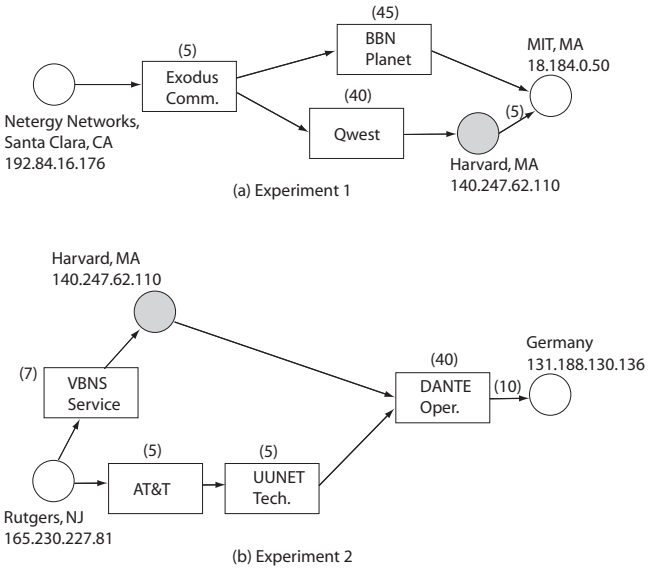

Figure 3: Experimental setup. Source and destination hosts are shown as white circles and relay servers as gray circles, all labeled with their IP addresses. Intermediate service providers are represented by boxes. The numbers in parentheses show the average time in ms required for the packets to traverse corresponding providers or other interconnected networks.

Table 2: Delay and loss statistics of the two streams in the Internet experiments.

Exp.	Path	Delay median (ms)	Delay STD (ms)	Link loss rate (%)	Delay corr. coeff.
1	1	49.6	130.6	0.02	0.028
	2	52.1	19.9	0.85	
2	1	55.0	17.9	0.6	0.034
	2	61.3	10.6	1.1	

source to the relay server and that from the relay server to the destination are determined by routing algorithms without our intervention. The setup of Experiment 1 is shown in Fig. 3(a), with the intermediate service providers shown. It can be observed that, although Harvard and MIT are close neighbors, the streams sent to them from the same source follow very different routes, which provides us with two independent paths. The only link shared by the two paths is operated by Exodus Communications, the service provider of Netergy Networks. The shared link constitutes only a very small part of the total routes and does not contribute to violently varying behavior of the channels. This can be observed from the small normalized correlation coefficient of the delay of the two streams listed in Table 2. Other statistical quantities of the two streams also listed in Table 2 include delay median, link loss rate, and delay standard deviation, which characterizes the delay jitter. It should be noted that Stream 1 in this experiment has very high delay jitter and a few packets experience delays of up to 2000-3000ms, which results in virtual outage periods.

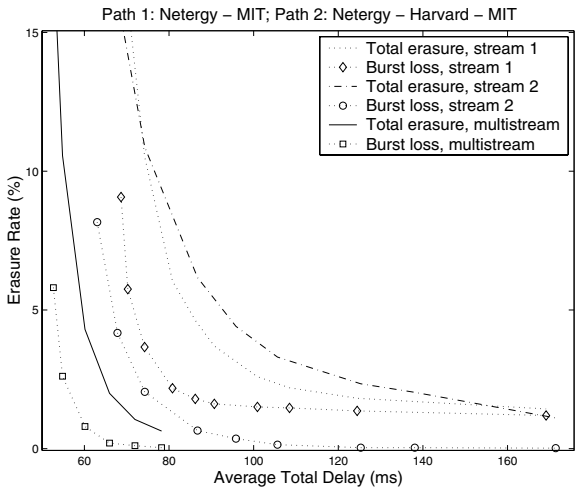

Figure 4: Loss - delay tradeoff, Experiment 1.

