

Evaluation of VoIP Quality over the Pakistan Internet Exchange (PIE) Backbone

M. Amir Mehmood
Pakistan Internet Exchange,
IT Infrastructure Division,
PTCL, Lahore, Pakistan
amir.mehmood@ptcl.net.pk

Tariq M. Jadoon
Lahore University of
Management Sciences,
Lahore, 54792, Pakistan
jadoon@lums.edu.pk

Noor M. Sheikh
University of Engineering
and Technology, Lahore,
54890, Pakistan
adstec@mailcity.com

Abstract

This paper assesses VoIP quality over the Pakistan Internet Exchange (PIE) backbone between major cities of Pakistan. Passive measurements for voice calls in the presence of background Internet data traffic for different speech codecs such as the ITU-T G.711, G.723.1 and G.729 were carried out. Inter-packet delay and jitter were measured over the PIE IP/ATM backbone and the corresponding inter-packet delay densities were calculated for evaluating the perceptual quality of computer telephony calls using the ITU-T G.107 E-model. The transmission quality (R-factor) and resultant Mean Opinion Scores (MOS) were calculated for both backbone and access links of ISPs. The study highlights the growth of VoIP as a significant component of the PIE backbone traffic in Pakistan as well as reconfirms the presence of congestion hot spots in the access links.

1. Introduction

For the Internet to evolve into a true Integrated Multimedia Broadband Network it is essential that it provides a similar level of service that we are used to on conventional dedicated legacy networks. Converging telephone and IP networks entails providing similar to “toll-quality” service over best-effort IP networks.

As voice quality is ultimately adjudged by the listener, speech quality is inherently *perceptual* or subjective in nature. A widely accepted measure for the subjective speech quality is the Mean Opinion Score (MOS). MOS ranks perceptual quality on a scale ranging from 1 to 5, 1 meaning bad and 5 excellent [1]. However, assessing speech quality through surveys is a time consuming and expensive process. A viable alternative is to develop *quality models* that simulate human rating behaviour by correlating perceptual QoS

with other quantifiable parameters. A number of quality models that provide objective MOS measures have been developed. Two of these models have been standardised by the ITU-T i.e. the Perceptual Speech Quality Measure (PSQM) [2] and the Perceptual Evaluation of Speech Quality (PESQ) [3]. Both models compare a reference and degraded version of the speech samples. An alternative transmission planning tool is the E-model defined in ITU-T G. 107 and provides a transmission quality factor (R-factor) on a scale of 0 to 100 [4]. The R-factor maps into an MOS through a simple transformation.

Voice over IP (VoIP) essentially entails sending speech as packets over the Internet. The original acoustic signal is sampled at a fixed rate and generates a stream of bits. This bit stream is either directly filled in packets or is processed further to achieve compression before being prefixed with RTP/UDP/IP headers. Thus, a sample must wait for an algorithmic, processing and packetization delay before it can be placed on the wire. VoIP packets that traverse the Internet are subject to two principal impairments namely, packet loss and packet delay. Loss is either a consequence of congestion or a failure of network elements such as links and/or nodes. Packet delay has a fixed component as a consequence of the propagation and transmission delays as well as a variable component as a result of the potentially variable queueing delay that packets encounter due to randomly full-buffers at intermediate hops whilst traversing the Internet. Thus, inter-packet delay at the receiver is not the same as at the sender resulting in jitter. An appropriately sized de-jitter or playout buffer can compensate for most of the jitter at the cost of an added delay. Delay jitter indirectly manifests itself as packet loss for packets that arrive latter than a maximum threshold and adversely contributes to the perceived listening quality of one-way speech. The two-way conversational quality of a call is primarily affected by the end-to-end delay in addition to the packet loss and

delay jitter. These parameters constitute the network quality of service (QoS) parameters and can be mapped to a perceptual QoS measure such as the MOS through quality models. An example of these is the E-model. The QoS ultimately depends on the end-to-end conditions or mouth-to-ear (M2E) delay and delay jitter. A debate that often ensues amongst backbone and access service providers is that who is to blame for poor QoS. Is it the backbone service providers? or is it the access service providers? Thus, collecting network statistics that can help resolving such conflicts are essential. This work is an effort in this direction and provides essential statistics of the inter-packet delay variation (jitter) distributions of both access and backbone links for a number of different scenarios and subsequently calculates their MoS for VoIP calls using the E-model.

