A new method for VoIP Quality of Service control using combined adaptive sender rate and priority marking

Zizhi Qiao, Lingfen Sun, Nicolai Heilemann and Emmanuel Ifeachor
Centre for Signal Processing & Multimedia Communications, School of Computing, Communications and Electronics
University of Plymouth, Plymouth PL4 8AA, UK
{z.qiao;l.sun;e.ifeachor}@plymouth.ac.uk;n.heilemann@gefahrstoffeinheit.de

***Abstract* — Quality of Service (QoS) control is an important issue in Voice over IP (VoIP) applications because of the need to meet technical and commercial requirements. The main objective of this paper is to propose a new QoS control scheme that combines the strengths of adaptive rate and speech priority marking QoS control techniques to provide a superior QoS control performance, in terms of perceived speech quality. A second objective is to propose the use of an objective measure of perceived speech quality (i.e. objective MOS score) for adaptive control of sender behaviour as this provides a direct link to user-perceived speech quality, unlike individual network impairment parameters (e.g. packet loss and/or delay). Our results show that the new combined QoS control method achieved the best performance under different network congestion conditions compared to separate adaptive sender rate or packet priority marking method. Our results also show that the use of an objective MOS as the control parameter for the sender rate adaptation improves the overall perceived speech quality. The results reported here are based on a simulation platform that integrates DiffServ enabled NS-2 network simulator, a real speech codec (AMR codec) and the ITU-T standard speech quality evaluation tool (PESQ).**

Keywords — Voice over IP, Quality of service (QoS) control, Adaptive bit rate, Priority marking, Perceived speech quality, AMR codec, PESQ.

I. INTRODUCTION

The convergence of communications and computer networks has led to a rapid growth in real-time applications such as Internet Telephony or Voice over IP (VoIP). However, IP networks are not designed to support real-time applications and factors such as network delay, jitter and packet loss lead to unpredictable deterioration in the perceived voice quality. A major challenge that faces network operators/service providers is how to control Quality of Service (QoS) to meet technical, legal and commercial requirements.

QoS control mechanisms for VoIP should aim to make optimum use of available network/terminal resources and to minimise the effects of network impairments on voice quality. Several approaches exist to realise QoS control, but most seek to control the information flow from the audio/video sources, adaptively, in accordance with significant changes in the network. An important class of QoS control technique involves rate control (i.e. QoS control is achieved by automatically adjusting the send bit rate depending on network congestion conditions). However, current rate control mechanisms [1-3] are based largely only on network impairments such as packet loss rate or delay during congestion. The strategy is to control the sender behaviour, using the network impairments, from the receiver or the network node but this may not be sufficient to provide optimum QoS, in terms of the voice quality delivered, because the control information is directly linked to user perceived quality.

A second important class of QoS control techniques exploits knowledge of the fact that different parts of speech have different perceptual importance and so do not contribute equally to the overall voice quality [4-6]. In this approach, voice packets that are perceptually more important are marked, i.e. given priority, and so are less likely to be dropped than packets that are of less perceptual importance, if there is congestion. The priority marking based QoS schemes are open loop and do not make use of changes in the network impairments.

The main objective of this paper is to investigate the possibility of combining rate adaptation control technique with priority marking, to exploit the advantages of the two approaches to provide a robust control scheme which delivers optimum QoS in terms of voice quality. In rate control schemes, the cost of adapting the data flow to changes in the network is that some packets may be dropped randomly when congestion occurs and this will increase the packet loss rate. However, in priority marking schemes important packets are dropped less and delayed less. Thus, the combined scheme should provide improved overall user perceived quality. DiffServ is used to implement the scheme and employs different queuing methods, the most important of which is a variation of random early drop queue (RED queue). RED not only gives different packets different drop probabilities, it also gives the receiver hints about whether congestion has occurred or about to occur. With a proper feedback mechanism, this information can be used to control the send bit rate.