Experiment 2 is performed in a similar way. The source host is at Rutgers University, New Jersey, and the destination is at Erlangen University, Germany. We use the same relay server at Harvard for Stream 2. In this experiment a long cross-Atlantic link is shared by the two channels. However, the delay correlation between the two streams is still low, meaning that the channels have largely uncorrelated delay statistics. Although the cross-Atlantic link contributes most to the total end-to-end delay, the delay introduced by this link is nearly constant due to the high bandwidth and good quality of the fiber connection. The 40ms delay shown in Fig. 3(b) is mainly the propagation delay from US east coast across the Atlantic to Europe, which is a constant component. On the contrary, although the delay introduced by domestic service providers is only a very small part of the total delay, it exhibits large variation due to limited bandwidth of and heavy load on these networks. These variations are uncorrelated since the routes are different.

4.2 Results

We compare the schemes of using FEC protected Stream 1 only, using FEC protected Stream 2 only, and using both MDC coded streams. The adaptive playout technique [13], which already achieves state-of-the-art performance for single stream transmission, has been applied to all schemes under comparison.

We first compare the delay - loss tradeoff by setting λ_2 to zero and varying λ_1 in (1) during playout, which suggests that we are less concerned about full reconstruction of both descriptions at the receiver. The results are plotted in Fig. 4. The *average total delay*, $\mathcal{E}(d^i_{play})$, is the average value of d^i_{play} of all the received packets in a playout session. The *total erasure rate* ε is the percentage of lost packets (neither description played out), no matter if the loss is a result of channel erasure or late arrival. We also define the *burst loss rate* ε_b as the percentage of burst erasure occurrences in a session. In calculating ε_b, M consecutively lost packets are counted as $M-1$ occurrences, where $M \geq 2$. Burst loss is one of our greatest concerns because it is difficult to conceal and it impairs voice quality severely. Burst loss is a result

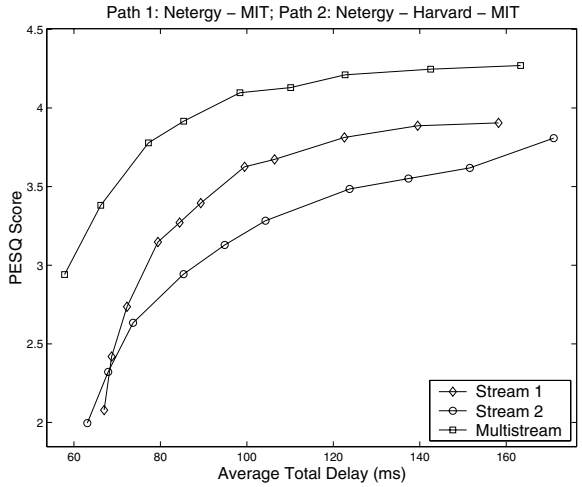

Figure 5: PESQ score vs. delay, Experiment 1.

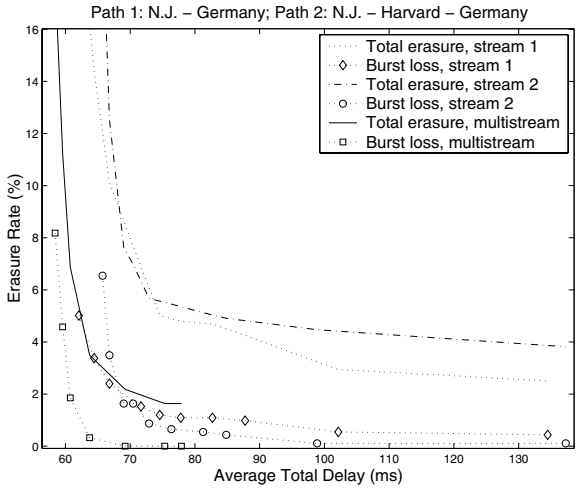

Figure 6: Loss - delay tradeoff, Experiment 2.

of not only consecutive channel erasure, but also delay jitter and sustained outage periods.

From Fig. 4, we observe a significant reduction of the packet erasure rate for a fixed target delay when using our proposed multi-stream path diversity scheme. At the same average delay of 70ms, the total erasure rate is reduced from more than 16% to less than 2%, compared with using FEC protected single-stream transmission. More importantly, the burst loss rate is reduced from more than 3.5% to 0.5%, which is significant for burst loss rate reduction. The majority of packet loss in this experiment is not a result of channel erasure according to Table 2, but late loss caused by the high delay jitter. The loss rates are reduced by a great amount due to the independent jitter characteristics of the multiple paths and independent outage periods.

