STUDIES TOWARD IMPROVED VoIP SERVICES FOR FUTURE COMBAT SYSTEMS

Robert G. Cole and Subramaniam Kandaswamy
JHU Applied Physics Laboratory
Laurel, MD *

Alan Clark
Telchemy, Inc.
Suwanee, GA †

ABSTRACT

In this paper, we propose adapting a well proved approach from the telephony world to characterize Voice over IP (VoIP) performance and develop improved capabilities for tactical VoIP deployments. The approach relies upon the development of an equivalent E-Model [14] specific to military, tactical deployments of VoIP services. We define the approach necessary to develop such a model, and describe it's potential benefits to the DoD.

1. INTRODUCTION

The Department of Defense (DoD) is moving to an information centric style of warfare in order to improve its war fighting capability. To support this reliance on improved information dissemination, the DoD is designing and deploying the Global Information Grid (GIG). The GIG is built upon a single, ubiquitous IP network supporting all DoD applications. As part of this migration to a common IP network, the military plans to deploy Voice over IP (VoIP) services. This deployment will include VoIP services over tactical, battlefield networks such as the Joint Tactical Radio System (JTRS). As wireless Mobile Ad Hoc Networks (MANETS) are expected to have rather complex channel characteristics, i.e., a broad range of delay and loss statistics, it is not fully known how effective VoIP services will perform over these tactical, battlefield networks. Therefore, a set of studies aimed at a quantitative characterization of these environments and evaluation of the quality of their VoIP services are needed. Specifically, we propose adapting a well proved approach from the telephony world in characterizing VoIP performance in order to develop improved capabilities for tactical VoIP deployments.

Extensive use of subjective quality-of-voice studies has been made in the design and deployment of commercial services over the past fifty years. Typically, subjective evaluation of voice quality has been performed within the context of Mean Opinion Score (MOS) measurements in the commercial world and Rhyme Tests in the military world [2]. Based upon extensive MOS tests on the Conversational Quality of commercial Compression/Decompression (CODEC) devices under a range of conditions, the ITU has standardized an analytical model for predicting the performance of transport connections supporting voice services. This model is referred to as the E-Model [14] [15] [18]. The E-Model provides an analytical method to relate the underlying parameters describing the voice path to the expected voice quality on the reference connection. The E-Model has proved to be an extremely useful tool for engineers in assessing facility design options for voice transport. More recently, the E-Model has been used in assessing the quality of VoIP services [9] [11], in enhancing VoIP monitoring systems and tools [9] [7] [10], and in designing new adaptive CODEC schemes in VoIP networks [5]. An extremely interesting prediction of the E-Model is that it may be possible to simultaneously improve voice quality while decreasing the bandwidth for the VoIP flow, under conditions of packet dropping in the underlying IP transport. Thus the E-Model can be used to assess the Adaptive Multi-Rate (ARM) CODEC designs proliferating within cellular network systems [17]. This amazing capability would be extremely useful in designing adaptive CODECs for deployment in low bandwidth, tactical networking environments.

*R. Cole and S. Kandaswamy are with the Power Projection Systems Department, JHU/Applied Physics Laboratory, 12000 Johns Hopkins Road., Laurel, Maryland, 20723. Phone: (443) 778-{6951,1122}, E-mail: {robert.cole, kanda.kandaswamy}@jhuapl.edu

†A. Clark is with Telchemy, Inc., One Focus Plaza, Suite 200, 3360 Martins Farm Road, Atlanta, Georgia, 30024. Phone: (770) 614-6944, E-mail: alan.d.clark@telchemy.com

We propose applying these commercial methods to develop an equivalent E-Model for military CODECs under tactical, battlefield environments. With such a Military E-Model in hand, we would expect the DoD to reap the same benefits that the E-Model has produced in commercial applications, as identified in the previous paragraph. This would require five phases of investigation: 1) select a the set of tactical environments and systems for the study, 2) perform channel characterization of packet performance in tactical, battlefield MANETS, 3) perform the appropriate subjective studies of military CODECs under the derived channel characteristics, 4) perform data analysis for the Military E-Model development for the tactical environments and CODEC systems, and 5) design algorithms for dynamically switching between optimal VoIP CODEC settings to maximize voice quality.

