Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality

A. D. Clark, Ph.D. Fellow IEE, Member IEEE[?]

This paper describes VQmon, a non-intrusive monitoring technique for Voice over IP networks that is computationally efficient and suitable for integrating or embedding into VoIP gateways or IP Phones. This uses an extended version of the ITU G.107 E-Model incorporating the effects of time varying packet loss and "recency". A 4 state Markov model is used to represent the time distribution of packet loss during a VoIP call.

QoS, Voice over Packet, E model, subjective quality

A. INTRODUCTION

Voice over IP networks differ from conventional telephone networks in that voice quality is affected by a wider variety of network impairments and can vary from call to call and even during a call. It is therefore desirable to monitor call quality in order that service providers can properly provision networks and that network resources are properly allocated.

Passive monitoring systems examine operating characteristics of a system in order to assess or measure performance level. This may involve examining elements of the system, for example buffer levels, or examining the data stream being transmitted through the system. This contrasts with Active measurement systems in which test data is inserted into the system and used to obtain performance measurements.

This paper describes a passive monitoring system (VQmon) for Voice over IP networks that is able to monitor per-call quality, providing feedback to a service management or CDR (Call Detail Record) system. The VQmon monitoring system also considers the effects of time varying impairments such as bursty packet loss and recency.

B. EMBEDDED PASSIVE MONITORING

Passive monitoring systems examine operating characteristics of a system in order to assess or measure performance level. This may involve examining elements of the system, for example buffer levels, or examining the data stream being transmitted through the system. This contrasts with Active measurement systems in which test data is inserted into the system and used to obtain performance measurements.

Embedded passive monitoring systems employ some form of monitoring function embedded into the equipment that comprises the system under test. This has the advantage of a closer relationship with system elements, allowing access to real time data and control information however has the disadvantage that implementation cost and complexity must be low. This contrasts with external passive monitoring systems which may, for example, be connected to T1 trunks or Ethernet LANs.

Within the context of a VoIP network embedded passive monitoring can be integrated into VoIP Gateways, IP Phones or other end-systems, providing access on a per-call basis to CODEC selection, packet loss and delay information. This permits per-call estimates of transmission quality to be made with minimal impact on the service being monitored.

Active monitoring systems typically make test calls through the VoIP network, transmit speech files and compare transmitted and received files using PSQM, PESQ or some similar method. This approach allows the CODEC performance to be directly measured however provides only a snapshot of network performance on a single connection.

C. USING THE E MODEL TO ESTIMATE VOICE QUALITY

The E Model [8] is a well established transmission quality model for telephone networks. This provides an objective method of assessing the mouth-to-ear transmission quality of a telephone connection and is intended to assist telecom service providers with network planning and performance monitoring. The E Model is described in some detail in ETSI Technical Report ETR 250 [9] and in ITU Recommendations G.107 [10] and G.108 [11].

The E Model has the following components:-

$$R = Ro - Is - Id - Ie + A$$

Which results in an R factor of between 0 and 100. The components of R are:-

Ro - representing the effects of noise and loudness ratio

[?] Telchemy Incorporated, 3360 Martins Farm Road, Suite 200, Suwanee, GA 30024, alan@telchemy.com

Is - representing the effects of impairments occurring simultaneously with the speech signal

Id - representing the effects of impairments that are delayed with respect to the speech signal

Ie - representing the effects of "equipment" such as DCME or Voice over IP networks

A - the advantage factor, used to compensate for the allowance users make for poor quality when given some additional convenience (e.g. cellphone)

The equipment impairment factor Ie is generally used to represent the effects of Voice over IP equipment. Certain CODECs have been characterized through subjective testing to give a profile of the variation of Ie with packet loss [11].

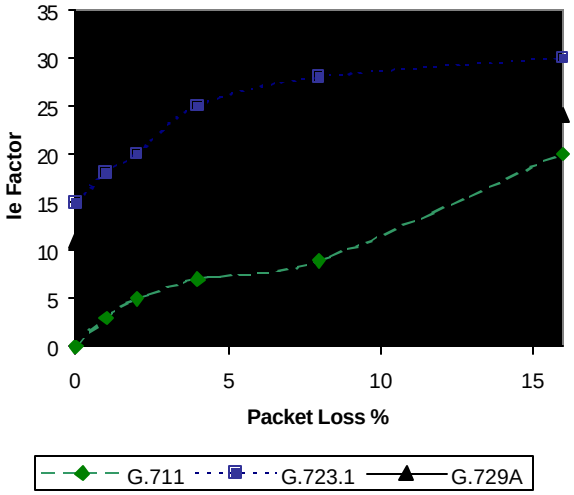

Figure 1. Mapping Packet Loss to Ie

D. TIME VARYING IMPAIRMENTS

The network impairment that has greatest effect on voice quality is packet loss. Packet loss may occur due to buffer overflow within the network, deliberate discard as a result of some congestion control scheme (e.g. Random Early Detection) or transmission errors. Several of the mechanisms that can lead to packet loss are of a transient nature and hence the resulting packet loss is bursty in nature. Bolot [6] studied the distribution of packet loss in the Internet and concluded that this could be represented by a Markovian loss model such as the Gilbert or Elliott models.