In Fig. 4, the average delay is also much lower for the multi-stream packet path diversity scheme. At the same erasure rate of 5%, the average delay is reduced by more than 20ms. The delay reduction can be explained by the possibility to play out the description with the lower delay if obtaining full voice quality is not a priority.

One important issue to be addressed in this context is whether the improved tradeoff between delay and loss is achieved at the cost of compromised speech quality, e.g., when we decide to play out the description which arrives earlier while discarding the other one most of the time. In the next experiment, we measure the quality of the reconstructed speech signal for the different schemes. We use PESQ (perceptual evaluation of speech quality), which is an objective measure for narrow-band speech recently adopted by the ITU-T [9]. Unlike previous objective measures, PESQ is applicable not only to speech codecs but also to end-to-end measurements, since it takes into consideration factors such as filtering, variable delay, coding distortions and channel errors. The range of the PESQ score is -0.5 to 4.5, but for most cases the output is a MOS-like score between 1.0 and 4.5 [9]. In our experiments, erased packets are concealed using information from a prior packet [12], while in other situations speech packets are reconstructed

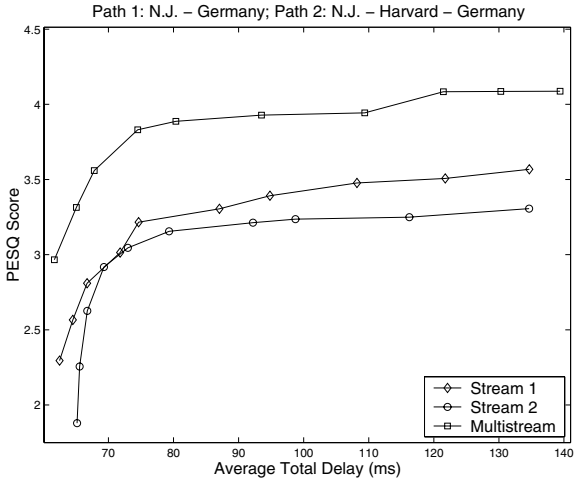

Figure 7: PESQ score vs. delay, Experiment 2.

depending on how many MDC descriptions are received by the playout deadline.

In Fig. 5 we have plotted PESQ score vs. delay as we vary λ_1 and λ_2 while keeping their ratio fixed. It is obvious that as delay increases, speech quality is better due to the lower erasure rate and higher probability that both MDC copies are played out successfully. In Fig. 5, the voice quality corresponding to multiple streams is better than that using single-stream FEC by more than 0.4 PESQ score for all delays. This indicates that the improved tradeoff between delay and loss using MDC path diversity transmission is obtained without compromising voice quality. This can be explained by the fact that the benefit (such as the lower erasure rate) obtained from the path diversity scheme outweighs the loss of one out of the two MDC descriptions, since packet erasure introduces much higher perceptual distortion than quantization noise.

Fig. 6-7 show the corresponding performance results for Experiment 2. Similar improvements for our proposed path

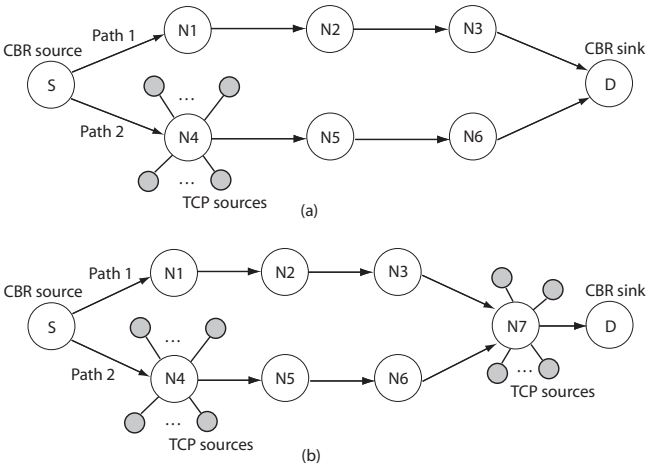

Figure 8: Multi-hop topologies for network simulations: a) independent paths; b) paths sharing a common link. Each of the intermediate nodes N1 through N6 has a number of TCP data sources attached.