The balance of the paper is organized as follows. Section 2 briefly reviews the E-model. Section 3 describes the Pakistan Internet Exchange (PIE) as well as the experimental set-up for the measurements. Section 4 presents the experimental results and finally Section 5 concludes the paper.

2. E-Model

The E-model is a computational model that determines a transmission quality rating called the ‘R’ factor from the transmission parameters to predict the quality of the M2E speech path. The typical range for the R factor is 0-100 for narrow-band telephony and corresponding values of the MOS are shown in Fig 1.

The basic principle of the E-model is that the various impairments contributing to the overall perception of voice quality are additive when converted to the appropriate psycho-acoustic scale (R). The basic formula for the E-model is:

$$R = R_0 - I_s - I_d - I_e + A \quad (1)$$

where R_0 is the signal to noise ratio of the connection. I_s are impairments simultaneous to voice signal transmission. I_d are impairments delayed after voice signal transmission. I_e are equipment impairments. A is the advantage factor given to a particular voice connection. The default values result in $R = 93.2$. (using the year 2000 revision Annex A of [4]). The E-model is a widely used computational model and several enhancements have been proposed to the basic model to cater for VoIP traffic [5]. Cole and Rosenbluth [6] outline a methodology for incorporating packet loss and delay variations as a reduction in base transport metrics through I_e , the equipment impairment factor.

R	User Satisfaction	MOS
100	Very Satisfied	5
93.2		4.4
90	Satisfied	4.3
80		4.0
70	Some users dissatisfied	3.6
60		3.1
50	Many users dissatisfied	2.6
0		1

Fig. 1: Voice quality classes

Basically, the end-to-end delay variations result in loss at the de-jitter buffer due to the arriving packet stream under-flowing or overflowing the decoder's de-jitter buffer and this is incorporated by adjusting I_e appropriately. Other components that contribute to a further reduction in QoS are the end-to-end delay as well as the network loss.

3. The Pakistan Internet Exchange (PIE)

The Pakistan Internet Exchange (PIE) was created in 2000 to cater for connectivity needs via a single core backbone for the whole country by providing peering points to ISPs. Domestic Connectivity to other cities within Pakistan is either through STM-1 (155Mbps) or E3 (34Mbps) links. Data was collected at the *PIE Lahore Node* over a period of two months from mid-March to mid-May 2005 using Sniffer Portable from Network General in four fortnightly batches capturing all traffic traversing the Lahore node. Each batch consisted of approximately 1.1×10^{10} packets. A protocol wise break up showing 95% confidence intervals of the traffic mix is presented in Fig. 2. The Protocol wise break-up reveals that approximately 30% of the total data traffic is web traffic. SMTP, POP, DNS, MSN and Yahoo Messenger add up to another 20%. The remaining 50% traffic is categorised as “OTHERS”. In order to investigate how much of the “OTHERS” traffic was VoIP. The packet sniffer was once again set to capture RTP, RTCP and SIP during peak hours (1500-0230 hrs). The average RTP traffic captured is 6.5%. Most of this traffic was G.723.1, G.729 or G.711. This indicates that VoIP traffic is growing to become a significant component on the PIE backbone. Fig. 3 shows the distribution of packet sizes passing through PIE Lahore node for four consecutive days

