

significant implications to not only the circuit core but also the design, deployment, planning and, of an equal importance, the short-term and long-term investment of the backbone transmission networks.

While a debate can be staged on the merits of architectural attributes in R4 and R99 circuit core design, one of the most important concerns of operators is how the design impacts and enhances the delivery of revenue-generating services. Chief among these services is voice applications. Today, voice is still considered as the “killer” application, generating the bulk of service revenue for wireless network operators. In addition, second generation (2G) Global System for Mobile Communications (GSM) wireless subscribers increasingly expect a decent voice quality, as they are accustomed to having wired-line service quality as a benchmark for acceptable voice quality in wireless services.

The first objective of this paper is to summarize the results and findings of a series of tests designed to compare, quantitatively, the voice quality under the R4 architecture against the time division multiplexing (TDM)-based R99 design. In this paper, detailed descriptions of the test environment are also given.

IP is gaining acceptance in both the wired-line and the wireless world. With the explosive growth in Internet traffic and the maturing of IP technologies, IP has increasingly become the transmission medium of choice among telecommunication network operators. Wired-line voice over IP (VoIP) traffic is becoming ubiquitous as the reliability and voice quality of VoIP applications starts to rival that of legacy, TDM-based voice services. On the wireless side, while 3GPP Release 5 embarks the migration to all-IP environment, the “IP-evolution” has already started in the circuit core. With the introduction of decomposed media gateway environment in R4, IP can be used for both bearer and signaling transport. Real-time transport protocol (RTP) is specified to be one of the transportation options for voice bearer traffic in the Nb interface in R4.

The nature of IP network environment raises concerns among wireless network planners about the reliability of IP as a transport medium for voice traffic and its impact to voice quality. Unlike a TDM-based network, an IP network is a “shared” environment for any application, data and voice and there is no dedicated circuit for voice applications. The voice quality in an IP network is subject to the network conditions in the IP network, which can be characterized by a few key attributes, namely, packet loss, mean delay and delay jitter. Packet loss and delay distribution function of an IP network dictate how the packet traffic will behave in such a network environment. It is of particular importance to network planners who contemplate to deploy an IP-based UMTS circuit core to understand to what extent these attributes impact voice quality in R4 environment.

The second objective of this paper is to describe the findings and conclusions of tests designed to validate several hypotheses regarding the impact of these attributes (i.e. packet loss, mean network delay, and delay standard deviation) on the voice quality.

2. KEY CONCEPTS

A few key concepts that are worth discussing briefly first are stated in the following.

2.1. Voice quality degradation due to codec conversion

The principle motivation behind codec conversion in wireless networks is to reduce radio-link bandwidth requirements since the scarcest network resource is radio-link bandwidth. Pulse code modulation (PCM) voice sample at 64Kbps, while offers excellent voice quality, is not efficient for air-interface transmission. As a result, different standards-based codec conversion schemes have been developed to compress and decompress PCM voice samples (i.e. encoding and decoding schemes).

The increased bandwidth efficiency as a result of PCM coding compression is accomplished at the expense of voice quality and encoding/decoding delay. Note that voice quality will degrade when the 64Kbps PCM samples are compressed to a lower bit-rate, e.g., Adaptive Multi-Rate (AMR) 12.2Kbps. The extent of degradation varies from codec to codec. Additionally, individual codec compression schemes have different level of complexity, which will result in different processing delays (encoding or decoding). In addition to the processing delay, the compression process will incur scheduling delay as well as packetization delay (i.e. converting PCM samples into packets). To optimize voice quality in wireless networks, unnecessary codec conversion should be avoided to preserve the quality level as much as possible.