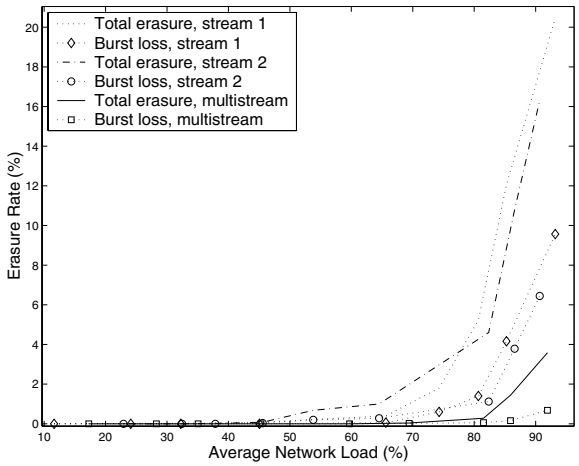

Figure 9: Link loss reduction.

diversity scheme are observed. The delay reduction is less than in Experiment 1, which can be explained by the lower STD of network delay and milder jitter in this experiment (Table 2). However, Fig. 6 shows that the reduction of total erasure rate and burst loss rate is still considerable and Fig. 7 shows an improvement of more than 0.5 PESQ score for all delays.

5. PERFORMANCE ANALYSIS

From the Internet experiments presented in the last section we observe a significant improvement of voice quality and reduction of delay by using packet path diversity. In those examples, however, we have not observed high packet erasure rates or strong correlation between the channels due to the network condition at the time of our experiments. In order to evaluate the effectiveness of multi-stream transmission using path diversity in a more general setting, we study how much gain, in terms of loss and delay reduction, can be obtained from the proposed scheme using a network simulation tool for different network topologies and varying network load.

5.1 Simulation setup

We simulate sending two CBR voice streams from source to destination via two paths, with TCP data traffic contending for network resources at the same time. Two different network topologies are shown in Fig. 8, with the top one showing a setup of independent paths while the bottom one showing a setup with a shared link. Each CBR stream is transmitted in 30ms UDP packets at a rate of 40kbps. The first stream follows the route from node N1 through N3 to the destination, while Stream 2 from N4 through N6 to the destination. The intermediate nodes N1 through N6 represent access points on the routes for data traffic. Each of these nodes has a number of data sources attached, with a large amount of incoming TCP traffic heading for different destinations.

We use the log-normal model proposed in [3] for the file size of each TCP session. Network load is the ratio of average data throughput to the bandwidth of the link. The load is controlled by varying the number of sources attached to each access point, as well as the intermission between successive TCP transmissions.

The simulations are performed using the ns-2 network simulator[2]. In the simulations all links except the shared link in Fig. 8(b) have 10Mbps bandwidth and all links introduce 20ms propagation delay (which will vary in later simulations). The queue buffer size is 100kbyte per port. Each simulation runs for 300 seconds and QoS parameters are monitored from end to end. We will first study loss and delay reduction for independent paths and then study the case with a shared link in Section 5.4.

5.2 Link loss reduction

In Fig. 9, we have plotted the link loss rates of using Path 1 only, using Path 2 only, and using both paths, respectively, as a function of average network load. For each individual path, loss rates resulting from channel erasure increase as the link utilization goes up, which is explained by more packets being dropped by the routers as the network becomes congested and queues fill up.

With multi-stream multi-path transmission, the loss rates shown in Fig. 9 are much lower due to the uncorrelated characteristics of the erasure channels. Even at a high load of 90% on both paths, the total erasure rate of the two-path scheme is still below 4%, while a single-stream could result in 16%. Using path diversity, the hostile burst loss rate can be kept below 1% at high load, which could otherwise go up to 6% if using only one stream.