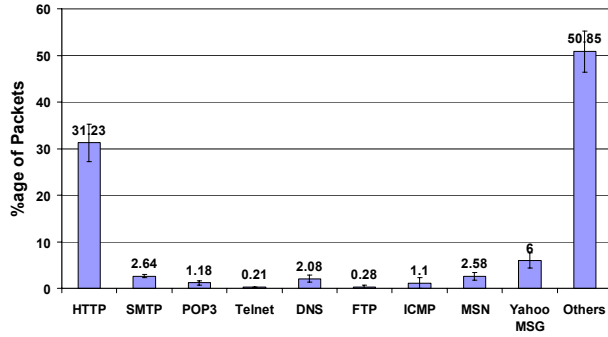


Fig. 2: Protocol-wise break-up

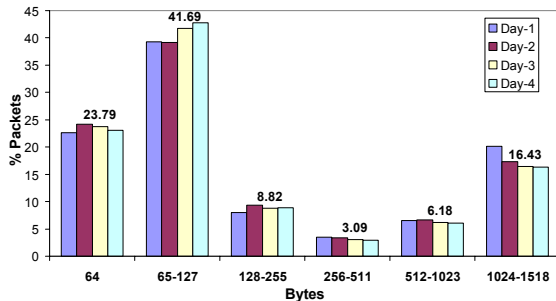


Fig.3: Packet-size distributions

From Fig. 3 it is evident that the packets are predominately either small-sized packets (real-time & ACKS) or large-sized web packets.

4. Measurement Setup

For assessing the VoIP quality over the PIE backbone, five different points of presence (POPs) on PIE were selected. The five POPs were:

- LER – Egerton Road node Lahore
- LGB – Gulberg node Lahore
- KHI – Karachi node
- RWP – Rawalpindi node
- GJR – Gujranwala node

Four sets of experiments were performed by setting-up a VoIP call from a soft-phone connected to the LER distribution router to another soft-phone connected to:

- A) a core/distribution router in another PoP on the PIE backbone.
- B) to an access router
- C) to a dial-up modem
- D) to a cable modem

as indicated by scenarios A, B, C and D in Fig. 4.

Voice sessions were established using a Session Initiation Protocol (SIP) based soft phone client, i.e. Microsoft MSN Messenger ver 7.0 as shown in Fig. 4.

The MSN Messenger was selected due to its low average mouth-to-ear M2E delay as compare to other IP phones or soft client phones [7]. The calls were established from an Intel 1.5 GHz Centrino processor based laptop ‘the monitoring station’ connected through a fast Ethernet link to a Cisco 7500 series distribution router placed at LER to other machines with soft phone clients connected to core/distribution routers in other cities. It is important to note here that the traffic characteristics presented to the VoIP applications are a snapshot of the instantaneous traffic conditions presented to the VoIP streams. The voice streams were captured through an open source packet capturing software Ethereal Ver. 0.10.10 installed on the monitoring station. The VoIP stream from the soft phone is encoded as RTP/UDP/IP packets and is transmitted as the forward RTP stream while the captured voice streams from the ‘far’ side is the reverse RTP stream. Jitter calculations are carried out in accordance with RFC 3550 [8] for both directions. The inter-packet delay (IPD) or time-spacing and the IPD variation also called the ‘jitter’ for VoIP packets coming from the far end is calculated for all scenarios by capturing VoIP packets at the monitoring station and analyzing packets using network protocol analyzer Ethereal Ver 0.10.10. Subsequently, the R-factor and MOS of voice calls on each scenario is calculated using the E-model in the same vein as suggested by Cole and Rosenbluth [6].