Compression scheme	PESQ Score
Clean ISDN Network	4.3
Analog Network (G.711)	4.1
G.728 Codec (16kbps)	3.8
G.729 Codec (8kbps)	3.6
G.723.1 (6.3 kbps)	3.5
GSM EFR (Enhanced Full Rate) Codec (12.2 kbps)	3.9
GSM FR (Full Rate) Codec (13 kbps)	3.5
GSM-EFR mobile network in typical operating range	3.6 to 3.1
GSM-EFR mobile network in very poor condition	2.2

Table 1: PESQ scores of different codec compression schemes and mobile networks [6]

3. HYPOTHESES

1) Voice quality of mobile-to-mobile (MTM) calls in R4 in an IP network under “no impairment” must be better than that in R99, and voice quality under “good” conditions is no worse than that in R99

While R99 has the advantage of utilizing dedicated circuits for transporting Nb user plane traffic, it has to perform transcoding (i.e. AMR to/from PCM transcoding) in both outbound and inbound direction, thus incurring voice quality degradation. In R4, TrFO implementation avoids unnecessary transcoding in MTM calls, which enhances voice quality. On the other hand, the voice quality is subject to IP network conditions.

There is no universal definition of “good” IP network conditions. In this paper, the definition specified by China’s Ministry of Information Industry is referenced. The following table describes the definition of good and poor network conditions.

Definition	Mean Delay (ms)	Delay Jitter (ms)	Packet Loss (%)
No Impairment	0	0	0
Good	100	10	1%
Poor	400	30	5%

Table 2: Reference definition of network conditions

To counter the impact of network delay fluctuation (i.e. jitter) and packet loss, MGWs typically implement a jitter buffer and voice quality enhancement mechanism such as error concealment technique. A well-design buffer with a generous buffer size (e.g. 80ms) should be able to wither the impact of delay distribution (mean and standard deviation) and packet loss when they are relatively moderate. By implementing an error concealment technique, MGWs can smooth out voice samples to mitigate the impact of packet loss.

It is expected that the advantages and disadvantages of transcoding and transmission in R4 and R99 be balanced out, at least under “good” IP network conditions. The net result is that the voice quality in R4 under “good” IP network conditions should be at least as good as that in R99.

2) Voice quality in R4 under “poor” IP network conditions is inferior to that in R99

Under “poor” IP network conditions, packet loss and delay jitter become severe enough that the voice quality will suffer significantly. First, when excessive packet loss occurs, packet extrapolation or packet duplication by an error concealment technique will not be able to adequately compensate the lost packets. Voice quality will be compromised. In addition, excessive packet loss will cause packets to arrive out of sequence, which degrades the voice quality. When the fluctuation in inter-packet arrival time (i.e. network delay jitter) becomes so severe, the MGW buffer will no longer be able to compensate for the variation of the arrival time. The buffer may have to treat the late arrival as dropped packets. The net result is voice quality in such a poor network environment will be inferior to that in R99.

3) Network delay has no impact to voice quality in R4

There are two aspects of call quality – conversation quality and voice clarity. Conversation quality deals with whether a conversation occurs naturally and interactively over the phone. Voice

In the test configuration of R4 CS domain, two MGWs are connected using IP transport for the Nb interface. Voice bearer traffic exchanged between the two MGWs uses RTP for the transportation protocol. An IP network impairment emulator is inserted to simulate impairment, such as packet loss, delay and jitter induced by an IP network. A sniffer using Ethereal (<http://www.ethereal.com>) is also attached for monitoring the RTP packets over the IP network. The IP network impairment emulator offers several options for different kinds of impairment. For example, packet loss can be generated periodically, randomly or in a burst. In the test, the Sniffer helps to verify the impairment introduced by the IP network emulator through monitoring the RTP packets carrying voice bearer traffic over the Nb interface. An AMR call generator and a VQT (Voice Quality Tester) are used for load generation and voice quality measurement respectively. In the tests, mobile-to-mobile calls are used. The AMR call generator emulates and provides loads in the access network so no radio network subsystem is required in the tests. As such, the voice quality measured is purely attributed to the circuit-switched core network.

It is known that voice quality measurement varies in a diverse range for different voice test samples even with the same voice codec, which is AMR in our case. In the test, a voice sample composed of a male, a female and a child's voice is used. The original recorded signal of the voice sample is in generic PCM .WAV voice format. The voice sample is then used and transformed into a digital format using 12.2 kbps AMR codec by the call generator.