5.3 Delay reduction

Multiple stream transmission reduces the delay by providing the opportunity of receiving and playing out the packets with lower delay. Here we study how much delay reduction can be obtained compared with using only one stream. To this end, we define a quantity, *average delay reduction*,

[2]http://www.isi.edu/nsnam/ns/

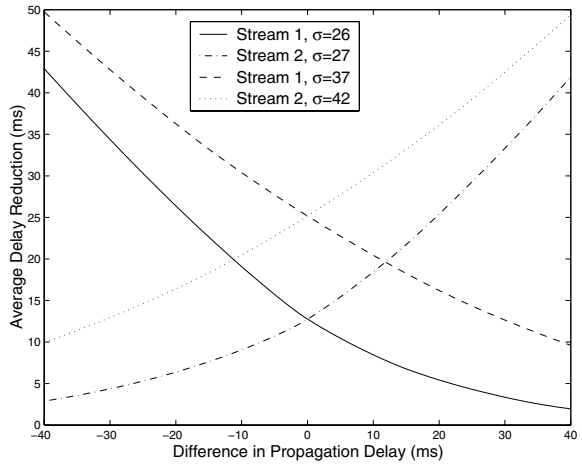

Figure 10: Average delay reduction vs. difference in propagation delay.

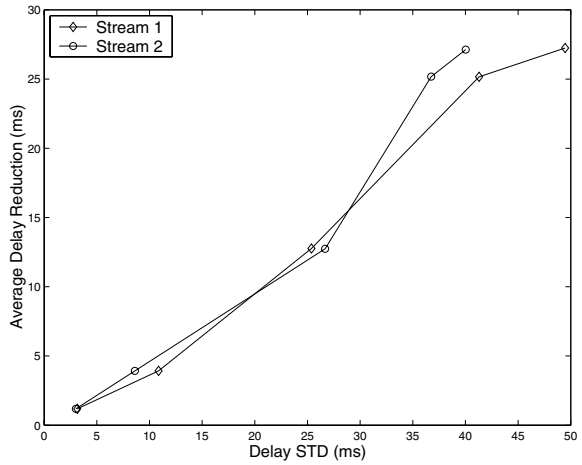

Figure 11: Average delay reduction vs. delay STD.

denoted by Δd_{S_l}, for stream l. Average delay reduction is calculated as the trace average of the difference between $d^i_{S_l}$ and d^i_{play}, by only counting packets with $d^i_{S_l} > d^i_{play}$. It characterizes the average delay reduction by using multiple streams instead of stream l only.

In this experiment we vary the propagation delay of links in the simulation and study its effect on average delay reduction. In Fig. 10, we have plotted Δd_{S_l} versus the difference in propagation delay of the two paths, which is defined as the difference between the trace average of the propagation delay of Path 2 and Path 1. Two sets of curves are shown in Fig. 10 for a low delay STD of about 26ms and a high delay STD near 40ms, respectively. For either set, it is observed that the two streams obtain equivalent gain in terms of delay reduction as they have the same propagation delay. As the difference in propagation delay increases, Stream 2 gains more while Stream 1 gains less. This indicates that if the alternative path has much higher average propagation delay than the direct path, the benefit from using path diversity becomes small since the alternative path becomes trivial. However, in practice the default path does not necessarily have lower delay than the alternative path according to [17]. Hence, the efficiency of multi-path transmission depends on the availability of an alternative path which has a mean delay not much higher than the default path. This is the case in our experiments in Section 4 (refer to Table 2). In both experiments in Section 4, we observe median delays close enough since both paths follow close geographical routes.

In Fig. 10, higher gains are observed with path diversity when the delay variation is higher. This can be seen more clearly in Fig. 11. The change of delay variation is obtained by varying network load in our simulations. For both streams, the delay reduction increases as the STD of network delay goes up. This can be explained by the effectiveness of the packet path diversity scheme in reducing the virtual variability of the channels.

5.4 Shared link

So far we have studied independent paths. Shared links in the multi-path setup potentially increase the correlation between the two paths. In practice this might be common when the bottleneck is on the "last mile", such as the DSL or T1 line connected to small offices or small homes (SOHOs). In these cases, packet path diversity can only be employed before the "last mile" and statistical correlation is expected on the shared link.

We now study the case in which the multiple paths share a common link with different bandwidth of 384kbps, 1Mbps, and 10Mbps, respectively. The simulation condition is the same as in the case of independent paths, except that we have a number of data sources attached to node N7 (Fig. 8(b)) that contend for the resources of the shared link with the voice streams.

