Voice Quality on the Internet in 2005 as Measured by www.TestYourVoIP.com

Mark Sylor, Nagarjuna Venna, and Harrison Ripps

Brix Networks, 285 Mill Rd, Chelmsford, MA, 01824
{msylor, nvenna, hripps}@brixnet.com

Abstract. How well can the Internet serve as a network carrying voice calls? To answer that question, we have been running a web site called TestYourVoIP where users can run tests of VoIP Quality from their PC. In this paper we describe how TestYourVoIP operates, the kinds of tests it performs, and the results we have found from those tests. Our findings indicate that while many users will get acceptable call quality, too many will not. We offer some insights as to the reasons for poor quality based on these real world experiences. Lastly, we note that over 2005, voice quality on the Internet has gotten worse.

1   Introduction

All communications, including voice and video, are converging on IP networking technology. IMS (IP Multimedia Subsystem) is but the latest trend towards converged services. While private IP networks run by service providers and large enterprises will be important carriers of converged services, the Internet will remain as the core network to which everyone connects.

While the Internet is ubiquitous, it was designed to carry best-effort data services such as email and the web. Many have long feared that the Internet was not suitable for carrying real time services such as voice calls. That fear raises the question at the heart of this work, "How well can the Internet serve as a network carrying voice calls?" This question is critical for VoIP Service Providers who route calls over the Internet as well as for network providers rolling out next generation services over IP such as IPTV, Metro-Ethernet, and VoIP peering services. After all, some fraction of the calls will traverse the Internet even in walled garden networks.

To answer the Internet Voice Quality question, we have been running a public web site at http://www.TestYourVoIP.com where users have been testing Internet Voice Quality since mid-2004. In that time users of TestYourVoIP have successfully run over 500,000 tests of Internet VoIP Quality. The software and hardware used in Test-YourVoIP are based on a Brix Networks product called BrixCare, which is normally used by VoIP service providers in their customer care departments to provide "self help" testing.

This paper describes the results of tests done in 2005, and describes some conclusions we have drawn from those tests.

2   Test Description

TestYourVoIP tests voice quality on the public Internet using hardware and software verifiers located worldwide. Each test run is performed between a software verifier run by a user and one of 7 hardware verifiers located in various cities around the world. Anyone with a web browser and access to the Internet can run a test. Users simply connect their web browser to http://www.TestYourVoIP.com, select a few test parameters and then run the test. The verifiers measure call signaling quality, media quality, and network quality. When the test is completed, the results are displayed to the user, and recorded in a database.

When a user runs a test the user's browser downloads a web page that contains a software verifier implemented as a Java Applet that runs in the user's web browser. The applet is signed by Brix Networks and the user must grant it the privilege to communicate with the hardware verifier.

2.1   Call Signaling Testing

The applet verifier calls the hardware verifier, using SIP as shown in Fig 1. This sequence simulates a call established directly between a calling party and a called party.
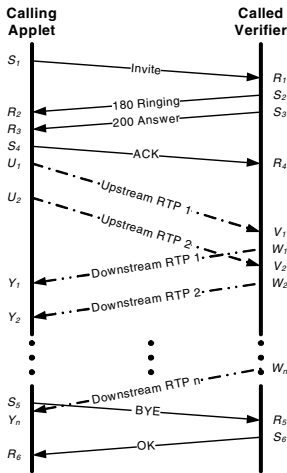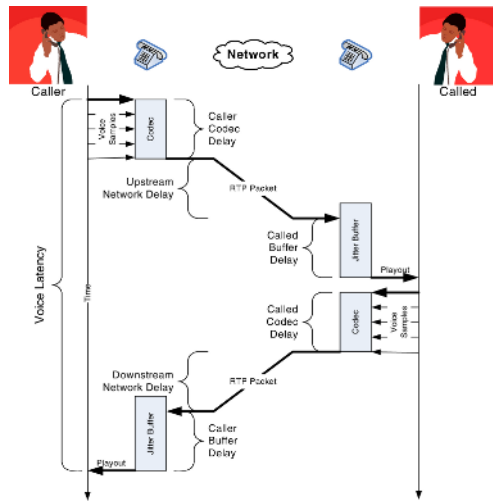

Fig. 1. SIP Call Messages

Fig. 2. Voice Latency

2.2   Voice Media Testing

Once the call has been set up, the verifiers exchange voice media traffic in RTP streams over UDP/IP packets. The voice signal is encoded with either the G.711 or the G.729 codec. The specific parameters used in the codecs are listed in Table 1. No silence is inserted in the signal. The call is maintained for 15 seconds. At the end of that time, each party disconnects the call; generally the applet verifier finishes first.