The main contributions of this paper are twofold. First, we propose a new QoS control scheme that combines the strengths of the adaptive rate control technique and speech priority marking QoS technique to provide a superior QoS control performance than hitherto possible. Second, we propose the use of an objective measure of perceived speech quality (i.e. objective MOS score [11]) instead of individual network impairments (e.g. packet loss and/or delay) to control

The work is supported in part by an EU Grant under the Sixth Framework Programme (BioPattern project, No. 508803).

sender behaviour as this provides a direct link to user-perceived speech quality.

Preliminary results show that by exploiting the strengths of both methods, the new scheme achieved the best perceived quality compared to rate-adaptive, marking and no control schemes under different network congestion conditions. The results are based on extensive simulation in an environment that integrates the NS-2 network simulator [12], adaptive speech codec (the AMR codec [8]) and an objective perceived speech quality measurement system, which is based on the ITU-T speech quality evaluation standard [7].

The remainder of the paper is structured as follows. In Section II, we present the three QoS control schemes – the rate-adaptive, priority marking and the new combined, QoS control schemes. The simulation system and experiments are described in Section III. The results and discussion are given in section IV, and Section V concludes the paper.

II. QoS CONTROL SCHEMES

A. Rate-adaptive QoS Control Scheme

The adaptive rate QoS control scheme is based on the AMR (Adaptive Multi-Rate) [8] speech codec. The AMR codec was developed by ETSI and has been standardized for GSM. It has been chosen by 3GPP as the mandatory codec. It is a multi-mode codec with eight modes (MR475 to MR122) with bit rates between 4.75 and 12.2 Kb/s. Mode switching can occur at any time (frame-based). Thus, the AMR codec is well suited to rate control.

The adaptive rate QoS control scheme is shown in Fig. 1. In the scheme, the send rate of the AMR codec is adjusted in accordance with the network conditions to achieve the best possible QoS. The bit rate control mechanism is based on individual network parameters (e.g. packet loss rate and delay) or on the predicted, perceived speech quality (e.g. MOS score). In VoIP applications, the feedback information can be sent via RTCP reports. The specification of RTP [10] stipulates that the RTCP traffic does not exceed 5% of the whole traffic and that the time between the reports is at least 5s.

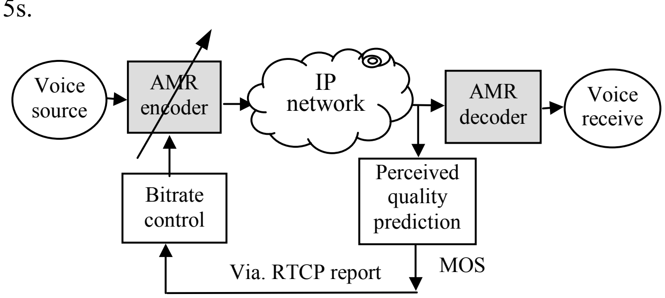

Fig. 1. Rate-adaptive QoS control scheme

The two important modules in Fig.1 are the bit-rate control module at the send side and the perceived quality prediction module at the receive side. Approximately, every 5s (the time interval between RTCP reports), a measure of the perceived conversational speech quality (i.e. MOS score) is predicted from network parameters (e.g. packet loss and delay) using a PESQ based method we proposed in a previous study [11].

The bit rate control module is used to adapt the send bit rate in accordance to the feedback information. The adaptive algorithm used in the module follows the 'additive increase/ multiplicative decrease' concept that has been successfully employed in other congestion control algorithms, such as TCP and ABR [3]. The basic idea is that the AMR codec can reduce its bit-rate (if possible) when there is network congestion and increase its bit-rate when no congestion is detected. The AMR rate is then used to predict the packet loss and the MOS. The predicted MOS score is compared with the existing MOS and the controller will choose the best step to change or keep existing AMR rate. The control mechanism used is presented in Section III.