Jitter (or packet delay variation) also has an effect however the use of a jitter buffer generally replaces jitter by delay and packet loss. Incoming packets are buffered and then read out at a constant rate; if packets are excessively late in arriving then they are discarded. For this reason it is advisable to measure packet loss (or rather frame loss) between the jitter buffer and the CODEC. Jitter buffers are often adaptive and adjust their depth dynamically based on either the current packet discard rate or current jitter level.

Cox and Perkins [3] compared the impact of random and burst packet loss on G.711 and G.729A CODECs. They found that for low packet loss rates a burst distribution gave a higher subjective quality than a non-bursty distribution whereas for high packet loss rates the converse was true. One explanation for this effect is that at low packet loss rates the distortion due to the loss of two successive packets is not much greater than that of one lost packet and is counteracted by the greater distance between packet loss events.

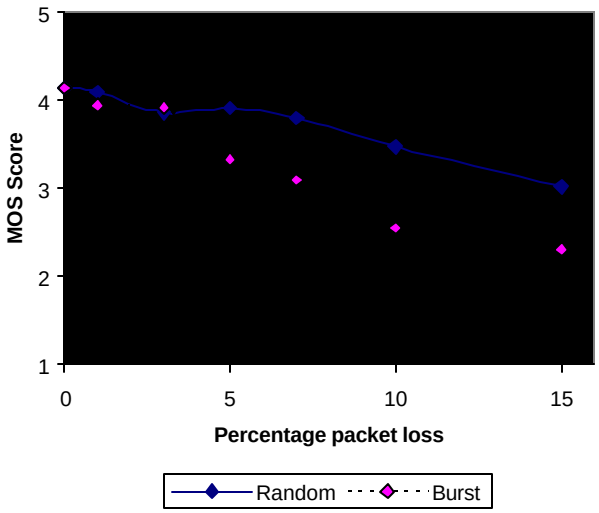

Figure 2. Effects of Random vs Burst Packet Loss

If the rate of packet loss varies during a VoIP call then the perceived call quality will also vary. The term "instantaneous quality" may be used to denote the measured or calculated quality due to packet loss or other impairments and the term "perceived quality" may be used to denote the quality that the user would report at some instant in time.

Intuitively, if instantaneous quality changes from "good" to "bad" at some moment in time then the listener would not immediately notice the change. As

time progresses the user would become progressively more annoyed or distracted by the impairment. This leads to the idea that the perceived quality changes more slowly than instantaneous quality.

In tests reported by Barriac et al [1] the packet loss rate during a 3 minute call was varied from 0 to 25%. In the example shown below the packet loss was set to 25% for most of the call and reduced to 0% for a 30 second period mid-call. Listeners were asked to move a slider to indicate their assessment of quality during the call and then asked to rate the call at the end. This showed the effect described above, with an approximately exponential curve with a time constant of 5 seconds for the good-to-bad transition and 15 seconds for the bad-to-good transition.

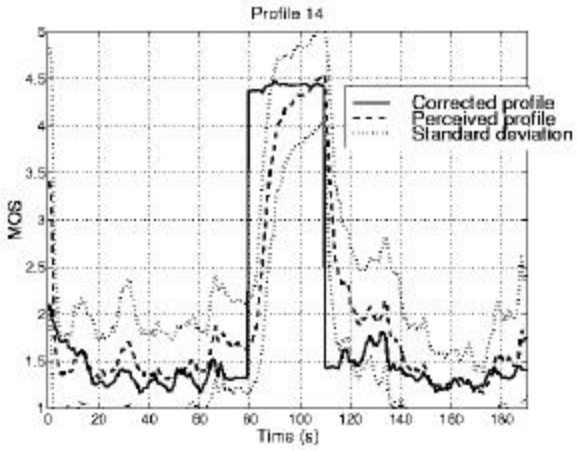

Figure 3. Relationship between Instantaneous and Perceived Quality Metrics (Source Barriac [1])

The "recency" effect reflects the way that a listener would remember call quality.

In tests conducted by AT&T [2] a 15 second burst of noise was moved from the beginning to the end of a 60 second call. When the noise was at the start of the call users reported a MOS score of 3.82 whereas when the noise was at the end of the call users reported a MOS score of 3.18, giving a change in MOS score of 0.64.