Table 1. Codec Parameters

G.711	G.729
RTP packet sent every 20 milliseconds	RTP packet sent every 20 milliseconds
160 Samples per RTP packet	2 10ms frames per RTP packet
64 Kbits/sec generated by the codec	8 Kbits/sec generated by the codec
80 Kbits/sec generated by IP	24 Kbits/sec generated by IP
87.2 Kbits/sec over the Ethernet link	31.2 Kbits/sec over the Ethernet link

During the call, we count and measure the quality of the received RTP streams. Numerous measurements are made on the RTP traffic; a few of the more important metrics are listed and described in Table 2. Each measurement is made on both streams. In the results that follow, we call the stream sent from the calling applet verifier to the called hardware verifier the upstream media stream and the stream from the called hardware verifier to the calling applet verifier the downstream media stream.

Table 2. RTP Media Stream Metrics

Metric	Description
CMOS	Conversational Quality Mean Opinion Score, an objective measure of voice quality in a conversation.
LMOS	Listening Quality Mean Opinion Score, an objective measure of voice quality when the user is just listening.
Voice Latency	The round trip time of a voice signal from caller to called party and back.
Lost Packets	The percentage of RTP packets not delivered by the network.
Late Packet Discards	The percentage of RTP packets delivered, but too late for them to be played out.
Out Of Order Packets	The percentage of RTP packets delivered out of order.
Duplicate Packets	The percentage of RTP packets received as duplicates of packets already received.
Average Jitter	A measure of variation in network delay in delivering packets.

Of these metrics, the most important is the objective measurement of the Mean Opinion Score (MOS) for voice quality. MOS takes into account multiple sources of degradation in voice quality. MOS is determined using the methodology described in the G.107 standard.

We also send RTCP packets between the verifiers. RTCP allows us to measure voice round trip latency. The user applet verifier may fail to establish the SIP call with the hardware verifier in some cases. One common cause is the presence of devices on the network that block traffic going out on the standard UDP port used by SIP (5060). If a test fails because of call setup problems, we run it again using an alternate port. In general, media is sent and received on a randomly chosen even

numbered port between 50000 and 51000. The RTCP stream is sent and received on the next odd numbered port. The tests send the media streams as normal, best effort IP traffic. No special QOS marking is done to the packets; neither in the Ethernet LAN header nor in the IP packet header.

After the media test completes, the call is disconnected using SIP. Results are reported back to the TestYourVoIP site by both verifiers, and the test results are displayed to the end user.

2.3 MOS

MOS (Mean Opinion Score) is the most commonly used measure of Voice Quality. What follows is a summary of the techniques used to derive MOS. For a more complete and detailed explanation, see [1,2].

MOS scores are based on subjective tests of voice quality done by representative telephony users in a lab environment. A group of users are asked to listen to a recorded voice sample through the system under test and score the quality of the voice call on a scale from 1 to 5 where 1 is Bad (or Unacceptable), 2 is Poor, 3 Fair, 4 Good, and 5 Excellent. The individual scores are averaged to derive the Mean Opinion Score.

The user's perception of voice quality depends on the task being performed. Simply listening to a voice sample for understandability and legibility is less demanding than attempting to hold a conversation over a network. Thus two types of MOS scores have been developed, Listening MOS and Conversational MOS.

Subjective testing is difficult and expensive to operate. It also is limited to laboratory test environments. So a number of objective techniques for measuring voice quality have been developed over time. They all take measurements of voice calls, either test calls, or observed calls, and compute an estimate of the MOS score that would have been given to that call by users. The technique used in TestYourVoIP is based on the E Model standardized in ITU G.107. The E Model is packet based, in that it measures statistics such as packet loss and delay and uses those metrics to estimate CMOS and LMOS in an objective fashion. The E Model is efficient to measure and compute, thus making it a practical for large scale testing.