The correlation coefficients of the delay of the two streams are found to be 0.92, 0.88, 0.19 at 66% load for shared bandwidth of 384kbps, 1Mbps, and 10Mbps respectively. Compared with the delay correlation coefficient of 0.002 of the independent path case (Fig. 8(a)) at the same load, the correlation between the two paths is much stronger with the existence of a loaded shared link. This is because congestion takes place on the shared link and the delay jitter caused by the queuing delay before the shared link has a similar pattern. Congestion is more severe when a shared link has lower bandwidth, which results in even stronger delay correlation.

The link loss rates observed when using path 1 only and when using both paths are plotted in Fig. 12 for different bandwidth as a function of the average network load. The average network load shows the average utilization of all the 10Mbps links, as well as of the shared link. To make the figure easier to read, we omit the loss rate of Stream 2 here, which is similar to that of Stream 1 due to the symmetric simulation topology. It is easy to understand that the increased packet loss rates correspond to the lower bandwidth of the shared link. Although the delay between the two streams exhibits strong correlation, it is observed from Fig. 12 that the packet erasure rate can still be significantly reduced by using the path diversity scheme, especially at high load. With a low shared bandwidth of 384kbps, the loss rate when using one stream is 27% at 85% network load. When using two streams at the same conditions the

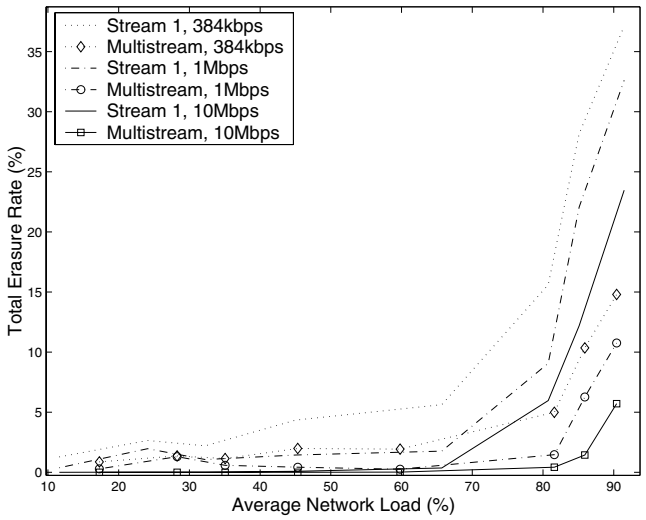

Figure 12: Link loss rate reduction with a shared link of different bandwidth.

total erasure rate is reduced to 11% [3]. Similar reduction in loss rates can be observed for links of higher bandwidth.

6. CONCLUSIONS

In this work, we propose a scheme for real-time voice transmission using packet path diversity. Experiments over the Internet show that with this scheme, the mean end-to-end latency and loss rate can be greatly reduced, and the overall PESQ score can be improved by more than 0.4, when compared with a scheme that uses FEC protected single path transmission at the same payload data rate.

Independent multiple paths provide channels with uncorrelated network behavior such as loss, delay variation, and outage periods. By taking advantage of path diversity, packets can be used and played out from a backup channel when the default path suffers from negative disturbance. For this reason, the end-to-end delay can be reduced by taking the packet with lower delay, and late loss and burst loss can be avoided. A Lagrangian cost function in combination with adaptive playout scheduling is introduced that trades off delay, late loss, and speech reconstruction quality for multi-path transmission where the Lagrange multipliers can be used to control the relative importance of each of these quantities.

The efficiency of packet path diversity depends on the statistical correlation of the paths. It also depends on the availability of an alternative path which has a mean delay not much higher than the default path. This can be obtained in practice by sending streams over networks with close geographical routes but serviced by different ISPs. Simulation results also show that the obtainable path diversity gain depends on the difference in propagation delay of the multiple paths and the variance of the network delay. If a slower link has to be shared by the multiple paths, it is found that the delay correlation between the streams is very strong. Despite of this, the link drop rate can still be reduced significantly by using packet path diversity.

With current Internet protocols, multi-path transmission can be realized by a dedicated overlay network of relay servers or by exploiting future peer-to-peer architectures. The source routing feature of the next-generation IP protocol IPv6 also promises a practical way of implementing packet path diversity.