In the remainder of this paper, we expand upon the ideas discussed within the *Introduction*. There is some technical background information on the E-Model that is useful in understanding this proposal, which we present in the next section. We then follow this with a more detailed presentation of the specific steps toward the development of a Military E-Model.

2. BACKGROUND ON THE E-MODEL

In the commercial world, the evaluation of CODEC performance has generally been based upon Mean Opinion Score (MOS) testing. These are subjective tests where test subjects, i.e., humans, are asked to rate quality of speech after it has been passed through the encoding and decoding methods specific to the CODEC. Typically, subjects are asked to rate the speech quality on a 0 to 5 scale and the results are averaged, resulting in the MOS. CODEC performance is evaluated in the presence of impairments, e.g., packet delay, delay jitter and loss in a VoIP transport. Depending upon the specific question asked of the subjects, i.e., 'What is the quality of the conversation?' versus 'Can you understand what was spoken?', the MOS reflects either the Conversational Quality or the Intelligibility Score for the CODEC. For the commercial world, Conversational Quality is often the question to be assessed in MOS tests. For military applications, in addition to Conversation Quality, an Intelligibility Score is relevant and is typically assessed in the context of Rhyme Tests [2].

In the commercial world, these collective MOS studies are culminating in the development of an analytic engineering tool called the *E-Model* [14] [15] [18]. The development of this tool is the result of extensive MOS testing under conditions experienced in commercial telephony services, extensive characterization of transport services in order to quantify these test conditions, and data analysis resulting in analytical characterization of MOS results.

The E-Model generates an R-factor which ranges from 0 to 100; 100 being a perfect rating of the overall performance of the voice transport connection. High quality is typically equated with a score of 90 or higher, while poor quality is associated with a score of 60 or lower. There exists an analytic, one-to-one relationship between the R-factor and the MOS. In a simplified form [14] [15] [9], the E-Model can be expressed as

$$\begin{aligned} R \;\approx\; & \alpha - \beta_1 d - \beta_2 (d-\beta_3) H(d-\beta_3) \\ & - I_e + (95 - I_e)\frac{e}{e+B_{pl}} \end{aligned} \tag{1}$$

where α is the highest possible R-factor rating given a particular transport reference connection, the βs are related to the impact of ear-to-mouth delay on conversational quality, and I_e and B_{pl} are related to the CODEC performance. Here d is the ear-to-mouth delay of the reference connection and e is the IP packet loss ratio for the reference connection. β_1 reflects the impact of delay on a conversation when the $d < \beta_3$ and its small value implies that delay has little impact at this range. However, above $d > \beta_3$, the interactive conversation becomes more and more difficult and will eventually degenerate into a simplex conversation where the participants explicitly hand over control, through, e.g., the word "over", once they complete their current statement. The rate at which this awkwardness in the conversation increases with delay is captured by β_2. I_e reflects the information loss of the CODEC compression scheme and the performance of the associated Voice Activity Detection (VAD) algorithms. The lossier the compression scheme, the higher the value of I_e. The Packet-loss Robustness Factor, B_{pl} reflects the resilience of the CODEC's Loss Concealment and Encoding algorithms in masking packet losses from the ear of the listener.

Eq.(1) gives the relationship between the ear-to-mouth delay and loss ratio to the expected conversational quality. However, one typically associates

Table 1: E-Model parameters for a few CODECs.

α	β_1	β_2	β_3
94.2	0.07	22	177
CODEC	I_e	B_{pl}	Notes
G.711	0	25.1	w/ LC*
G.728	7	-	B_{pl} not listed
G.729a	11	19	w/ VAD**

* Loss Concealment (LC)
** Voice Activity Detection (VAD)

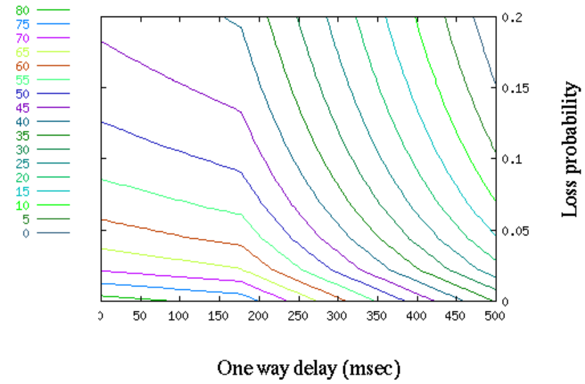

Figure 1: The contour plot of constant R-factor values for a G.729a CODEC versus delay and loss.