The E Model computes a value called the R Factor, a transmission quality rating, based on the assumption that the causes of degradation are additive. The formula is

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

Where R_0 is a base factor that depends on noise and loudness, I_s represents signal impairments, I_d represents impairments that are delayed with respect to speech like echo, I_e represents equipment impairments, and A represents an advantage factor, the users willingness to tolerate poorer voice quality in exchange for convenience. These impairments are derived from: the codec, audio signal loss due to lost packets, audio signal loss due to late packet discards, and voice round trip latency (delay). For details on how the impairment factors are computed, see [1].

The R Factor is then converted into a MOS score. The relationship between R Factor, MOS, and user satisfaction is given in Table 3. In this paper, we use the average of upstream and downstream CMOS scores as the CMOS score for each test.

Too put this value into perspective, a MOS score of 4.4 is generally considered to be toll grade quality. The maximum MOS that can be achieved with the G.711 codec is 4.4. While there is no universal agreement on what constitutes acceptable call quality for VoIP, we count tests that achieve an average CMOS score of 3.6 or better as having acceptable call quality.

Table 3. R Factor to CMOS Conversion

R Factor	CMOS	User Opinion	Users scoring call Good or Excellent
90 – 100	4.3 – 5.0	Very Satisfied	97%
80 – 90	4.0 – 4.3	Satisfied	90%
70 – 80	3.6 – 4.0	Some Users Dissatisfied	70%
60 – 70	3.1 – 3.6	Many Users Dissatisfied	50%
50 – 60	2.6 – 3.1	Almost All Users Dissatisfied	20%
0 – 50	1.0 – 2.6	Not Recommended	0%

3 Related Work

Numerous studies of Internet performance have been made, but we are unaware of any studies similar to the study done at TestYourVoIP. A recent survey of publicly available measurements can be found in [5]. One source of such measurements can be found at CAIDA[6]. Most Internet performance studies have set up a mesh of active testers. Examples include AMP and NLANR. These studies are based on a limited number of testers located in universities and other network hosting sites, and so does not have the scope of TestYourVoIP. Two research efforts, DIMES [8] and NETI@home[7], are attempting to build a larger mesh of testers by asking users to download software that runs in the background following the SETI@home model. DIMES is collecting data on Internet topology through traceroute data. NETI@home is focused more on performance data collection for tests between sites. None of these studies are specifically designed to measure VoIP quality.

A number of publicly available sites on the Internet allow users to run speed tests, often to assess if a user can expect acceptable voice quality. These tests measure available bandwidth upstream and downstream using a variety of techniques. They do not take into account the effects of loss, discards, or voice delay, nor do these sites compute MOS. None of the sites we are aware of have published results.

Keynote Systems conducted one test that does focus on VoIP Quality over the Internet[10]. This test was run between San Francisco and New York over a number of Internet service providers and VoIP providers. The purpose of the test was to compare the quality provided by those ISPs. It did not cover the scale (in either space or time) covered by this study.

The goals of the study reported on in Markopoulous, et.al.[4], are similar to those of our work. However, they are focusing on the Internet Backbone between 5 cities and thus do not include the user access network. Because the user access link is likely to be the main source of loss and delay variation, we believe our study is more reflective of actual user experience. Similarly, the study reported on by Jiang, et.al.

[9] looks at the availability of VoIP on the Internet. It too uses a mesh of active tests between 14 sites, half of which are Internet2 sites and half of which are connected to the Internet via commercial ISPs via Cable Modem or ADSL links. It does not cover voice quality once a call has been made, nor does it cover the scope of this study.

4   Results

In this paper we report on the following questions.

1. What voice quality can users of VoIP over the Internet expect?
2. What are the causes of degradation?
3. What is the affect of codec choice on voice quality?
4. What is the trend in voice quality over time?

The body of data we have collected offers a wealth of information on other questions as well. See the section on Future Work below for some of the other questions we are exploring.