Tests reported by France Telecom [1] showed a similar effect. An improvement in MOS score of 0.68 was reported when a period of high packet loss was moved from the end to the beginning of a 60 second call.

The effect is believed to be due to the tendency for people to remember the most recent events [4] or possibly due to auditory memory which typically decays over a 30 second interval [1].

E. Extending the E Model to reflect time varying impairments

In many Voice over IP network implementations the connection between CODEC and telephone handset may be transient. For example a user may be dialing through an existing local loop and being routed to a Gateway located at the Central Office. This means that some elements of the E Model may not be measurable by equipment located within the network. Default values for many of the E Model parameters can be assumed (per G.107), giving an effective value for Ro – Is of 94.

The E Model can then be represented as:

$$R = 94 - Id - Ie$$

The average value of Ie may be determined by taking the average of the perceived quality for the call.

For each time interval t(i), the instantaneous quality $I_{inst}(i)$ is determined by measuring the post-jitter buffer packet loss for the time interval and mapping the packet loss to an Ie value using the curves shown in Figure 1.

The perceived quality can be estimated from the instantaneous quality by assuming an exponential decay, modeling the effect described in Section D. A time constant of 5 seconds is assumed for a deterioration in quality and 15 seconds for an improvement in quality.

Over a series of N samples the average perceived quality is therefore

$$I_e = \text{sum}(I_{perceived}(i))/N$$

The recency effect can be modeled by assuming that perceived quality decays exponentially with time constant t3 from the "exit" value I_{exit} from a burst of noise or distortion towards the average Ie. The following model is proposed:

$$I_e \text{ (end of call)} = Ie + (k(I_{exit} - I_e)) e^{-y/t3}$$

F. Modeling packet loss

In order to meet the requirements of real time implementation within a VoIP Gateway it is essential to minimize processing overhead. The approach used in VQmon is to obtain some minimal amount of

information during a call and perform most computation at the call end.

A 4-state Markov model is used to represent the burst packet loss characteristics of the call. The four states represent the conditions of receiving or losing a packet within burst or gap conditions.

State 1 – Gap state- receive packet
State 2 – Burst state – receive packet
State 3 – Burst state – lose packet
State 4 – Gap state – receive packet

A gap state is defined by the requirement that g_{min} successive packets must be received.

This model is similar to the more normal Gilbert or Elliott models however includes a state representing the loss of an isolated packet within a gap. The rationale for this is that packet loss concealment (e.g. replay last packet), can mask the effects of isolated lost packets.

A packet loss event driven model is used to count a minimum number of key transition events. It is assumed that Voice Activity Detection is being used and hence that packet loss reports relate to packets containing speech energy. When the call is completed then remaining transition counts can be derived and then the counts normalized to give probabilities. This model holds considerable information and can be used to determine average gap and burst size and density, successive lost packet distribution etc.

Packet loss event:-

$c_5 = c_5 + pkt$
if pkt >= g_{min} then
    if lost = 1 then
        $c_{14} = c_{14} + 1$
    else
        $c_{13} = c_{13} + 1$
    lost = 1
    $c_{11} = c_{11} + pkt$
else
    lost = lost + 1
    if lost > 8 then $c_5 = 0$
    if pkt = 0 then
        $c_{33} = c_{33} + 1$
    else
        $c_{23} = c_{23} + 1$
        $c_{22} = c_{22} + pkt$
pkt = 0

pkt is an input parameter representing the numbr of packets received since the last lost packet event.The series of counters c_{11} to c_{14} are used to determine the corresponding Markov model transition probabilities (i.e c_{11} is used to calculate p_{11}). Counter c_5 is used to measure the delay since the last "significant" burst of lost packets. Parameter g_{min}, the minimum gap size, is typically 16.

The equipment impairment value for the burst and gap condition is determined using the curves shown in Figure 1, giving I_{eb} and I_{eg} respectively.

Let I_1 be the quality level at the change from burst condition I_{eb} to gap condition I_{eg} and let I_2 be the quality level at the change from I_{eg} to I_{eb}

$$I_1 = I_{eb} - (I_{eg} - I_2)\, e^{-b/t1} \quad \text{where } t_1 \text{ is typically 5}$$

$$I_2 = I_{eg} + (I_1 - I_{eg})\, e^{-g/t2} \quad \text{where t2 is typically 15}$$

Combining these gives

$$I_2 = (I_{eg} (1 - e^{-g/t2}) + I_{eb} (1 - e^{-b/t1}) e^{-g/t2}) / (1 - e^{-b/t1-g/t2})$$

Integrating the expressions for I_1 and I_2 to give a time average gives

$$I_e(av) = (b\, I_{eb} + g\, I_{eg} - t_1 (I_{eb} - I_2) (1 - e^{-b/t1}) + t_2 (I_1 - I_{eg}) (1 - e^{-g/t2})) / (b + g)$$

This may be used to determine an R factor from the expression:-

$$R = 94 - I_e(av)$$

This R factor does not yet include the effects of delay or recency however is useful when examining the effects of packet loss, jitter and CODEC type on transmission quality. Within the context of VQmon this is the *Network R Factor*.