B. Priority Marking QoS Control Scheme

In rate-adaptive QoS control scheme, it is assumed that all the packets within a flow are equally important. Previous research [4][9] has shown that some speech segments are more important than others and is the basis for the priority marking control scheme, which is depicted in Fig.2. Each speech frame is marked differently depending on its perceptual importance. For example, the priority-marking module marks the beginning of a voiced segment (e.g. the first 5 or 10 frames of a voiced segment for the AMR codec) as high priority (e.g. marked as a 'premium' class), while others are marked as perceptually unimportant (e.g. marked as a 'best-effort' class). When there is network congestion, the perceptually unimportant frames have a higher drop probability. This protection scheme results in a lower loss probability for packets with high priority and can lead to a better perceived QoS compared to no-control scheme.

Priority marking QoS control scheme can be implemented in networks that support Differentiated Services (DiffServ) architecture [6], e.g. a simplified 2-bit marking DiffServ implementation [13].

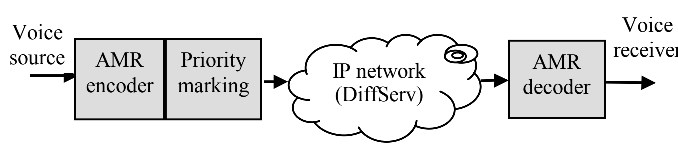

Fig. 2. Priority marking QoS control scheme

C. Combined Rate-Adaptive and Priority Marking QoS Control Scheme

As discussed above, the rate-adaptive QoS control scheme is based mainly on an objective MOS score at the receiver and packet loss contributed significantly to the measured or predicted MOS score. To reduce the long delay caused by simple over buffered drop tail queue, the network operators commonly use the RED queue or a similar queue management method to provide a better congestion notification and control. The use of RED queue method in the system makes it logical to try to link send rate adaptive control with packet priority marking because it is relatively easy to set different queues or virtual queues in a RED queue management system to provide different treatment for different priority packets.

Priority marking should reduce the loss or delay of important packets. An important goal is to investigate whether the overall perceived speech quality can be improved further by combining rate-adaptive and priority marking control schemes. This is the motivation of the proposed combined QoS control scheme, which is shown in Fig.3.

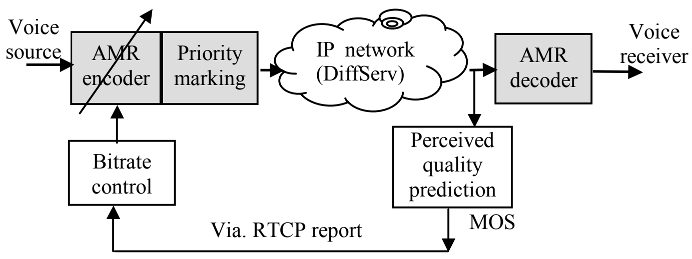

Fig. 3. Combined QoS control scheme

As shown in Fig.3, the bit rate of the AMR codec is adjusted in accordance with the objective, predicted MOS and, at the same time, the perceptually important segments of speech are protected by priority-marking. Potentially, this should make it possible to optimise the perceived speech quality for VoIP applications using AMR codec.

III. SIMULATION SYSTEMS AND EXPERIMENTS

A. *Simulation system*

The combined QoS control scheme was set up as shown in Fig.4. It consists of three main parts: (I) An NS-2 network simulator [12] to simulate multiple VoIP flows and IP networks with congestion; (II) a VoIP simulation system to simulate VoIP flow, which includes an AMR encoder/marker, loss simulator, decoder, and control modules, and (III) a perceived quality evaluation system to provide a measure of the overall speech quality and quantify the performance of each control method.