VoIP call quality with network delay, loss and delay jitter. The impact of delay jitter is reflected through the packet loss ratio based upon the performance of the CODEC's dejitter buffer. When the buffer is too small, there are times when the CODEC is expecting the next packet's encoded bits but they are not available due to variation in the packet delays across the transport network. In this case, even though the transport network may eventually deliver the packet, the CODEC will consider this as a packet loss event and count this loss toward the ear-to-mouth loss ratio.

Table 1 shows some representative values for these parameters for a few commercial CODECs, including the G.711 [13] [1], G.728 [13], and G.729a [12]. Figure 1 shows the R-factor predictions of voice quality from a typical G.729a VoIP Gateway. Here the connection's performance is defined in terms of the mean delay and packet loss ratio. This plot is a contour plot showing the constant R-factor curves versus the delay and loss metrics. For example, the curve intersecting the left hand axis at the upper left hand corner is the R = 45 contour curve. Following this curve down to the right, it eventually intersects the lower axis at a delay value of 420 msec. Curves to the lower left of this curve represent higher R-factor values and hence higher quality VoIP conditions, while curves to the upper right represent lower R-factor values and hence lower quality VoIP conditions.

The E-model continues to be enhanced and refined based upon further MOS testing. One blaring issue with the current E-Model is its characterization of the channel's packet loss in terms of the mean loss ratio. It is well known that packet losses typically occur in bursts [3] [4]. Further, it is believed, but based upon little empirical evidence, that CODEC performance can be greatly impacted by the nature of the channel packet loss statistics. One impact of burst losses on CODEC performance may be due to the effectiveness of the packet loss concealment algorithms built into the CODEC. When a packet is late in arriving to the decoder, the loss concealment algorithm attempts to mask the loss, e.g., by repeating the last packet, etc. For packet losses which are spaced out into, primarily, single loss events, the loss concealment algorithms tend to be very effective in hiding the packet losses from the listener. However, the effectiveness of the loss concealment algorithms degrades in the presence of multiple, consecutive losses typical of a bursty loss channel [6]. This explains the belief that bursty losses will negatively affect perceived speech quality relative to a smoother packet loss channel.

Having an accurate Military E-Model describing the performance of military, tactical battlefield CODECS would be an invaluable tool in the design, deployment and maintenance of high quality VoIP services. In the commercial world, the E-Model is already finding many useful applications beyond its original role as a facility planning and design tool. In the network monitoring realm, it has been proposed to use the E-Model for continuous monitoring of voice services, see, e.g., [9]. Companies offer products and tools which provide this type of monitoring functions, see, e.g., [7]. The IETF has begun to embed this monitoring capability into their protocol standards, e.g., [10] [8]. The E-Model has recently been used as an objective function for the design of adaptive de-jitter buffers for VoIP transport [5]. Previous adaptive de-jitter buffer schemes have focused on objective functions which attempt to minimize end-to-end delay while maintaining reasonable packet loss. However, it seems more natural to use the E-Model as the objective function in the design of new adaptive CODEC schemes.

