Early Estimation of Voice over IP Quality

Pravesh Biyani, McMaster University, Canada
Olof Hagsand, KTH, Sweden
Gunnar Karlsson, KTH, Sweden
Ian Marsh, SICS, Sweden
Ignacio Más, KTH, Sweden

Abstract

Users of Voice over IP (VoIP) applications are sensitive to the quality of an ongoing call. We hypothesize that the quality of a VoIP session can be estimated from the first few seconds of the session and this can be generalized to other VoIP calls. Our approach is an in-band probing mechanism and does not require any external monitoring schemes or network support. We show by post processing VoIP data from globally distributed sites that it is possible to determine the quality after an initial number of seconds. One application is admission control, where it would be possible to reject poor quality calls before they are fully established.

1 Introduction

The scope of this work is to investigate the quality of VoIP sessions over best effort IP networks. Our contribution is a method for estimating the quality of a VoIP session from the early seconds of the session. We provide intuitive reasoning, supported by empirical measurement, that the loss can be estimated early.

We state that measurements between a set of paths can produce data useful for predicting future quality. It is important to state we only use packet loss as a quality indicator of a VoIP session in this work and we are only advocating that the quality can be estimated early for the same Internet paths we have measured. Packets arrive at a receiver 50 times per second (assuming no loss) in our VoIP scheme [1], so we have frequent sampling and observation of the network state. We measure the loss by counting gaps in the sequence numbers and dividing the count with the total number of expected packets. We refer to this as the *loss ratio* of the call.

Given the loss ratio for one call we use our empirical measurements to show that the loss ratio can be estimated for many connections across many paths. The measured loss after an initial number of seconds (0 to 10) is compared to the loss measured over the whole session. We use the correlation between the two measurements to determine how accurate the estimation is. This is possible as we have the whole session recorded at the receiver stored in data files available for post processing [2](see also http://www.sics.se/˜ianm/COST263/cost263.html).

The structure of this paper is as follows, the next section gives some background on how the empirical measurements were taken. We also give a short explanation of why we reduced the sample set of the data and describe how we measure the loss ratio for one call. Section 3 shows the results for all considered calls showing the accuracy of the loss estimation for different initial time periods. Section 4 gives some conclusions of our work, some applications and pointers to future work.

2 Measurement description

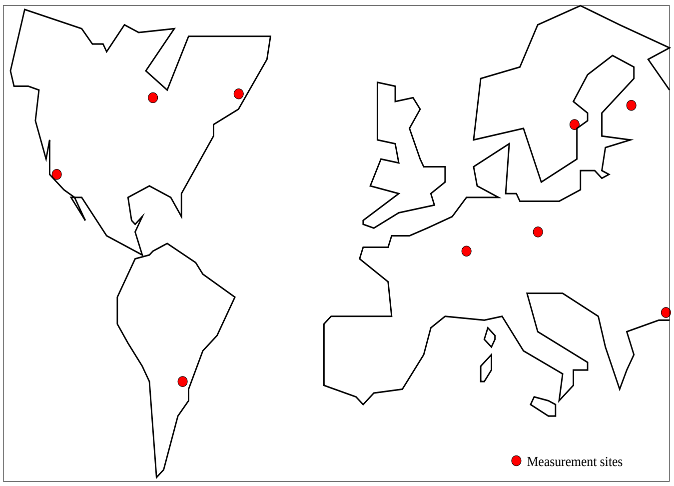

Figure 1: Measurements were made between nine academic sites worldwide.

This paper uses the results of previous work where over 18000 VoIP calls were measured between hosts at nine academic sites [2]. The location of the sites is shown in Figure 1. The sites were connected as a full-mesh allowing us, in theory, to measure the quality of 72 different Internet network paths. These paths exhibit large differences in timezones, hop counts and geographic distances.

The measurements were performed in the following way: on an hourly basis, a call was established between a sending host and a receiving host. The measurements were performed over 15 weeks. A sequence of pre-recorded speech samples was transmitted at 64 kbps as a stream of RTP/UDP/IP datagrams. The receiver made a detailed log of the arrival process, recording the reception time of each datagram.

Each call was initiated by scripts that started one sender and one receiver at the two hosts. The sender packetized the pre-recorded audio samples and transmitted them to the receiver.

2.1 Reducing the sample set

For the purposes of this paper, we needed a common basis for our analysis, and therefore we selected a subset of the 18000 calls based on the following restrictions:

1. 20 ms packetization. The measurements contain sessions with different packet sizes, so we excluded the sessions that were not packetized every 20 ms. We cannot state that sessions with larger packet sizes would induce the same loss as for 20 ms packetization, hence they were discounted.

