Real-Time Non-Intrusive VoIP Evaluation Using Second Generation Network Processor

Dorel Picovici, Adil Raja and Colin Flanagan

Department of Electronic and Computer Engineering, University of Limerick, Ireland.

ABSTRACT

In today's telecommunications market there is a clear trend to adopt a flexible infrastructure to minimize costs and facilitate the introduction of new services. Typical examples are represented by dynamically managed packet-based Next generation IP networks (NGN) and 4G mobile networks. In general the rate of success of these new networks will depend on whether they can deliver at least the same quality of service (QoS) than the networks they replace at lower overall cost. Voice quality represents a major dimension of perceived QoS. Measures that predict voice quality are essential in monitoring and managing the performance of such networks. Traditionally the perceived quality of voice has been evaluated by expensive and time-consuming subjective listening tests. Several attempts have been made to supplement subjective tests with objective estimators of voice quality. From the viewpoint of the measurement procedure, objective voice quality estimation methodologies can be categorized either as speech-layer objective models or packet-layer objective models.

1. INTRODUCTION

This paper introduces packet-layer objective models for voice quality assessment and proposes a new real-time packet-layer evaluation technique for conversational voice over IP (VoIP) quality. The proposed model is based on using IXP2400 network processor (NP).

Speech-layer objective models require speech signals as inputs and produce estimates of subjective scores such as Mean Opinion Score (MOS) [1]. Examples of such models include the Bark Spectral Distortion (BSD), the Perceptual Speech Quality (PSQM), the Modified BSD (MBSD), the Measuring Normalizing Blocks (MNB), the PSQM+, the Telecommunication Objective Speech Quality Assessment (TOSQA), the Perceptual Analysis Measurement System (PAMS), and the Perceptual Evaluation of Speech Quality (PESQ).

This work has been funded by SFI under the NCNRC project 03/CE.2/I315

Packet-layer objective models are different than speech-layer objective models in the sense that they exploit IP packet characteristics rather than speech signals. ITU-T is currently in the process of standardizing a packet-layer objective speech quality measure for use in real-time quality monitoring. This process is now in the algorithm selection phase and there are two candidates: PsyVoIP [4] and VQmon [5].

Typical communications involve real-time two-way conversations. In VoIP and mobile networks the voice quality is affected by a wide variety of network impairments and can vary from conversation to conversation and even during the conversation. For example due to the increasing delay the mobile networks can reduce the QoS, which will increase the possibility of double-talk and increase the user's perceptibility of echo. There is an increased need for a real-time voice quality evaluation method. Ideally, this method should combine "conversational" impairments such as noise level, echo and delay, together with "listening quality" measures to estimate the overall (conversational) quality perceived by the user at either end of the connection. A new approach for developing a real-time, non-intrusive and objective assessment of conversational voice quality is suggested. The underlying principle of the proposed method is based on using Intel's second generation NP (IXP2400) to "bridge" IP traffic as well as to extract relevant information from the packet headers. The extracted information is used in assessing the conversational voice quality.

Following this introduction the paper is organized as follows: Section-2 describes the hardware and software description of the test-bed. Section-3 describes the choice of transport protocol for the proposed work. Section-4 describes the implementation details of various algorithms for the computation of packet-based quality parameters on IXP2400 Network Processor. Section-5 discusses the tests and results. Finally, Section-6 presents conclusions and future work directions.

2. TEST BED DESCRIPTION

The approach described here uses a network planning voice quality evaluation model known as the ITU-T Rec. G.107 [6]. The target is to implement this model on the Intel IXP2400 NP [7]. The IXP2400 NP is a 32bit X-Scale RISC processor compatible with the ARM Version 5. It has eight Microengines which are specifically targeted at optimizing traditional network operations such as packet forwarding. The X-Scale core is specifically targeted at optimizing packet-layer objective models.

The software architecture of the system (shown in Figure 1) includes two main planes: a) a data plane, which includes all the microblocks used to receive, process and transmit packets, and b) control plane, which includes all the X-Scale core components that are used for processing the data extracted by packet processing microblocks. The data plane processing starts with forwarding IP packets as well as measuring intermediate quality parameters such as packet loss, jitter and delay. The MOS score can be calculated by using the intermediate quality parameters with the ITU-T Rec. G.107 model.