4.1   Filtering Out Bad Network Connections

One reason users run a test on TestYourVoIP is to see whether their Internet connection is good enough to use for VoIP calls. For some tests, the user's Internet connection is not good enough. Tests from bad connections skew the results, and so we have filtered them out of the results we report here. The criterion used to filter the tests was to exclude tests where the round trip voice latency was greater than 1 second or more than 20% of the packets were lost or late packet discards in either direction. Note that in these tests, the MOS score is usually very low (in the Unacceptable range). 9.1% of the tests were filtered out by this criterion.

4.2   Basic Results and ACQ

During 2005, the Average Conversational Objective MOS score across all the included tests run using TestYourVoIP was 3.94. This is a good score, which while not toll quality would satisfy users.

Averages can hide important details. While the users may be happy with the average call, a few bad calls can outweigh the average experience. The distribution of CMOS scores thus gives a better picture of user voice quality experience than the average. Fig 3. shows a cumulative probability distribution of the CMOS scores seen from best (5.0) to worst (1.0). Table 4 below lists the percentage of calls that fall into the quality ranges listed in Table 3.

Most users are satisfied with a call where the CMOS is 3.6 or better. The percentage of test calls that achieve a CMOS score of 3.6 or better is the Acceptable Call Quality Percentage (or $ACQ_{3.6}$). The tests show that during 2005 the tests run on TestYourVoIP achieved an $ACQ_{3.6}$ of 81.1%. While commendable, there remain 18.9% of tests whose CMOS fell below 3.6. Or to put it another way, about 1 out of 5 test calls did not have acceptable call quality. This voice quality is unlikely to satisfy most users of VoIP on the Internet. An $ACQ_{3.6}$ of 95% or 99% would be closer to the quality that users would expect for Telephony over the Internet.

Table 4. CMOS Cummulative Probability Distribution (CPD) by Range

CMOS	Density	CPD
4.3 – 5.0	31.1%	31.1%
4.0 – 4.3	31.9%	63.0%
3.6 – 4.0	18.1%	81.1%
3.1 – 3.6	9.2%	90.3%
2.6 – 3.1	5.1%	95.4%
1.0 – 2.6	4.6%	100.0%
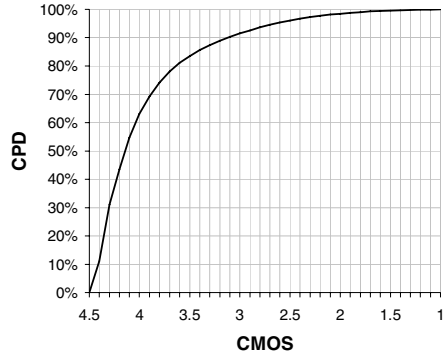
Fig. 3. CMOS Cumulative Distribution

Table 5. Degradation Factors for all calls

Cause	Absolute	Percentage
Codec	0.75	70.0%
Late Discards	0.17	16.3%
Delay	0.09	8.5%
Lost	0.06	5.2%

Table 6. Degradation Factors for poor quality calls

Cause	Absolute	Percentage
Codec	0.77	37.4%
Late Discards	0.49	23.6%
Delay	0.46	22.4%
Lost	0.33	15.8%

4.3   CMOS Degradation Factors

A number of factors affect the computed CMOS score using the E Model. They are: the codec used, lost packets, late packet discards, and the round trip voice latency (delay). The codec degradation is fixed by choice of codec; the others vary with network effects. Because these impairments are additive in the E Model computation of the R Factor, we can compute the relative importance of the degradation causes by the percentage of degradation they cause. By comparing the degradation factors, we can identify the largest contributor to a poor MOS score, the one that should be improved first.

Table 5 shows the CMOS degradation factors that determine the Average CMOS for 2005 as 3.94 both in absolute terms (CMOS points degraded) and as a percentage. The largest factor is the codec, followed by discarded packets, delay and lost packets.

Table 6 shows the degradation factors for tests with low voice quality (CMOS < 3.6). Each factor has roughly the same magnitude, implying there is no single answer to improving those calls with unacceptable voice quality. Each degradation cause must be addressed if we hope to raise the $ACQ_{3.6}$ of the network.

Lost and Discarded Packets. The effect of a lost packet and a packet that arrives too late to be played out is the same, lost audio and degraded call quality. However the cause of lost packets and late packets is quite different. A packet is lost if it is discarded somewhere in the network between the two endpoints. Packets are lost in the network due to congestion or errors. Packets are late because the network delayed delivering the packet until its time for playout has passed.