We simulated a simple bottleneck network topology using NS-2, as shown in Fig.4 (I). A total of N adaptive AMR sources were simulated for VoIP traffic (this assumed that the available bandwidth was shared among these UDP sources). All the sources were set as constant bit rate (CBR) UDP source (in order to match with the simulation of VoIP flow in part (II), as VAD for AMR codec was not activated there). The sender bit rate (plus header) was set according to the required bit rate for adaptive AMR codec (CBR source can change the send rate by request but still using the name CBR). All flows sent by traffic source was traced and the loss information collected and sent back to the loss simulator in part (II). A VoIP flow was simulated via encoder/marking, loss simulator and decoder. The loss information was also sent to the quality prediction module to obtain a MOS score. The MOS score was then fed back to the send side for bit rate control. A single hop of 2Mbit/s bandwidth, representing a bottleneck link, was set in a DiffServ enabled IP network. With the increase in the number of simultaneous users sharing the bottleneck link, we were able to investigate the performance of different QoS control methods under different network congestion situations. The overall performance of each of the different QoS control methods is evaluated by the evaluation system (III).

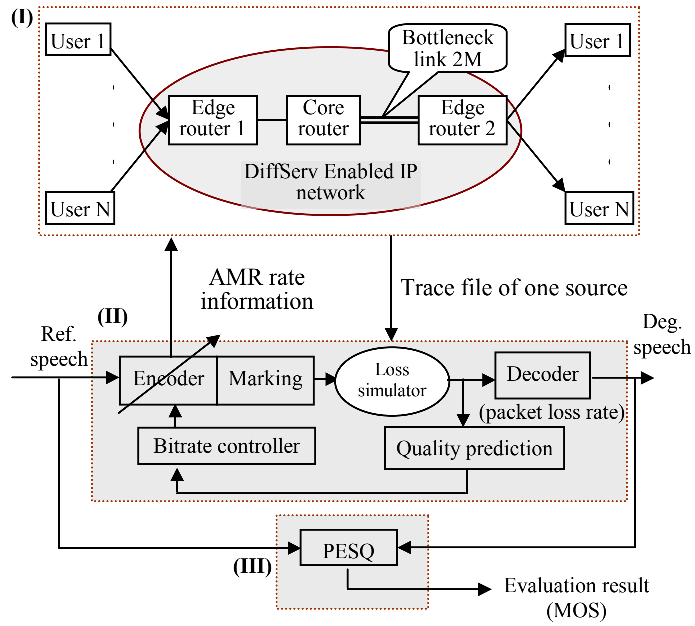

Fig.4. Simulation system for combined QoS control scheme

Perceived speech quality prediction is based on the ITU PESQ (Perceptual Evaluation of Speech Quality). For every loop (e.g. every 5 seconds), speech quality (MOS) is calculated in the 'quality prediction' module, depending on the network packet loss and AMR mode. In practice, this predicted voice quality would be used non-intrusively as described in [11]. The predicted speech quality is also used to evaluate the overall quality of the control schemes as shown in Fig.4 (III).

In the study, the MOS value was computed with the ITU-T PESQ by comparing a reference sample speech file with the degraded sample speech file, which was generated by processing the reference sample speech in accordance with the AMR rate and packet loss rate. More detailed explanation was described in [15].

B. *Priority marking and loss simulator*

Every frame generated from the AMR encoder was marked as perceptually important or unimportant, depending on the information from the AMR coder. In the simulation, the parameter that indicates whether it is voiced/unvoiced for each frame was extracted directly from the decoder's voiced_hangover flag for simplicity. Reference [4] and [5] gave more detailed explanations of packet marking.

The priority-marking scheme can be readily implemented in DiffServ supported networks. DiffServ is implemented in NS-2 version higher than NS2.1b8a and our simulation used NS2.1b9a to support the DiffServ simulation part. For simplicity, the DiffServ policyer used for our simulation is a Time Sliding Window with 2 colour marking policyer (built-in function supported by NS-2). It uses CIR and a drop precedence of two levels. The basic idea is that a lower precedence is used when the CIR is exceeded. The default scheduling mode is Round Robin.