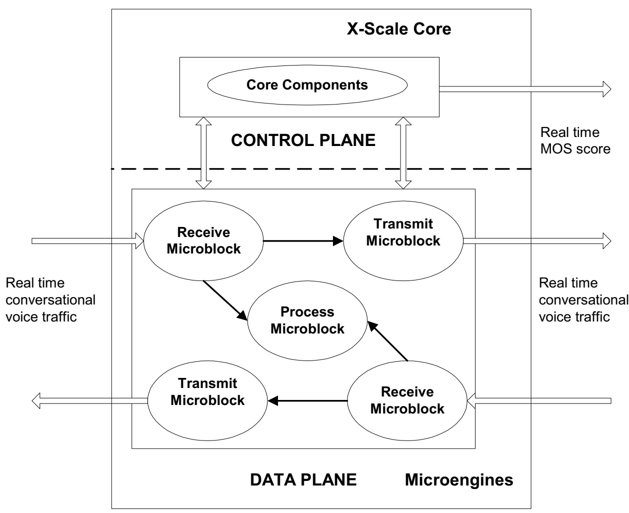

Figure 1: Block diagram of a new real-time VoIP evaluation method.

3. REAL-TIME TRANSPORT PROTOCOL

Real-time Transport Protocol (RTP) [8] has been chosen as the underlying transport protocol in this approach. RTP is the most widely used and deployed network transport protocol for VoIP applications. Initially designed by Internet Engineering Task Force (IETF), RTP is now an integral part of the H.323 suite of protocols for packet-based multimedia communications systems proposed by ITU-T [9].

A RTP packet is composed of a header and payload. The payload carries the voice frames while the header contains packet information. Various header fields in an RTP packet can be used to extract information about a particular voice stream.

Intermediate quality parameters such as packet loss and jitter can be calculated by using the "sequence number" and "timestamp" fields of the RTP packet header respectively. The computation procedures are described in detail in [8]. These procedures have been implemented into the packet processing microblock (see Figure 1). Similarly Round-Trip delay for a voice stream over RTP can be computed using the Real-time Transport Control Protocol (RTCP) packets [8]. RTCP packets supplement RTP by carrying control information of a VoIP call. This information can be shared among VoIP end-points. Fields used for the computation of Round-Trip Delay are "NTP Timestamp", "Last Sender Report (LSR)" and "Delay Since Last Sender Report (DLSR)".

4. IMPLEMENTATION DETAILS

This section discusses the implementation details of computation of packet loss, jitter and round-trip delay on the packet processing microblock.

4.1. Packet Loss

The packet processing microblock makes use of the sequence number field (16 bits) of the RTP header to calculate the packet loss. Initially, the sender increments by one the sequence number for each successive packet sent. The receiver uses this sequence number to detect packet loss.

Packet processing microblock calculates the packet loss in accordance with Appendix A.3 of [8]. Following this method the packet loss can be calculated between the sender and the IXP2400 NP.

For a given session an RTP transmitter appends sequence numbers to the packets before sending over to the destination(s). The sequence number for the first packet is first chosen at random and then incremented by one for the remaining successive packets. Adding sequence numbers to packets has the advantage that packet sequence numbers can be used to calculate the packet loss. The procedure to calculate the packet loss is as follows.

For a given stream of VoIP packets, the packet processing microblock notes the sequence number of the first packet it observes as the "base_seq" (base sequence number). For any subsequent packets the packet processing microblock stores the sequence numbers as the "max_seq_num" (maximum sequence number).

The packet processing microblock stores a variable named "received" which has an initial value of zero. The "received" variable is incremented by 1 every time a packet is received. In the light of above information the packet loss is calculated in the following way.

$$\text{Packet loss} = \text{num\_packets\_expected} - \text{received} \quad (1)$$

Where "num_packets_expected" is another variable which represents total number of packets expected and is given by:

$$\text{num\_packets\_expected} = \text{max\_seq\_num} - \text{base\_seq} + 1 \quad (2)$$

4.2. Packet Jitter

Packet jitter is defined as the variation in time between packets arriving. Jitter for the VoIP packets is calculated in accordance with Appendix A.8 of [8]. The following formula has been used in [8] for the computation of jitter.

$$J(i) = J(i-1) + (|D(i-1, i)| - J(i-1))/16 \qquad (3)$$

Where: J(i) is the value of jitter for current packet and J(i-1) is the value of jitter for the previous packet. |D(i-1, i)| is the absolute value of the difference between transit times of two consecutive packets in the network. |D(i-1, i) | is the absolute value of the difference between transit times of two consecutive packets in the network and it can be written as.

$$D_{(i,j)} = (R_j - R_i) - (S_j - S_i) = (R_j - R_i) - (R_i - S_i) \qquad (4)$$

Where: R_j, R_i, S_j and S_i are reception and sending times of j^{th} and i^{th} packets respectively.