Voice quality is measured at both directions of a call and at least ten replications are done to compute the average PESQ. A random distribution for the packet loss and a Gaussian distribution for the delay and jitter at the IP network are used in the tests. Tests are conducted with a combination of packet losses, different means and standard deviations for the Gaussian distributed network delay. In light of the importance of TrFO for voice quality improvement in R4, tests are conducted with TrFO. Voice quality test with TDM transport between the two MGWs (i.e. R99) is also performed as a baseline for comparison.

5. TEST RESULTS AND ANALYSIS

- *Hypothesis (1) - Voice quality of MTM calls in R4 under “no impairment” IP network conditions is better than that in R99; at least as good as that in R99 under “good” IP network conditions.*

Assumption:

- “Good” IP network conditions in this test is defined as
 - ♦ Packet loss at no greater than 1%
 - ♦ Delay distribution of $\mu = 100\text{ms}$ and $\sigma = 10\text{ms}$ (Normal distribution)

Test results:

Circuit core design	Packet loss (%)	Delay and Delay Jitter	Average PESQ
R4 (TrFO)	0	0	4.04
R4 (TrFO)	0	$\mu = 100\text{ms}, \sigma = 10\text{ms}$	3.96
R4 (TrFO)	1	$\mu = 100\text{ms}, \sigma = 10\text{ms}$	3.70
R99	Not Applicable	Not Applicable	3.80

Table 3: PESQ of R4 (TrFO) and R99 – “good” network conditions

As shown in the table above, R4 with TrFO enabled scores better than R99 in PESQ, under “no impairment” and “good” network conditions, with packet loss less than 1%. On the other hand, R99 outperforms R4 with TrFO enabled when the packet loss is at 1%. However, it should be noted that the numeric differences between the scores are relatively minor.

Test conclusion:

The test results partially validate hypothesis (1). Under 0% packet loss condition, R4 with TrFO actually outperforms R99 in PESQ. On the other hand, when packet loss is at 1%, the reverse is true.

However, R4 with TrFO and 1% packet loss, in practical sense, can be considered as in parity with R99 in PESQ. It is doubtful that the minor difference in PESQ will result in noticeable voice quality gap. From a practical sense, the hypothesis is validated by the test results.

- *Hypothesis (2) - Voice quality in R4 under “poor” IP network conditions is inferior to that in R99*

Assumptions:

As shown in Table 5 above, the average PESQ scores are almost identical between 100ms and 400ms delay mean value when packet loss and delay jitter are held constant. The results are further illustrated in Figure 3.

Test conclusion:

Based on the test results, it is concluded that voice quality is not influenced by the network mean delay value.

- *Hypothesis (4) - Higher network delay standard deviation (i.e. delay jitter) results in lower voice quality*

Test results:

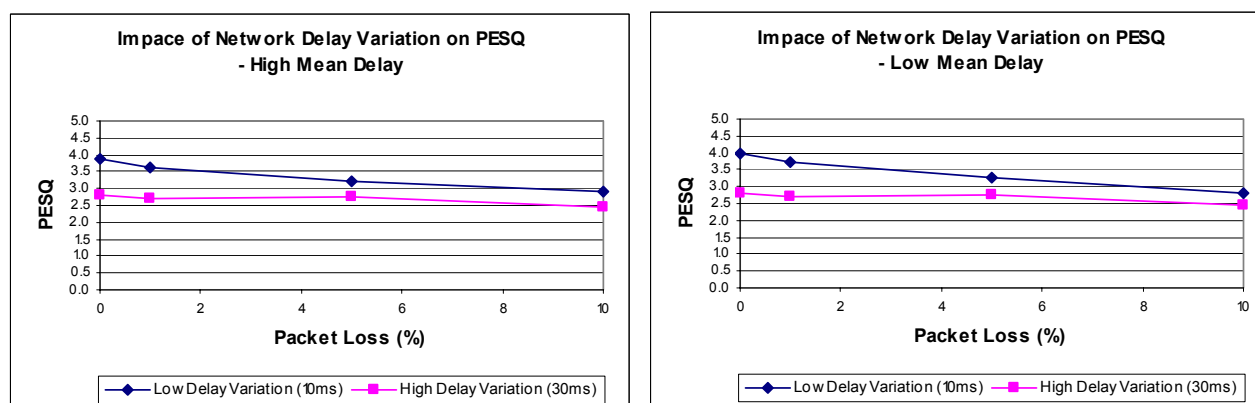


Figure 4: Impact of network delay variation on PESQ

To compare the effect of network delay variation on PESQ, both packet loss and network delay are held constant. As shown in Table 5 and Figure 4 above, PESQ scores are consistently higher in low delay jitter environment, regardless of the mean delay value.

Test conclusion:

Evidence from the tests supports the hypothesis that delay standard deviation adversely impacts the voice quality as measured in PESQ.

- *Hypothesis (5) - Higher packet loss ratio in IP network results in lower voice quality*

Assumptions:

For ease of comparison, the delay distributions are categorized into two groups, as described in the table below:

Category	Mean delay	Delay std. deviation
Good (100)	100	10
Good (400)	400	10
Poor (100)	100	30
Poor (400)	400	30

As proved above, mean delay has no significant impact to the voice quality in an R4 environment. To provide more data points, PESQ scores under different mean delay values are treated as separate data points.

Test results:

As shown in Table 5 and Figure 5 below, the PESQ scores are directly inversely correlated with the packet loss percentage in “good” network delay environment. In other words, lower packet loss results in higher PESQ in “good” network delay environment. However, when the network conditions deteriorate, the PESQ scores trended downward. In addition, the PESQ scores of different packet loss levels except 10% become converged and almost indistinguishable.

Test conclusion:

Evidence from the tests appears to support the hypothesis that packet loss is inversely correlated with PESQ across different network conditions. Test results also indicate that the inverse relationship seems to weaken when network conditions deteriorate significantly as the scores start converging.

Secondly, PESQ measures voice quality, not conversation quality (e.g. crosstalk). Therefore, though it is shown in the test results that network delay has no impact to PESQ, it is necessary to consider conversation quality in defining the network delay requirement.

7. CONCLUSION

3GPP Release 4 offers a different network design and architecture in the circuit core. In a decomposed media gateway environment in R4, voice bearer traffic can be transported over IP between media gateways. In such an environment, the traffic is subject to the conditions of the IP networks.

In this paper, the test results of an experiment conducted by us show that the voice quality of R4 circuit core, when IP transport is used, performs as good as that of R99 circuit core and that mean network delay has no significant impact on the voice quality as measured in PESQ. On the other hand, network delay variation and packet loss are shown to have inverse relationship with PESQ. In other words, a lower PESQ is observed with higher network delay variation and packet loss.

With such information as revealed in the experiment, operators can derive the requirements for the IP bearer transmission networks to meet certain voice quality targets. It should be noted that the impact of network delay variation and packet loss is dependent on the capability of the MGW in the R4 circuit core. Specifically, the jitter buffer and the error concealment techniques employed in a MGW will influence the PESQ under a given level of delay variation and packet loss. It is recommended that the analysis results be used as an initial reference point in defining the IP network requirements and that the requirements be further refined to incorporate the specific implementation in the jitter buffer and the error concealment employed in the media gateways in the R4 circuit core.

REFERENCES

1. "Bearer-independent circuit-switched core network; Stage 2", 3GPP Technical Specification 23.205, Rel-4.
2. "Core network Nb data transport and transport signaling", 3GPP Technical Specification 29.414, Rel-4.
3. "Out of band transcoder control, Stage 2", 3GPP Technical Specification 23.153, Rel-4.
4. "Methods for subjective determination of transmission quality", ITU-T P.800, Aug. 1996.
5. "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs", ITU-T P.862, Feb. 2001.
6. <http://www.gl.com/ITUalgorithms.html>.
7. "Guidance on one-way delay for voice over IP", ITU-T G.114 Appendix II, Sep. 2003.