As previously mentioned in this context, the current E-Model for commercial CODECs predicts that under certain channel conditions, e.g., high packet loss, dynamically switching to a lower bit-rate CODEC with forward error correction will simultaneously improve the perceived speech performance compared to the higher bit-rate CODEC while reducing overall network bandwidth consumption. This somewhat surprising result gives a quantitative evaluation of the benefits of Adaptive MultiRate (ARM) scheme prevalent in cellular networks. This effect has also been qualitatively discussed in the context of video encoding studies. An example of this effect is shown in Figure 2. Here the voice quality of two CODEC designs are compared by taking the difference in their predicted R-factor values. One CODEC scheme is the standard G.711 CODEC with loss concealment. The other scheme is a G.729a CODEC with loss concealment and a two fold Forward Error Correction (FEC) redundancy. The G.711 CODEC generates 64Kbps data rate while the G.729a CODEC with redundancy generates a data rate of 16Kbps. The E-model predicts that the optimal CODEC scheme between these two, is the G.711 CODEC for packet loss ratios of 5% or less, while the G.729a CODEC with error correction is favored for higher packet loss ratios. It is currently unknown what the impact that burst errors will have on this prediction due to the lack of an existing E-Model under these channel conditions. Further, the current E-Model attempts to predict speech quality results, while the DOD for military applications may be more interested in speech intelligibility based upon Rhyme Tests [2].

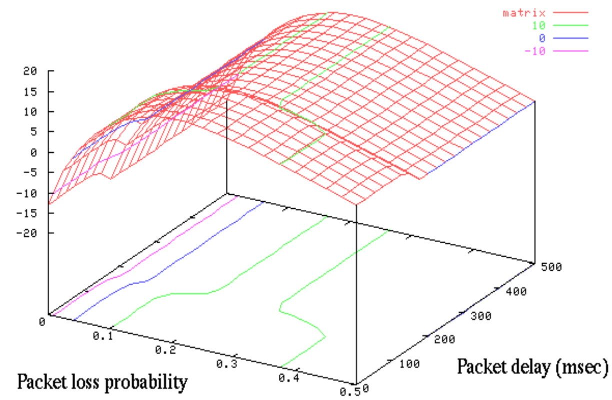

Figure 2: The zone of advantage for lower bit-rate encoders with error redundancy.

3. PROPOSED STUDIES

We propose that a Military E-Model tool designed for tactical environments and systems would be of great value to the DoD in supporting its drive toward VoIP on the GIG. Such a tool would aid in the design of VoIP architectures, in CODEC trade off studies, performance management and adaptive CODEC designs for tactical environments and systems, e.g., the JTRS. In order to develop this tool and its derived benefits, the following investigations into the development of a Military E-Model for DoD applications are required. A five phase study is necessary, as shown in Figure 3.

- *Phase 1. Identify Tactical Environments and Systems* - important to the ultimate utility of a Military E-Model tool for tactical environments is the understanding of the environments and systems where the tool is to be applied. The initial phase of the study is to identify a set of military CODECs and tactical systems, e.g., JTRS, deemed critical to the DoD's deployment plans for VoIP services. These choices will then drive the specific channel characterization studies and the test planning for the MOS and Rhyme tests discussed below.

- *Phase 2. Characterization of the Transport Channels* - the performance of the transport channel and its reflection in packet level metrics must be understood. The specific communications media, e.g., JTRS Waveform 1, Singars, etc. must be identified and empirical measurements made in order to develop an understanding of the packet level performance of these systems. Based upon the measurement data, channel models must be derived which accurately reflect/generate channel performance characteristics. These channel models are required in order to develop a meaningful set of subjective voice quality studies for military CODECs. In previous work, we have developed IP network models based on an impulse driven time series model. These models more accurately represent IP network conditions than previous models. Potentially this type of model could be applied to a wider variety of network types, including assorted tactical and mobile networks.

- *Phase 3. MOS and Rhyme Tests Studies* - Mean Opinion Score (MOS) and Rhyme Test measurements are necessary to begin characterizing the performance of military CODECs and perceived performance by those individuals typical working in tactical, battlefield environments. Several work items are necessary in order to setup meaningful subjective tests. These include a) developing the appropriate question set for the measurements, e.g., should voice quality or speech intelligibility measurements be made, b) developing a set of speech samples for the testing, and c) defining the test plans, building the test environment and executing the tests. This phase would draw from previous work in MOS and Rhyme Tests.