C. Perceived speech quality driven rate-adaptive control simulation

The bit rate control module aims to detect the optimal bit rate settings that would yield the best perceived speech quality under a given network condition. A perceived speech quality MOS score is used as a control metric to drive the control mechanism. MOS is predicted at the receive side for each RTCP interval (e.g. 5 seconds) and then sent back via the RTCP report. A predicted MOS is then calculated and compared with reported MOS. If the MOS after rate adaptive control are predicted better, the rate will be changed otherwise it will stay unchanged. This mechanism can get the balance between rate reduction and congestion decrease. A MOS driven rate adaptive control loop pseudo code is shown in Fig.5.

```
For each RTCP report  //every 5 seconds
{ bitrate_old = bitrate_new;  //save existing status
  MOS_old = MOS_new;  // get the new status
  get MOS_new from RTCP report;
  // compare MOS scores
     if  (MOS_new > MOS_max) goto NOCHANGE
     else if ((MOS_new – MOS_old > threshold1) &&
                                    (bitrate_old != bitrate_max))
         bitrate_new=next_higher_bitrate ; //increase the bit rate
     else  if  ((MOS_old – MOS_new) > threshold2)
         bitrate_new=next_half_lower_bitrate; //halve the bit rate
     else if (( threshold1 < MOS_old – MOS_new < threshold2 ) &&
                                   (bitrate_old != bitrate_min))
         bitrate_new=next_lower_bitrate; //decrease the bit rate
     else
     NOCHANGE: bitrate_new=bitrate_old; // no change of bit rate
     // Predict MOS after rate change
     get MOS_predicted from PESQ{AMR_rate, lossrate_predicted}
     if (MOS_new>MOS_predicted)  bitrate_new=bitrate_old;
     // no change of bit rate
     eles send bitrate_new to sender;  //control the sender
}
```

Fig.5. Rate-adaptive control loop pseudo code

MOS prediction is based on the ITU PESQ measurements for a given AMR rate and packet loss rate (As described in section A of this Section).

$$\text{MOS} = \text{PESQ}\ \{\text{AMR rate, loss rate}\} \qquad (1)$$

The predicted packet loss in (1) is based on the following equation:

$$\text{loss rate} = (\ MR * N - BW) / (MR*N) * 100\% \qquad (2)$$

Where loss rate is the predicted packet loss rate for next control step, MR is the single user transmit rate considering header and next step AMR frame payload, N is the number of users and BW is the bandwidth they are shared. This is a simplified equation and does not consider the effect of a limited buffer size and the distribution of arriving packets but this will not affect the main predict MOS control idea.

In the simulation, the threshold1 in Fig.5 was set to 0.2 in order to avoid unnecessary oscillation, and the threshold2 was set to 0.5 to indicate an obvious decrease in perceived speech quality. The maximum MOS (MOS_max) achievable for the AMR codec was set to 4. For the AMR codec, the maximum bit rate, bitrate_max, was set to 12.2kbit/s and the minimum bit rate, bitrate_min, was set to 4.75kbit/s. For every control loop, the modified send bit rate was sent back to NS-2 simulator to adjust the source bit rate. For simplicity, we assume that all N sources in the NS-2 use the same AMR sender bit rate at the beginning and are adjusted to the same bit rate when adaptation occurs.

D. Simulation of combined rate-adaptive and priority marking method

The modules described above can be integrated to support the simulation of the new combined quality of service control method. Each user's packets are traced and recorded for evaluation. The packet size and packet loss information is used to process a reference speech to get a degraded speech. The degraded speech is then compared with reference speech using PESQ to get the evaluation result. The results of the simulation are discussed in next Chapter.

IV. RESULT AND DISCUSSION

In order to investigate how the QoS control schemes affect perceived speech quality under different network conditions, different network congestion scenarios are simulated using the network simulator NS-2. The bandwidth of the bottleneck link was set to a fixed value (2Mbit/s) with a delay of 1ms. The number of the streams sharing the link was increased from a small number to a large number to simulate different congestion scenarios. The starting point was 70 streams sharing the bottleneck when there is no congestion at all. The number of users was increased from 70 to 140 in steps of 5. By 140 users, almost every stream suffered from a very high loss rate and all the control methods were unable to cope with the impairments well. (Packet loss rate for non-control scheme was measured more than 40%). The latest investigation about PESQ method's performance in high packet loss situation [14] suggested that MOS score is much lower from PESQ result so we stopped to increase the user number after 140, as the result is meaningless.