The effects of delay are well known [5] and easily modeled. Delays of less than 175mS have a small effect on conversational difficulty whereas delays over 175mS have a larger effect. A simple delay model is used in VQmon:

If delay < 175 mS then
    $I_d = 4$. delay
Else
    $I_d = 4 + (\text{delay} - 175) / 9$

The Recency effect is modeled using the approach described in E above. It is assumed that the I_1 represents the exit value from the last significant burst of packet

loss, y represents the time delay since the last burst, t3 is a time constant of typically 30-60 seconds and k is a constant (set to a nominal value of 0.7).

$$I_e \text{ (end of call)} = I_e(av) + (k(I_1 - I_e(av))) e^{-y/t3}$$

The *User R Factor* is determined from the expression below. This is intended to more closely approximate the user's perspective of quality and therefore does take into account both recency and delay.

User R Factor = 94 - I_e (end of call) – I_d.

G. EXPERIMENTAL RESULTS

Some initial subjective comparison was made to validate the VQmon model. An audio file was corrupted using a burst error process which comprised a low loss state and a high loss state, the loss and state transition probabilities being selected randomly. A 10mS packet size was used and packet loss concealment applied.

Sets of five test files were created, and a group of six listeners used to rank the files from 1(best) to 5 (worst). The ranking was compared with that predicted by the algorithm described above.

File	Mean user rank	R factor rank
Vgnf	1.0	1
dgcs	3.0	2
cnkb	3.5	3
mxhr	2.5	4
gwav	5.0	5

Table 1 – Comparison of R factor and User Ranking – data set 1

File	Mean user rank	R factor rank
Ecen	1.2	1
Fhpc	1.8	2
Rlgd	3.2	3
Xknc	3.8	4
Dlwx	5.0	5

Table 2 – Comparison of R factor and User Ranking – data set 2

File	Mean user rank	R factor rank
Mvui	1.8	1
Fyok	1.2	2
Rkdi	3.5	3
Mtwt	3.5	4
okdu	5.0	5

Table 3 – Comparison of R factor and User Ranking – data set 3

The results showed reasonable correlation with user ranking however there were several obvious exceptions, for example file **mxhr** from data set 1. The locations of packet loss events for this file were reviewed and it became apparent that some loss bursts occurred either during silence periods or during periods when the sound produced by the speaker was not changing significantly, for example during an extended "aaaah".

Further comparisons are being made using both ranking tests of the type described above and comparisons with well known objective test measures such as PSQM and PESQ.

H. SUMMARY AND CONCLUSIONS

VQmon represents a novel approach to embedded passive monitoring that incorporates the effects of burst packet loss and recency. The algorithm provides a computationally efficient method for estimating the transmission quality of a Voice over IP network, and produces results that correlate well with user ranking of impaired files.

One problem that arises as a result of analyzing only packet related statistics is that it is not possible to identify the exact effects of lost packets during talkspurts. Additional work is in process to incorporate "smart" packet loss events which would be generated by the CODEC.

I. REFERENCES

[1] France Telecom Study of the relationship between instantaneous and overall subjective speech quality for time-varying quality speech sequences: influence of a recency effect. ITU Study Group 12 Contribution D.139, May 2000

[2] Rosenbluth, J. H., Testing the Quality of Connections having Time Varying Impairments. Committee contribution T1A1.7/98-031

[3] Cox, R., Perkins, R., Results of a Subjective Listening Test for G.711 with Frame Erasure Concealment, Committee contribution T1A1.7/99-016, May 1999

[4] Baddeley, Human Memory

[5] Britt, R., Armstrong, M., Voice Quality Recommendations for IP Telephony. Committee contribution TR41.1.2/00-05-004, May 2000

[6] Bolot, J., Fosse-Parisis, S., Towsley, D., Adaptive FEC based Error Control for Interactive Audio in the Internet. Infocom 99

[7] ETSI Speech Communication Quality for Mouth to Ear for 3.1 kHz Handset Telephony across Networks. Technical Report ETR250, 1996

[8] ITU Recommendation G.107

[9] ITU Recommendation G.108