For periodically generated voice traffic like PCM, the RTP time stamp contains a nominal sampling instant as determined from the sampling clock and not a reading from the system clock [8]. As an example, for fixed rate voice traffic the time stamp for each successive packet is increased by the number of samples in the previous packet. This information can be used to derive the relative sampling in units of time. In this approach the voice traffic was encoded using PCM µ-law codec. PCM has a sampling rate of 8kHz. This means, for instance, that given a data payload of length 80 samples (where 1 sample is represented by 1 byte) per packet, the payload contains 10 ms of voice. Given this, the timestamp for subsequent packets are increased by 10 ms. PCM µ-Law encoded voice was used for this study which has a sampling rate of 8 kHz.

The packet processing microblock calculates the jitter by implementing equations (2), (3) and the values received in the timestamp field to compute the value of jitter.

4.2. Round-trip Delay

Round-trip-delay is defined by [8] by using the RTCP-SR (Real Time Transport Control Protocol – Sender Report) and/or RTCP-RR (Real Time Transport Control Protocol – Receiver Report). A number of VoIP applications were tested for computation of round-trip delay. Amongst them are Speak Freely [10], OpenPhone [11], and WinRTP [12]. Personal investigation showed that none of them has a complete implementation of RTCP therefore computation of round-trip delay is not possible. An alternative to this issue was addressed by the authors by using a UDP (User Datagram Protocol) based client/server application written in Java. The client sends a UDP datagram containing an RTCP-SR to the server. The server receives the RTCP-SR and sends its own RTCP-SR back to the client. While sending a RTCP-SR the client or the server encodes the idle time spent by the machine between sending the RTCP-SR and receiving RTCP-SR from the remote side. This idle time is encoded into the DLSR field of the RTCP-SR being sent. Along with this the client or the server also encodes the LSR into the RTCP-SR being sent. This LSR is obtained from the middle 32 bits of the “NTP timestamp” field of the previously received RTCP-SR.

Procedure for measuring the round-trip delay between IXP2400 NP and VoIP endpoint-A is shown in Figure 2. Endpoint-B sends a RTCP-SR packet to endpoint-A. As the packet is received by IXP2400 NP the processing microblock notes the time t_1 and middle 32 bits of the “NTP-timestamp” field. After reception of the RTCP-SR end-point-A sends a RTCP-SR to endpoint-B. The middle 32 bits of the “NTP timestamp” from the previously received RTCP-SR are encoded in this RTCP-SR’s LSR field. The idle time between receiving the previous RTCP-SR and sending this LSR is also encoded in the DLSR field. Upon receiving this RTCP-SR the packet processing microblock compares its LSR with the previously noted middle 32 bits of the NP timestamp. If these values match it computes the time t_2 of arrival of this RTCP-SR and also notes the DLSR value. The round-trip delay ($D_{RTCP\text{-}A}$) between endpoint-A and the IXP2400 NP can be calculated by the following equation.

$$D_{RTCP\text{-}A} = t_2 - t_1 - DLSR \qquad (5)$$

In a similar manner the $D_{RTCP\text{-}B}$ between endpoint-B and IXP2400 NPU can be computed. The sum of these two values is equal to the round-trip delay (D_{RTCP}) between endpoint-A and endpoint-B

$$D_{RTCP} = D_{RTCP\text{-}A} + D_{RTCP\text{-}B} \qquad (6)$$

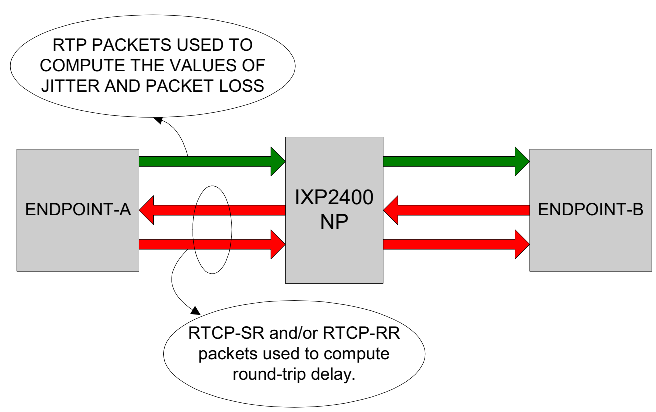

Figure 2: The figure schematically shows IXP2400 NP bridge and the flow of RTP and RTCP packets between two VoIP end-points.