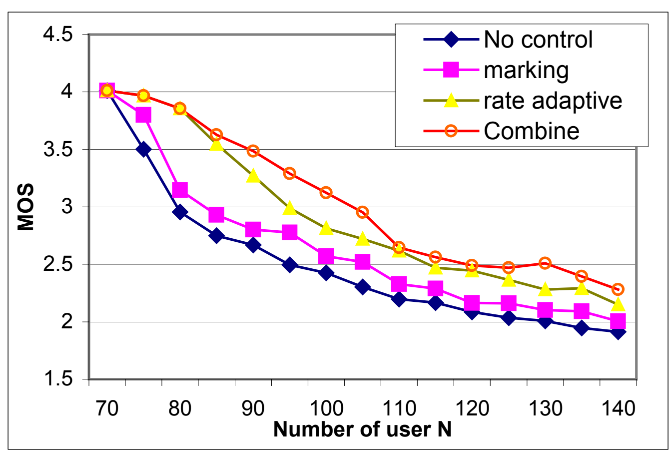

Fig.6 MOS vs. different control and non-control methods

In order to compare the performance between the different QoS control schemes and a "no control scheme", we also implemented the priority marking, the rate-adaptive and no control schemes. For the priority marking and no control schemes, the send bit rate of the AMR codec was set to a fixed mode (12.2 Kb/s). For rate-adaptive-only control method, the bottleneck link was set to a non-DiffServ link with the same delay parameters and the rest of the system remained the same. The simulation was carried out using the same scenarios as described previously. The number of simultaneous users was increased from 70 to 140 as before. Fig.6 compares the results for all four methods. The results show that for 70 simultaneous users, i.e. when there is no congestion, all four methods have the same performance. The MOS scores represent the highest score obtainable from an AMR codec.

In general, as shown in Fig. 6, the drop of the speech quality follows the similar pattern for all four schemes because they all suffering from the packet loss happened in the bottleneck link (see Fig.4). For the adaptive rate control scheme, the drop of speech quality is less steep compare with the "non-control" scheme. This is because the MOS driven rate adaptation can choose the best-optimised AMR rate to minimise the affect of codec rate decrease and packet loss increase. For the priority-marking scheme, the improvement over the non-control scheme is stable although not very significant. This is because although the DiffServ method can be used to treat different packets with different priority (i.e. loss rate), higher priority packets still have chance to be dropped, especially when the congestion is higher than CIR. From the figure, the performance of the new combined scheme is always better than those two different control schemes and the non-control scheme.

We also simulated a packet loss driven rate adaptive control to compare with the MOS driven rate adaptive control. The single packet loss rate control uses the same idea as shown in Fig.5 but use packet loss rate instead of MOS as the parameter to control the sender rate. Fig.7 gives the results.

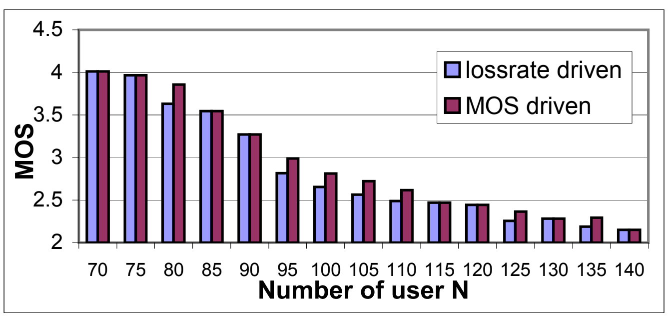

Fig.7 MOS vs. predict MOS driven and single parameter driven rate adaptive control

The results show that the predicted MOS driven method is slightly better than the single parameter driven rate adaptive control approach. The result is not linear because the control steps and AMR rate are not linear, and at some situation point, the predicted MOS driven method gives the same decision as the single packet loss rate driven method. Only the AMR rate and packet loss rate are considered in the MOS prediction and this is why the improvement is not very significant. But when more parameters are considered (e.g. delay, jitter etc.) in the MOS predicted process before rate adaptive control, the predict MOS driven rate adaptive control performance should provide a much larger improvement.