4.1 Scenario- A - Domestic Backbone links.

The measurements were made from the monitoring station to soft phones connected to a distribution router on four other PoPs, i.e. KHI, RWP, LGB, GJR. Essentially, these are domestic backbone measurements on PIE. The calls were made using the ITU-T G.711 PCM μ -law codec during peak hours in order to estimate the worst jitter that VoIP packets can experience. During these calls very low packet loss (0.02% to 0.11%) and variation in the IPD spacing was observed as most of the links were not loaded to capacity as can be seen Table 1. Jitter is a smoothed function of the IPD spacing [8]. The jitter for the four voice connections is shown in Fig. 5 while, Fig. 6 shows the inter-packet delay (IPD) density distribution. From Fig. 5 it is clearly evident that the jitter is fairly low on the PIE backbone albeit there are some spikes of large amplitude similar to those discussed in Markopolou et. al. [9]. Moreover, Fig. 6 shows that the distribution of the IPD density is very similar for the voice connections on the core network between a number of cities for a wide range of loads as indicated by Table 1.

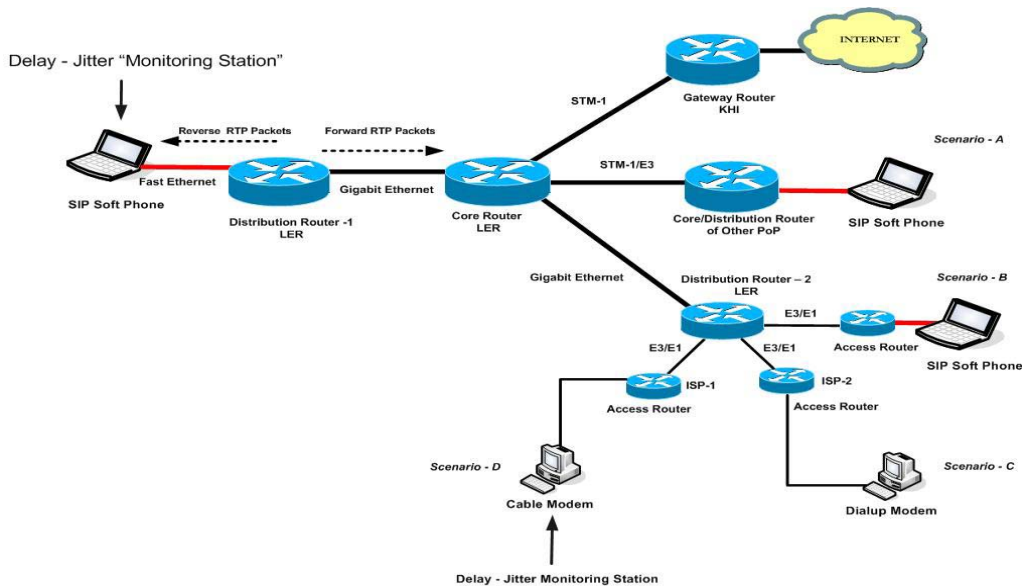


Fig. 4: VoIP delay jitter measurement set-up

4.2 Scenario-B - Domestic Access links

The second set of measurements was taken for evaluating the delay and jitter characteristics for access links. The soft-phone was connected to an access router connected to a remote PoP by a fast-Ethernet link as shown in the reference configuration of Fig. 4. Fig. 7 shows the IPD density for VoIP calls at two loads on the access links, i.e. between the distribution router of the remote PoP and its corresponding access router. The two loads are 50 and 95% of the respective E1 link capacities. Once again the voice calls were made using the ITU-T G.711 PCM μ -law codec with a similar packet size as in scenario A. It is evident from Fig. 7 that as the load increases from 50 to 95% the distribution of the IPD density splits from a single hump resembling the IPD of the backbone link to a bi-modal distribution with two distinct humps. This bimodal distribution indicates that a significant number of packets get severely delayed and make up the second hump. A delayed packet followed by a packet on time results in a clumping of packets at an IPD earlier than the mean and consequently the first hump. The measured average IPD is still close to 20ms. The packet loss varies from essentially no loss at a 50% load to over 2.5% packet loss at 95% load.