- *Phase 4. Results Analysis and a Military E-Model Development* - The data from the subjective tests are to be statistically analyzed and the equivalent Military E-Model developed for the critical set of DoD systems and CODECs. Aspects of this analysis can draw upon the collective commercial work in developing the ITU-T's E-Model. However, due to potential differences in the nature of the CODEC performance characterization between the commercial world and the DoD and due to certain deficiencies in the current E-Model, aspects of this data analysis will be new. To our knowledge, the E-Model methodology has not been applied to the prediction of voice intelligibility, i.e., Rhyme Tests. Depending upon the focus, the Military E-Model development for the tactical environment can be tailored to the specific application, i.e., Quality, Intelligibility or Both.

- *Phase 5. Monitoring and Control Algorithms* - Once the above tasks are completed, several applications of these tools can be pursued. As examples of the type of applications, we list the following.

  - It is possible to begin investigations of optimal, adaptive CODECs for anticipated tactical, battlefield environments. Design parameters would include methods of channel monitoring, type and level of forward error correction capabilities, encoding rates, CODEC types and adaptive play-out methods. To date, most studies have utilized objective functions which are not strictly based upon estimates of voice quality, e.g., [20] [16] and [21]. One exception is found in [5]. Further, previous CODEC development typically has assumed that constraints related to code size, memory usage and constant bit rates apply. However, in tactical implementations, it is quite possible to design CODECS that use different approaches to the typical CELP-type of algorithm but do not require DSP hardware and can be more robust in operation. The channel characterization and resulting Military E-Models would be useful in these investigations.
  - Another area of application of these capabilities is in the area of network monitoring of VoIP services. We have developed software [7] that is able to recognize the signatures of IP problems on individual VoIP streams. Within an Army Unit of Employment (UoE) and Unit of Action (UoA) deployment, this could, for example, identify remote link congestion as begin a root cause of a voice quality problem. Methods and capabilities such as these could be used in a variety of applications, e.g., letting a distributed management function isolate faults, and providing guidance to endpoints on what strategies to use for mitigation. Further, standards work is necessary to incorporate Military E-Model monitoring information into existing IETF capabilities [8] [10].

4. CONCLUSIONS

We propose a systematic study to access the performance of military CODECs for Voice over IP services deployed in tactical, battlefield environments. The study would build upon an extensive body of work

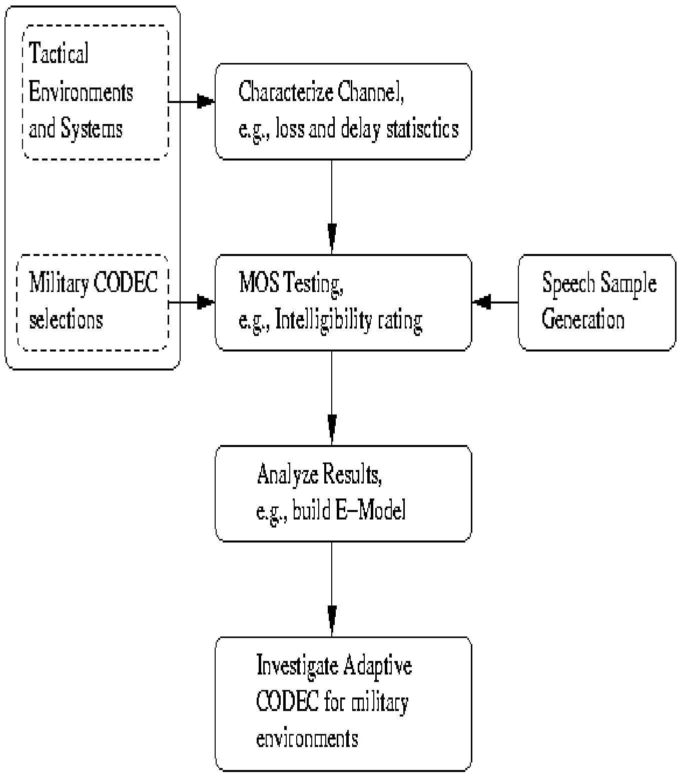

Figure 3: The method toward a Military E-Model.

in the commercial telephony world in accessing VoIP performance, in designing VoIP monitoring systems and in designing adaptive algorithms for improved CODEC performance. We believe that the results of such a study will provide the DoD with an improved understanding of expected performance and utility of VoIP deployments in tactical MANETS such as the JTRS system and other critical networking domains. We also believe that such a study will provide a set of tools of utility in designing VoIP services over tactical networks, in managing these deployments and in designing optimal adaptive CODECs for military applications.