Table 7 shows the percentages of packets discarded and lost for three cases: all tests, tests with acceptable call quality (CMOS ≥ 3.6) and tests with unacceptable call quality (CMOS < 3.6). We also show the percentage of packets that are lost or discarded, as well as the duplicate and out of order packet percentages. Out of order packets do not in general affect VoIP because the playout buffer reorders the voice samples into their proper order. Similarly, duplicate packets only waste bandwidth. They do not affect VoIP quality directly.

To achieve acceptable call quality, the sum percentage of lost and discarded packets must be kept less than 1%.

Table 7. Lost, Discarded, Duplicate, and Out of Order Packets

	All Tests	**Acceptable CMOS ≥ 3.6**	**Unacceptable CMOS < 3.6**
Late Packet Discards%	1.03%	0.72%	2.37%
Lost Packets%	0.43%	0.12%	1.78%
All Unplayed Packets%	1.46%	0.83%	4.15%
Out of Order Packets%	0.24%	0.17%	0.53%
Duplicate Packets%	0.06%	0.03%	0.17%

Voice Latency. Voice latency affects conversational quality in two ways. First, callers hear the echo of their own voice, and if the delay is too large, the echo becomes noticeable and disruptive. Second, long voice delay interferes with the ability of two people to know when to speak leading to collisions in the conversation.

Voice latency, L, comprises: codec delay, C, network delay, N, and playout delay which is also known as jitter buffer delay, B, in both directions as shown in Fig. 2 and the following formula.

$$L = C_u + N_u + B_u + C_d + N_d + B_d \tag{2}$$

The codec delays C_u and C_d are fixed by the choice of codec. For G.711, it is 20 milliseconds. For G.729 it is 30 milliseconds.

Network delays vary depending on the propagation delays due to distance, transmission delays due to link speed and packet size, forwarding delays in the routers, and queueing delays due to congestion. Network delays are measured using RTCP. RTCP packets are sent periodically during the call in both directions. Each RTCP packet i includes a timestamp T_i when it was sent. The receiver stores T_i and the time it received that packet R_i. When the receiver next sends RTCP packet j, it includes within the RTCP packet T_i, T_j-R_i, and T_j. When the sender receives that packet, it notes the time it received the packet R_j, and computes the network round trip delay using the formula:

$$(N_u + N_d) = (R_j - T_i) - (T_j - R_i) \tag{3}$$

The jitter buffer is a mechanism used to deal with variation in network latency. The delay introduced by jitter buffers B_u and B_d can be fixed if a fixed size jitter buffer is used or can vary if an adaptive jitter buffer is used. TestYourVoIP implements an

adaptive jitter buffer based on [3]. The verifiers measure the average jitter buffer delay and include the measured value in the voice latency they report.

Long voice round trip delay negatively impacts conversational voice quality. Its affect on Average CMOS accounted for 8.5% of the degradation. The average voice round trip latency measured was 263 milliseconds. For calls with low voice quality (CMOS < 3.6), the effect of delay increases to 22% of the total degradation. Average voice round trip latency has increased to 467 ms for the 18% of the test calls with unacceptable voice quality.

Jitter. A packet is discarded when it arrives too late to be played out by the receiver. Packets arrive too late if they are delayed more than the playout buffer can accommodate. At the beginning of each talk spurt, the jitter buffer duration and size is adjusted based on observed values of delay and delay variation for earlier packets. The playout time for each incoming packet is computed based on its transmit time and the current jitter buffer duration. Packets that are delayed in network more than the jitter buffer can accommodate will arrive after their scheduled playout time and so will be discarded as too late. While some jitter buffer implementations have a finite size, for TestYourVoIP, we have set up the jitter buffer to be of unlimited size and so we never have any early packet discards.