4.3 Scenario-C – Dial-up Links

The 3rd set of measurements was performed for dial-up links. This is still the most common way to access the Internet through PIE. The experiments were performed using the G.723.1 codec. Each VoIP frame was 78 bytes

in length and 54 of these bytes were overhead, while the remaining 24 bytes were compressed speech. The IPD spacing was 30ms. Fig. 8 shows the IPD density for VoIP packets received from the dial-up access link.

4.3 Scenario-C – Dial-up Links

The 3rd set of measurements were performed for dial-up links. This is still the most common way to access the Internet through PIE. The experiments were performed using the G.723.1 codec. Each VoIP frame was 78 bytes in length and 54 of these bytes were overhead, while the remaining 24 bytes were compressed speech. The IPD spacing was 30ms. Fig. 8 shows the IPD density for VoIP packets received from the dial-up access link.

Once again the bimodal nature of the density is clearly evident. However, this time the variance is much greater (49.31ms) as is evident from the spread of the distribution shown on a log-scale. Furthermore, approximately 2.57% packets are lost due to network congestion at the ISP link.

4.4 Scenario-D – Cable Modem Access Links

The final set of experiments used the G.729a to connect to a soft-phone attached to a cable modem indicated by *Scenario D* in Fig. 4. Each VoIP frame had a link-layer frame size of 94 bytes. 54 of these bytes made-up the RTP/UDP/IP/Ethernet header and the remaining 40 bytes were the payload containing compressed speech frames. The IPD spacing was 20ms.

Two sub-sets of results were obtained for link loads of 47.8% and 97.6% of the E3 links between the ISP and distribution routers. Once again an increase in the load results in greater jitter as is evident from Fig. 9 where the standard-deviation of the IPD increases from 1.455 to 6.161 when the link utilization increases from 47.8% to 97.6%. However, the distribution still does not degrade into a bimodal distribution for the cases studied.

4.5 E-Model and MOS Calculations

Calculating the E-model R-factor and MOS of voice calls for each scenario requires calculating an effective equipment impairment factor I_{e-eff} given by:

$$I_{e-eff} = I_e + (95 - I_e) \cdot \frac{P_{pl}}{P_{pl} + B_{pl}} \quad (2)$$

where I_e is the equipment impairment factor. P_{pl} is the packet-loss percentage and B_{pl} is the packet-loss robustness factor [4]. These values are codec dependent and have been taken from ITU-T Recommendation G.113 Appendix I [10]. A conservative relationship between the loss rate and buffer size can be derived using Chebyshev's inequality in the same vein as Cole and Rosenbluth [6]. One can calculate the loss as a result of jitter variations resulting in loss at the dejitter buffer as a function of buffer size b in packets. The total loss is given by: $P = P_{pl} + (1 - P_{pl}) \cdot P_{dejitter_buffer}$

(3)

and this replaces P_{pl} in (2) to calculate a new I_{e-eff} incorporating the loss due to jitter for a buffer size of $b = 3$ packets and becomes the ultimate I_e for (1)

I_d is not a significant component of (1) as the end-to-end delay is well within the 175ms bound [11]. The R values can now be calculated using (1) and the corresponding MOS scores can be obtained through the expressions:

For $R < 0$: MOS = 1

For $R > 100$: MOS = 4.5

For $0 < R < 100$:

$$MOS = 1 + 0.035R + 7.5 \times 10^{-6} R (R-60) (100-R) \quad (4)$$

Table II shows the corresponding values for the different scenarios described in the experiments. From the results in Table II, it is evident that the MOS values of the backbone links are highly acceptable while those of the access links have many users dissatisfied. The conditions in the dial-up access networks are unacceptable and would leave almost all users dissatisfied.

5. Conclusions

All in all, the conclusions are that VoIP is increasingly becoming a major component of the PIE backbone traffic. The PIE backbone links exhibits low mean jitter and support good quality voice. However, this reasonably well provisioned backbone network cannot achieve a satisfactory QoS on an end-to-end basis due to bottlenecks in the access networks. Thus, if the PIE backbone is to carry VoIP traffic with an acceptable QoS the access networks need to adequately provisioned in addition to the PIE backbone.