5. ACKNOWLEDGMENTS

We would like to acknowledge discussions with M. Perkins and J. Rosenbluth of AT&T Labs, and P. Chimento and N. Phamdo of JHU/APL.

References

[1] ANSI T1.521-1999, *Packet Loss Concealment for use with ITU-T Recommendation G.711*, December 1999.

[2] ANSI S3.2-1989 (R1999), *American National Standard Method for Measuring the Intelligibility of Speech Over Communication Systems*, December 1989.

[3] Bolot, J.-C., *Characterizing end-to-end packet delay and loss in the Internet*, Journal of High Speed Networks, vol.2, pp 305-323, 1993.

[4] Borella, M.S., Swider, D., Uludag, S. and G.B. Brewster, *Analysis of end-to-end internet packet loss: Dependence and asymmetry*, Technical Report AT031798, 3COM Advanced Technologies, March 1998.

[5] Choudhury, G. and R.G. Cole, *Design and analysis of optimal adaptive de-jitter buffers*, Computer Communications, 27 (6), pp 529-537, 2004.

[6] Clark, A., *Modeling the effects of burst packet loss and recency on subjective voice quality.*, IPTel 2001 Workshop, Columbia University, April, 2001.

[7] Clark, A., *Telchemy Corp.*, http://www.telchemy.com, May, 2001.

[8] Clark, A., *Proposed RTP Control Protocol Extended Reports (RTPCP XR) VoIP Metrics Management Information Base*, draft-ietf-avt-rtcp-xr-mib-00.txt, October 2004.

[9] Cole, R.G. and J. Rosenbluth, *Voice over IP performance monitoring*, ACM J. Computer Communications Review 31 (2), 2001.

[10] Friedman, T., et al, *RTP Control Protocol Extended Reports (RTCP XR)*, Internet Engineering Task Force (IETF), RFC 3611, November, 2003.

[11] Janssen, J., De Vleeschauwer, D., Buchli, M.J.C. and G.H. Petit, *Assessing voice quality in packet-based telephony*, IEEE Internet Computing Special Issue Internet Telephony, 6 (3), pp 48-57, 2002.

[12] ITU-T Recommendation G.729a, *Coding of Speech at 8 kbit/s using Conjugate Algebraic-Code-Excited Linear-Prediction (SC-ACELP)*, March 1996.

[13] ITU-T Recommendation G.711, *Pulse Code modulation (PCM) of voice frequencies*, November 1988.

[14] ITU-T Recommendation G.107, *The E-Model, a computational model for use in transmission planning*, December 1998.

[15] ITU-T Recommendation G.113, *General Characteristics of General Telephone Connections and Telephone Circuits - Transmission Impairments*, February 1996.

[16] Moon, S.B., Kurose, J. and D. Towsley, *Packet Audio Play-out Delay Adjustments: Performance and Algorithms*, ACM/Springer Multimedia Systems, 5:17-20, January 1998.

[17] Paksoy, E., et al., *An Adaptive Multi-Rate Speech Coder for Digital Cellular Telephony*, preprint, January 2005.

[18] Perkins, M., Dvorak, C., Leisch, B. and J.A. Zebarth, *Speech Transmission Performance Planning in Hybrid IP/SCN Networks*, IEEE Communications Magazine, December 1998.

[19] Private communications with M. Perkins of AT&T Labs.

[20] Ramjee, R., Kurose, J., Towsley, D. and H. Schulzrinne, *Adaptive Play-out Mechanisms for packetized Audio Applications in Wide Area Networks*, Proc. IEEE Infocom, 1994.

[21] Sreenan, C., Chen, J., Agrawal, P. and B. Narendran, *Delay reduction techniques for play-out buffering*, IEEE Trans. Multimedia 2 (2), pp 88-100, 2000.