Average jitter is a measure of the variation in network delay and is relatively easy to compute. Average jitter is often quoted in Service Level Agreements as a limit on the variation in delay experienced by packets. Average upstream jitter J_U, downstream jitter J_D, and total jitter J for a call are defined as:

$$J_U = \frac{\sum_{i=1}^{n-1}\left|(V_{i+1}-U_{i+1})-(V_i-U_i)\right|}{n-1}, \quad J_D = \frac{\sum_{i=1}^{n-1}\left|(Y_{i+1}-W_{i+1})-(Y_i-W_i)\right|}{n-1}, \quad J = (J_U+J_D)/2 \tag{4}$$

The jitter measured for all test calls, and calls with acceptable and unacceptable call quality are shown in Table 8. The measured average jitter is relatively small, 6.8 milliseconds for all tests. When we compare tests that achieved an acceptable voice quality with tests with unacceptable voice quality, we find the average jitter only increases from 6.1 to 9.9 milliseconds. Note that the average delay increases by roughly the same percentage. The ratio of jitter over delay stays roughly the same.

The data hints that the average jitter is not a very good measure for predicting whether or not a network can support VoIP. Instead we should examine late packet discards since they directly impact voice quality. Late packet discards takes into account the distribution of network delays, not just the average.

Table 8. Jitter Measurements

	All Tests	Acceptable CMOS ≥ 3.6	Unacceptable CMOS < 3.6
Upstream Jitter J_U	7.3 msec	6.4 msec	11.2 msec
Downstream Jitter J_D	6.2 msec	5.7 msec	8.6 msec
Total Jitter J	6.8 msec	6.1 msec	9.9 msec
Voice Delay	263 msec	216 msec	467 msec
Jitter/Delay Ratio	0.026	0.028	0.021
Late Packet Discards	1.03%	0.72%	2.37%

4.4   Codec Comparison

TestYourVoIP users can select the codec used in their test calls. While most choose the default G.711 codec, slightly more than 35,000 tests were run using the low bandwidth G.729 codec. Table 9 compares the results obtained for the two codecs. Because the G.729 codec degrades the voice quality more than the G.711 codec, it is not surprising that the average CMOS of the G.729 codec is less than that of G.711. However, one would hope that the G.729 codec would perform better under more challenging network conditions due to its decreased bandwidth. Fig. 4 shows the distribution of CMOS values for both codecs. The G.711 codec is more likely to outperform the G.729 codec through most of the range. Only when the CMOS score drops to below 1.4 is the G.729 codec more likely to have a better CMOS than that score. This only accounts for 0.3% of all tests.

Table 9. Codec Comparison

G.711	All Tests	Acceptable CMOS ≥ 3.6	Unacceptable CMOS < 3.6
Average CMOS	3.97	4.19	2.93
Voice Delay	256 msec	214 msec	457 msec
Lost Packets	0.43%	0.12%	1.92%
Late Packet Discards	1.01%	0.72%	2.38%
ACQ$_{3.6}$	82.6%		
Tests with bad network	8.2%		
G.729			
Average CMOS	3.63	3.97	2.94
Voice Delay	327 msec	235 msec	510 msec
Lost Packets	0.43%	0.09%	1.10%
Late Packet Discards	1.19%	0.63%	2.31%
ACQ$_{3.6}$	66.7%		
Tests with bad network	16.8%		

   These results suggest that the conditions where the G.729 codec outperforms the G.711 codec are rare. Given all the work done to develop low bandwidth codecs, we are surprised at this result.

4.5   Trends

Call quality over 2005 has generally been trending downwards as shown in Fig 5. When we fit linear least squares trends to the data, we find CMOS is decreasing by 0.007 points per month and ACQ is decreasing by 0.4% per month with R^2 values of 0.66 and 0.70 respectively. We find the trend in Acceptable Call Quality to be particularly distressing. If this trend continues, each year 5% more of the VoIP calls on the Internet will experience unacceptable call quality.
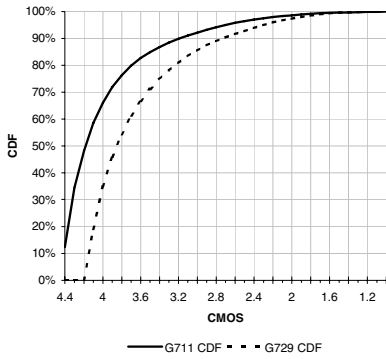
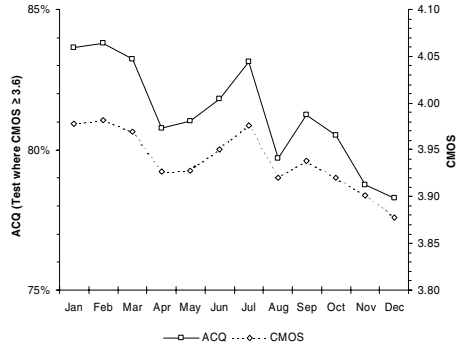**Fig. 4.** Comparing CMOS for G.711 and G.729