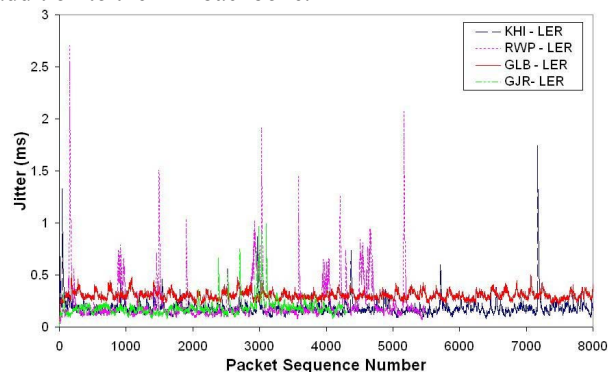


Fig. 5: Jitter Vs. Packet sequence number

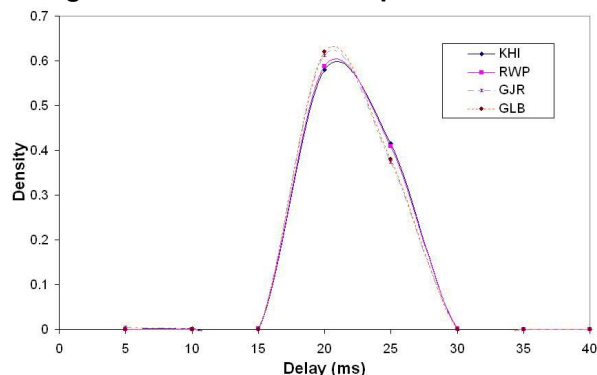


Fig. 6: IPD density of the core links (Scenario-A)

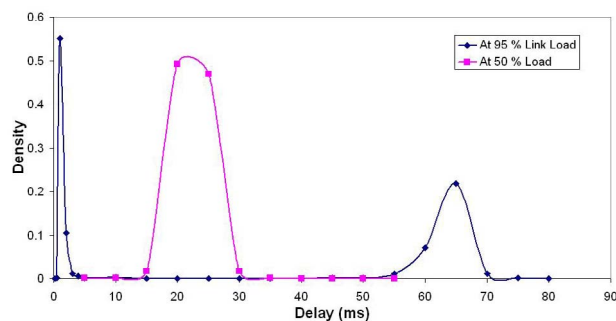


Fig. 7: Inter packet delay density for access links (Scenario-B)

Table 1. Mean jitter on backbone links

Link	Capacity	Load (%)	Mean Jitter (ms)
KHI-LER	OC-3	56.9	0.243
RWP-LER	OC-3	1.97	0.237
LGB-LER	OC-3	23.54	0.303
GJR-LER	E3	12.32	0.187