5. TESTS AND RESULTS

Tests were conducted using OpenPhone VoIP software [11] installed on two Pentium-4 machines with Microsoft's Windows XP operating system. Each machine had 512 mega-bytes of SRAM memory. Netperf benchmark was used as a traffic generating tool to vary the traffic load on the network while testing. A control component is used to display the accumulated results.

In all of the tests zero packet loss was detected. The results for packet loss were validated by comparing them with the ingress and egress counters of IXP2400 NP.

The tests for computing packet jitter were run for 60 seconds each and the samples for packet jitter were taken once every 100 milliseconds. Figure 3 shows the values of packet jitter plotted against time. Packet jitter varies drastically with time. The overall value of jitter is below 5 milliseconds.

The cumulative average round-trip delay between two VoIP endpoints was between 0.437 to 0.689 milliseconds depending on the network traffic.

The accuracy of implementation and execution of procedures for computing inter-arrival jitter and round-trip delay has been verified using the simulator of Intel Exchange Architecture (IXA) Standard Development Kit (SDK).

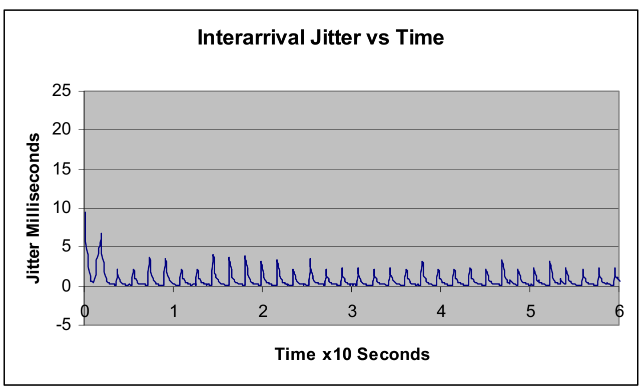

Figure 3: Plot of jitter against time.

6. CONCLUSIONS AND FUTURE WORK

A packet processing application for VoIP has been developed and tested to run on an Intel second generation network processor (IXP2400 NP). The packet processing microblock currently computes the values for packet loss, jitter and round-trip delay for an ongoing VoIP call. As a next step these values shall be used as an input to ITU-T Rec. G.107 model in order to calculate the MOS score for VoIP calls. The performance can be evaluated by comparing the correlation between the estimated MOS scores obtained by the proposed method and the predicted MOS scores obtained by the ITU-T recommended speech-based methods such as ITU-T Recommendation P.861. Testing the proposed system under high traffic loads of the order of a few gigabits per second is also a future consideration.

The authors also propose to implement a voice signal quality assessment module on IXP2400 NP in future. The objective of the study would be to perform perceptual evaluation of speech quality in a non-intrusive manner on a second generation Network Processor.

Acknowledgment

The authors would like to thank Wind River System Inc. for providing the software used in this work.

7. REFERENCES

[1] ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality", International Telecommunication Union, Geneva, Switzerland, August 1996.

[2] ITU-T Recommendation P.861, "Objective Quality Measurement of Telephone Band (300-3400 Hz) Speech Codecs", International Telecommunication Union, Geneva, Switzerland, August 1996.

[3] ITU-T Recommendation P.800.1, "Mean Opinion Score (MOS) Terminology", International Telecommunication Union, Geneva, Switzerland, 2003.

[4] S. Broom and M. Hollier, "Speech Quality Measurement Tools for Dynamic Network Management", MESAQIN 2003.

[5] A. Clark, "Modelling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality", IP Telephony Wksp. 2001.

[6] ITU-T Recommendation G.107., "The E-model, a computational model for use in transmission planning", International Telecommunication Union, Geneva, Switzerland, March 2003.

[7] "Intel IXP2400 Network Processor", [http://www.intel.com/design/network/products/npfamily/ixp2400.htm], Accessed on 8/26/2005.

[8] Schulzrinne, H. and S. Casner, "RTP: A Transport Protocol for Real-Time Applications", Request for Comments (RFC) 3550 (proposed standard), Internet Engineering Task Force, July 2003.

[9] International Telecommunication Union, "Packet-based multimedia communications systems," Recommendation H.323, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, July. 2003.

[10] "Speak Freely." [http://www.speakfreely.org]. Accessed on 8/26/2005.

[11] "OpenH323." [http://www.openh323.org/]. Accessed on 8/26/2005.

[12] "WinRTP." [http://www.vovida.org/applications/index.html]. Accessed on 8/26/2005.

[13] "NetPerf." [http://www.netperf.org/netperf/NetperfPage.html]. Accessed on 10/21/2005.