Fig. 5. Call Quality Trends in 2005

5   Future Work

TestYourVoIP has collected a rich dataset describing VoIP quality on the Internet. In this paper we have only scratched the surface of the lessons this data can teach us. Some other questions we can explore are:

- Does voice quality vary by time of day or day of the week?
- Is voice quality symmetric?
- Is voice quality the same worldwide, or are there geographic differences?
- How does voice quality vary by distance, hops, and route?
- How do different jitter buffer strategies perform?
- What the impact of the components of voice round trip latency?
- What is the cause of the differences seen in voice quality between codecs?
- How does average jitter statistically relate to late packets and voice quality?

6   Conclusions

While average CMOS on the Internet in 2005 was 3.94, only 81% of the test calls achieved an acceptable call quality of 3.6 ($ACQ_{3.6}$). Far too many calls had unacceptable call quality.

Of the three causes of degraded voice quality that network providers can affect; late packet discards, lost packets, and voice latency, the most important is late packet discards. However, if we focus on only those calls with unacceptable call quality, the pattern changes, and the three causes have roughly equal weights. Providers cannot focus on improving just one cause of poor voice quality to improve $ACQ_{3.6}$, rather, they must work to improve all three causes.

The data hints at two surprising results.

1. Average Jitter, a metric often used in service level agreements does not seem to be a good predictor of voice quality. A better choice for service level metrics is Late Packet Discards, a metric that directly affects voice quality.

2. When we compare the voice quality that codecs actually achieve on the Internet, we find that the G.711 codec is more likely to produce good voice quality than the low bandwidth G.729 codec. While 82.6% of the tests with the G.711 codec had acceptable call quality, only 66.7% of the tests with the G.729 codec had acceptable call quality.

Finally, the data clearly shows a slow but clear decrease in voice quality during 2005. Voice Quality on the Internet is getting worse.

References

1. The E Model: A computational model for use in transmission quality, ITU-T Recommendation G.107, Dec 1998.
2. Telchemy, Inc. Voice Quality Measurement, Jan, 2005,
   http://www.telchemy.com/appnotes/TelchemyVoiceQualityMeasurement.pdf.
3. Ramjee. R., Kurose, J., Towsley, D., and Schulzrinne, H., "Adaptive playout mechanisms for packetized audio applications in wide-area networks", in Proceedings of the Conference on Computer Communications (IEEE Infocom), Toronto, Canada, June 1994, pp. 680-688.
4. Markopoulou, A., Tobagi, F., and Karam, M., "Assessing the Quality of Voice Communications Over Internet Backbones", IEE/ACM Trans. Networking, v. 11, no. 5, Oct 2003.
5. Claffy, K.C., Crovella, M., Friedman, T., Shannon, C. and Sring, N., "Community-Oriented Network Measurement Infrastructure (CONMI) Workshop Report", ACM SSIGCOMM Comp. Comm. Review, V. 36, No. 2, Apr 2006.
6. CAIDA, the Cooperative Association for Internet Data Analysis, http://www.caida.org/
7. Simpson, Jr., C. and Riley G., "NETI@home: A Distributed Approach to Collecting End-to-End Network Performance Measurements", PAM2004 - A workshop on Passive and Active Measurements, Apr 2004.
8. Shavitt, Y. and Shir, E., "DIMES: A Distributed Architecture for Internet Measurement and Monitoring", IEEE Infocom 2005.
9. Jiang, W. and Schulzrinne, H. "Assessment of VoIP Service Availability in the Current Internet", PAM 2003 A workshop on Passive and Active Measurements.
10. "Keynote's Hard Numbers On VoIP Quality", VOIP Magazine, July 2005,
    http://www.voip-magazine.com/content/view/275/