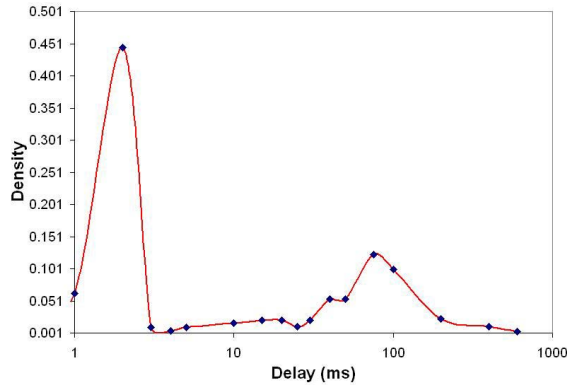


Fig. 8: IPD density for dial-up access links

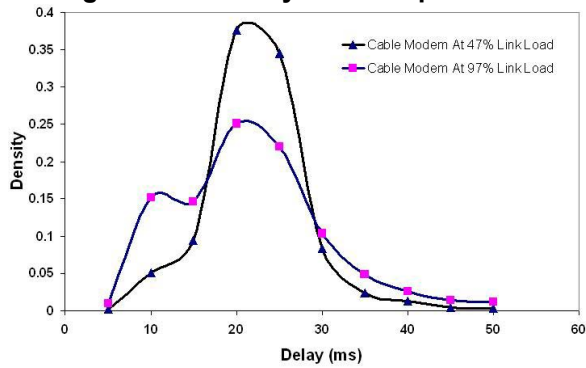


Fig. 9: IPD density for cable modem access

6. Acknowledgement

This work was supported in part by a research grant from the PTCL R&D Fund R&DF/Thematic-01/2004/06.

7. References

- [1] "Methods for subjective determination of transmission quality". ITU-T Rec. P. 800, Aug. 1996.
- [2] "Objective quality measurement of telephone-band (300-3400 Hz) speech codecs." ITU-T Rec. P. 861, Feb. 1998.
- [3] "Perceptual Evaluation of Speech Quality (PESQ): An Objective end-to-end speech quality assessment of narrow-band telephone networks and speech codecs." ITU-T Rec. P. 862, Feb. 2001.
- [4] "The E-Model, a computational model for use in transmission planning." ITU-T Rec. G.107, March 2003.
- [5] L. Ding, R. A. Goubran, "Speech quality prediction in VoIP using the extended E-model", *GLOBECOM 2003*, vol. 22, No. 1, Dec 2003
- [6] R. G. Cole, J. H. Rosenbluth, "Voice over IP performance monitoring." *ACM SIGCOMM Computer Communication Review*, vol.31 no.2, April 2001
- [7] W. Jiang, K. Koguchi and H. Schulzrinne, "QoS Evaluation of VoIP End-points." *IEEE International Conference on Communications (ICC)*, Anchorage, Alaska, May 2003.
- [8] H. Schulzrinne, et. al., "RTP: A Transport Protocol for Real-Time Applications", RFC 3550, July 2003.
- [9] A. Markopoulou, F. Tobagi & M. Karam, "Assessment of VoIP Quality over Internet Backbones", *IEEE/ACM Transactions on Networking* Vol. 11(5), Oct. 2003, pp. 747 – 760, 2003, ISSN:1063-6692
- [10] "Transmission impairments due to speech processing: Appendix I: Provisional planning values for the equipment impairment factor I_e and packet-loss robustness factor B_{pl} ", ITU-T Rec. G.113 Appendix I, May 2002.
- [11] "One-way transmission time", ITU-T Rec. G.114, May 2003.

Table 2: Perceptual quality assessment for VoIP for the PIE backbone

Scenario	Codec	Link	I_e	P_{pl}	B_{pl}	I_{e-eff}	μ ms	σ ms	$P\%$	I_e	R	MOS
A	G.711 (with no PLC)	KHI-LER	0	0.07	4.3	1.52	20	2.32	0.41	8.20	85.00	4.20
		RWP-LER	0	0.11	4.3	2.37	20	0.75	0.15	3.10	90.10	4.34
		GJR-LER	0	0.03	4.3	0.66	20	2.77	0.51	10.06	83.14	4.14
		LGB-LER	0	0	4.3	0	20	0.43	0.01	0.25	92.95	4.4
B	G.711 (with no PLC)	50%	0	0	4.3	0	20	2.24	0.31	6.46	86.74	4.25
		95%	0	2.5	4.3	34.9	20	27.8	49.59	87.42	5.78	1.00
C	G.723.1	dial-up	15	2.57	16.1	26.01	30	49.31	68.38	79.75	13.45	1.09
D	G.729a	48%	11	0	19	11	20	1.46	0.13	11.58	81.62	4.08
		97%	11	0	19	11	20	6.16	2.37	20.32	72.88	3.72