[3]At this packet loss rate the reconstructed voice quality can still be very good when using time-scale loss concealment techniques [12].

7. ACKNOWLEDGMENTS

This work has been supported by a gift from Netergy Networks, Inc., Santa Clara, CA. The authors would like to thank Nikolaus Färber and Rong Pan for their help on network simulations and John Apostolopoulos for helpful discussions. The authors would also like to thank Mack Hashemi, Hanming Rao, Bernd Westrich, Xiaowei Yang and Bin Zhang for providing us the remote machines to perform the experiments.

8. REFERENCES

[1] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. The case for resilient overlay networks. In *Proceedings of the 8th Annual Workshop on Hot Topics in Operating Systems (HotOS-VIII)*, May 2001. Online at: http://nms.lcs.mit.edu/projects/ron/.

[2] J. G. Apostolopoulos. Reliable video communication over lossy packet networks using multiple state encoding and path diversity. In *Proceedings Visual Communication and Image Processing*, pages 392–409, Jan. 2001.

[3] M. Arlitt and T. Jin. Workload characterization study of the 1998 World Cup web site. *IEEE Network*, 14(3):30–7, May 2000.

[4] J.-C. Bolot. End-to-end packet delay and loss behavior in the Internet. *Computer Communication Review, ACM SIGCOMM '93*, 23(4):289–298, Sept. 1993.

[5] J.-C. Bolot, S. Fosse-Parisis, and D. Towsley. Adaptive FEC-based error control for Internet telephony. In *Proceedings of IEEE INFOCOM '99*, volume 3, pages 1453–1460, Mar. 1999.

[6] V. Hardman, A. Sasse, and A. Watson. Reliable audio for use over the Internet. In *Proceedings of INET '95*, pages 171–178, June 1995.

[7] R. V. Hogg and E. A. Tanis. *Probability and statistical inference*. Macmillan Pulishing Company, 4th edition, 1993.

[8] ITU-T Recommendation G.726. *40, 32, 24, 16 kbit/s Adaptive differential pulse code modulation (ADPCM)*, Dec. 1990.

[9] ITU-T Recommendation P.862. *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, Feb. 2001.

[10] W. Jiang and A. Ortega. Multiple description speech coding for robust communication over lossy packet networks. In *International Conference on Multimedia and Expo*, volume 1, pages 444–7, Aug. 2000. New York, NY, USA.

[11] G. Kubin and W. Kleijn. Multiple-description coding (MDC) of speech with an invertible auditory model. In *1999 IEEE Workshop on Speech Coding Proceedings*, pages 81–3, June 1999. Porvoo, Finland.

[12] Y. J. Liang, N. Färber, and B. Girod. Adaptive playout scheduling and loss concealment for voice communication over IP networks. Submitted to *IEEE Transactions on Multimedia*, Feb. 2001. Online at: http://www-ise.stanford.edu/~yiliang/publications/

[13] Y. J. Liang, N. Färber, and B. Girod. Adaptive playout scheduling using time-scale modification in packet voice communications. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings ICASSP01*, May 2001. Salt Lake City, UT.

[14] Y. J. Liang, E. G. Steinbach, and B. Girod. Multi-stream voice over IP using packet path diversity. In *Proceedings of IEEE 4th Workshop on Multimedia Signal Processing*, Oct. 2001.

[15] D. Mills. *Internet Time Synchronization: the Network Time Protocol*. RFC-1129, Internet Engineering Task Force, Oct. 1989.

[16] H. Sanneck and N. T. L. Le. Speech property-based FEC for Internet telephony applications. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3969, pages 38–51, Jan. 2000.

[17] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson. The end-to-end effects of Internet path selection. *Computer Communication Review, ACM SIGCOMM '99*, 29(4):289–99, Oct. 1999.

[18] R. Singh and A. Ortega. Erasure recovery in predictive coding environments using multiple description coding. In *1999 IEEE Third Workshop on Multimedia Signal Processing*, pages 333–8, Sept. 1999. Copenhagen, Denmark.

[19] B. Wah and D. Lin. Transformation-based reconstruction for real-time voice transmissions over the Internet. *IEEE Transactions on Multimedia*, 1(4):342–351, Dec. 1999.