2. No silence periods. Silence suppression is used to reduce the load on the network by not transmitting packets when the speaker is silent. As was the case for different packet sizes, we cannot assume that sessions with and without silence suppression induce the same loss. Therefore sessions with silence suppression were also discounted. Additionally, if the speech data contains silence periods, it is difficult to make an analysis of continuous time intervals, since parts of the calls contain no data.

3. Lossy calls. Loss free calls do not provide any information for our analysis: both the probing and the total loss rate are zero which gives perfect correlation. A large percentage of the calls are in fact loss free. This is important since it reduces our sample set considerably. We attribute the large number of loss free sessions to the fact that the sites are located on well provisioned (academic) networks.

This set of restrictions resulted in a subset of only 564 calls. Even though this reduction is considerable, all nine sites are represented in the subset, since two hosts had poor connectivity or infrastructure. These two hosts were located in Turkey and Argentina.

2.2 Measuring a single call

The sender can be seen as a synchronous process that introduces no delay. The network introduces variable latency and packet loss, which is detected at the receiver. Only the receiver is aware of the quality of the transmission.

Figure 2 shows the loss process of an example call as observed by a receiver. The call was made between the Argentinian and Turkish sites. The figure shows a loss pattern that is representative of many other calls in the subset. The left plot shows the number of lost packets against time. In the plot, it can be seen that the number of lost packets increases almost linearly as the call proceeds. In fact, we have seen this linear behavior in most of the calls.

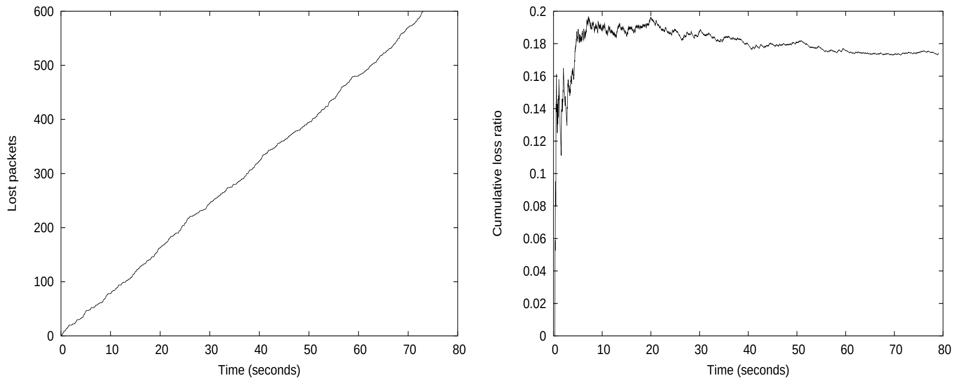

Figure 2: Loss behavior of a single example call between Turkey and Argentina. The left plot shows the number of lost packets against time. The right plot shows the cumulated loss as the call proceeds.

The right plot shows the cumulative loss ratio for the same call. The loss ratio is defined as the number of lost packets divided by the number of sent packets over the interval from the start to the specified time. We show the cumulative plot to clarify how long we need to measure to obtain a good estimation of the final loss ratio. From the plot, we see that the final loss ratio for the complete call is approximately $17\% - 18\%$.

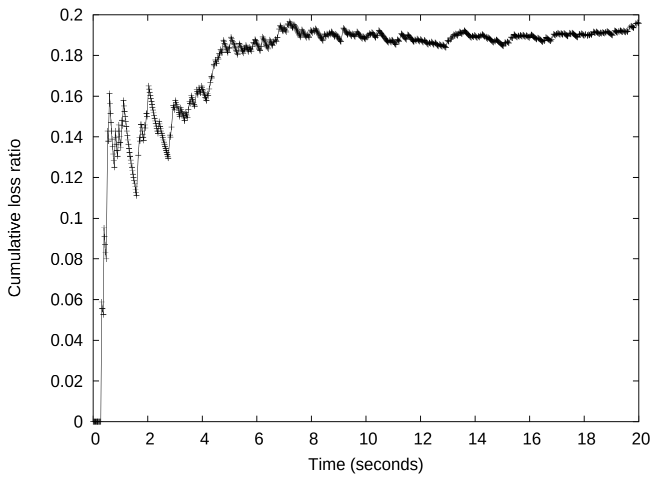

Figure 3: Cumulative loss of the same session showing only the initial portion of the call.