V. CONCLUSIONS

In this paper, we have proposed a new QoS control scheme that combines the strengths of rate-adaptive and priority marking QoS control schemes and uses an predicted objective measure of conversation speech quality as a control parameter. We investigated perceived speech quality for different QoS control schemes by integrating NS-2 network simulator with a real adaptive speech codec (the AMR codec) and a perceived quality evaluation system based on the ITU-T PESQ. We used the predicted perceived speech quality metric (measured by PESQ), instead of individual network parameters, to control the AMR codec's bit rate. Preliminary results show that the new control scheme achieved the best perceived speech quality compared with rate-adaptive, priority marking and no control schemes in different network congestion conditions. Our results also show that the use MOS as a parameter to control the send rate adaptation can improve the overall perceived speech quality.

In future, the investigation will be extended to include the application of the new combined method in a TCP/UDP mixed environment. The effects of delay in the DiffServ model and the use of conversational speech quality as metric to control AMR rate will be studied.

REFERENCES

[1] R. Eejaie, M. Handley and D. Estrin, RAP: An End-to-end Rate-based Congestion Control Mechanism for Realtime Streams in the Internet, Proc. IEEE INFOCOM'99, 21-25 March 1999, USA.

[2] F. Beritelli, G. Ruggeri and G. Schembra, TCP-Friendly Transmission of Voice over IP, Proceedings of IEEE International Conference on Communications, New York, USA, April 2002.

[3] A. Barberis, C. Casetti, J.C. De Martin and M. Meo, A Simulation Study of Adaptive Voice Communications on IP Networks, Computer Communications 24 (2001) 757 – 767.

[4] H. Sanneck, N T Long Le, M Haardt and W Mohr, Selective Packet Prioritization for Wireless Voice over IP, 4th International Symposium on Wireless Personal Multimedia Communication, Denmark, Sep. 2001.

[5] J. C. De Martin, Source-driven Packet Marking for Speech Transmission over Differentiated-Services Networks, Proceedings of IEEE ICASSP 2001, Salt Lake City, USA, May 2001.

[6] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, An Architecture for Differentiated Services, RFC 2475, IETF 1998.

[7] ITU-T Recommendation P.862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs.

[8] ETSI EN 301 704 V7.2.1 (2000-04), Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate (AMR) speech transcoding (GSM 06.90 version 7.2.1 Release 1998).

[9] L Sun and E Ifeachor, Impact of Packet Loss Location on Perceived Speech Quality, Proc. 2nd IP-Telephony Workshop, New York, 2001.

[10] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, RTP: a Transport Protocol for Real-time Applications, RFC 1889, IETF, 1996.

[11] L Sun and E Ifeachor, Prediction of Perceived Conversational Speech Quality and Effects of Playout Buffer Algorithms, Proceedings of IEEE ICC'03 , May, 2003, Anchorage, Alaska, USA.

[12] The Network Simulator - NS-2, http://www.isi.edu/nsnam/ns.

[13] K. Nichols, V. Jacobson and L. Zhang, A Two-bit Differentiated Services Architecture for the Internet, RFC 2638, July 1999.

[14] A. W. Rix, Comparison between Subjective Listening Quality and P.862 PESQ Score, Online Workshop MESAQIN (Measurement of Speech and Audio Quality in Networks) May 2003.

[15] A.E. Conway, A Passive Method for Monitoring Voice-over-IP Call Quality with ITU-T Objective Speech Quality Measurement Methods, Proc. of IEEE ICC2002, pp2583-2586, April 2002, New York, USA