In Figure 3 we show the first 20 seconds of Figure 2. From the figure, we see that the initial loss is approximately 14% after 1 second and 19% after 10 seconds. These are early estimations of the final loss rate. We want to know how accurate

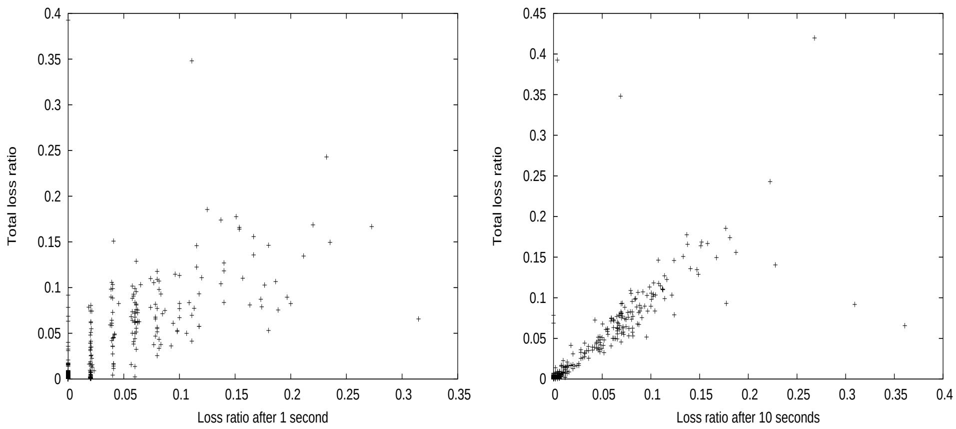

Figure 4: Two plots showing the relation between initial and total loss ratios for all calls. The left plot shows the relation with an initial interval of one second, the right plot after ten seconds.

such early estimations are. Therefore we need to study the relation between the loss ratio of an initial interval and the complete part of the call.

3 Results

In the preceding section, we considered only one call. In Figure 4, the loss ratio for the total call is plotted against the loss ratio of an initial interval, for all calls in our selected subset. In the figure, every point represents a call. The plots show that as the initial interval increases, the points group closer around the line $y = x$. In other words, the correlation increases and the estimation becomes better.

The plots in Figure 4 give an intuitive measure of the correlation between the loss ratio of the initial interval and the total call. In order to evaluate more precisely the accuracy of the estimation, we computed the actual correlation factor as a function of the initial interval. The result is plotted in Figure 5.

The figure shows that the correlation factor increases as the probing interval increases. From the figure, we can clearly see that the correlation stabilizes after four seconds. This is important, because after this point further measurements are not necessary.

4 Conclusions and future work

In this paper we have proposed a mechanism to obtain an early estimation of the quality of voice over IP calls. Based on empirical measurements, we have shown that an initial estimation of packet loss can be made for VoIP sessions. From our

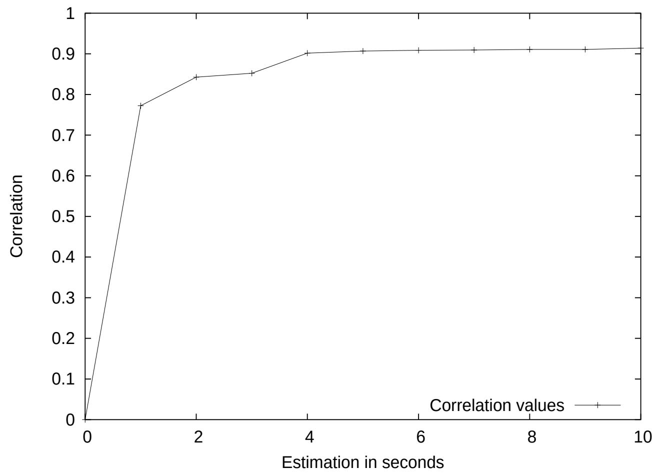

Figure 5: The correlation factor as a function of the initial probing interval

particular data, we have shown that it is sufficient to make an estimation after four seconds.

We believe that the primary application for initial estimation of VoIP quality is admission control, which can save valuable resources on congested links.

However, we believe that the number of calls experiencing loss were somewhat limited which reduced the number of useful calls for our analysis. We would therefore like to extend the number of calls experiencing lossy connections.

References

[1] Olof Hagsand, Ian Marsh, and Olof Hagsand. Sicsophone: A low-delay internet telephony tool. Technical Report T2002:26, SICS – Swedish Institute of Computer Science, December 2002.

[2] Ian Marsh and Fengyi Li. A VoIP measurement infra-structure. In *16th Nordic Teletraffic Seminar*, Helsinki, Finland, August 2002.