Optimal Audio Transmission over Wireless Tandem Channels

Ala' Khalifeh Homayoun Yousefi'zadeh
Department of EECS
University of California, Irvine
[akhalife,hyousefi]@uci.edu

Abstract

In this paper, we propose a statistical optimization framework for transmitting audio sequences over wireless links. Our proposed framework protects audio frames against both temporally correlated random bit errors introduced by a fading channel and packet erasures caused by network buffering. Forming a two-dimensional grid of symbols, our framework forms horizontal packets that are compensated only vertically against both types of errors. The utilized one-dimensional error correction coding scheme of our framework assigns parity bits according to the perceptual importance of frames such that the Segmented SNR of a received audio sequence is maximized. In addition, the proposed framework suggests an effective way of reducing the packetization overhead of small audio frames.

I. Introduction

With the emergence of broadband wireless networks, the use of audio and video streaming techniques has significantly increased. However, many challenges should be addressed in order to provide a good transmission quality over the error prone wireless networks. Some of these challenges include reducing the packetization overhead of small audio frames as well as protecting audio content against both bit errors introduced by wireless fading channels and packet erasures introduced by network buffering. The use of Forward Error Correction (FEC) codes represents popular alternative of mitigating the effects of bit errors and packet erasures. In addition, utilizing Multiple Input Multiple Output (MIMO) links can improve the quality of wireless links.

In what follows, a brief review of the literature work related to the transmission of audio is provided. Considering the depth of research conducted in the field, the review cannot be exhaustive. Rather, it includes a mentioning of the works more closely related to the subject of interest to this paper. As evidenced by the work of [9] and others, audio transmission is usually performed using a frame-based approach in which a frame consisting of a number of samples is taken from an audio source, encoded, packetized, and sent across a transmission channel. This approach introduces a potentially high packetization overhead due to the large size of the IP/UDP/RTP packet header compared to the small payload size of the audio frames. Transmitting small packets over most of today's Medium Access Control (MAC) protocols is subject to a significant overhead related performance degradation as well as a fairness problem as reported by [16] and [15]. In the literature, many different techniques have been proposed to decrease the overhead of transmission and increase the payload size. In [15], the authors propose concatenating small packets into larger packets such that the overhead is shared among a group of packets instead of being applied to single packets. In [16], the authors propose a voice/audio multiplex-multicast scheme in which multiple packets belonging to different users are combined into large packets. Large packets are received by all users, each user extracts his own corresponding packet and drops the other packets. Header compression is another

This work is supported in part by the Fulbright Fellowship Program.

technique widely used to reduce the packetization overhead [5]. In [8], we propose an optimization framework for protecting audio sequences against bit errors introduced by the wireless channel, such that more protection is applied to the most perceptual audio packets, in addition, the framework suggests a frame grouping technique to reduce the packetization overhead. In [20], utilizing Reed Solomon (RS) FEC codes is proposed in order to provide Unequal Error Protection (UEP) for audio streams. The authors propose two schemes, the first scheme is unequal frame protection where more protection is added to the header portion of the frame and less protection to the data portion of the frame. The second scheme employs UEP where more protection is assigned to the most significant bits of quantized data samples and less protection is assigned to the least significant bits. In [17], the authors suggest a Content-based UEP (C-UEP) framework for transmitting audio over wireless links. The framework protects the most perceptual audio frames against packet loss by generating a redundant secondary stream. The authors of [7] suggest a perceptually controlled error protection scheme for transmitting audio over IP networks. They present a UEP scheme in which the critical frames are transmitted twice at a full and a low bit rate version in order to achieve a high probability of delivering critical frames. A joint source-channel coding scheme of audio and video transmission over the Internet is proposed in [3]. While the sender sends multiple source and channel coding layers, each receiver subscribes to a number of layers that optimize the source-channel coding rate allocation according to its bandwidth and packet loss rate. In [12], a systematic study of FEC for audio packets over the Internet is presented. The authors emphasize that FEC should be added in a controlled way such that it reduces the network congestion and also constraints the source coding rate. Our literature review reveals that most of the cited works assign parity in a heuristic manner as opposed to an optimal manner. Further, they protect audio frames either against bit errors or packet erasures but not both. This paper proposes an optimization framework for transmitting audio streams over MIMO wireless links based on a detailed analysis of the wireless channel. The framework suggests an optimal way for assigning parity bits to audio frames according to the perceptual sensitivity of the frames. It protects audio frames against both bit errors and packet erasures. It also proposes an efficient way for packetizing and transmitting audio frames such that the packetization overhead is minimized.

The rest of the paper is organized as follows. In Section II, we describe our proposed framework based on a capturing of the wireless channel model. In Section III, we formulate the optimization problem and offer an effective solution to it utilizing dynamic programming. In section IV, we describe our experimentation setup and performance evaluation results. Finally, Section V concludes the paper and proposes future work.

II. Framework Description and Channel Analysis

In this section, we provide a description of our proposed framework and analyze the wireless channel model. Fig. 1 depicts the block diagram of our framework. As illustrated, the audio stream is first encoded and compressed using MPEG-4 Bit Slice Arithmetic Coding (BSAC). In specific, we use the MPEG-4 Natural Audio Coding Toolkit publicly available at the ISO website [1]. Next, Unequal Error Protection (UEP) is applied through which more parity bits are assigned to the more important audio frames. The parity assignment is done in an optimal manner as described in Section III such that it jointly protects the audio sequence against both bit errors and packet erasures. As illustrated by the figure, channel coding blocks are aligned on the columns of a grid such that each column corresponds to one block while packets are formed on the rows. As such, each symbol in every packet belongs to a different channel coding block and the loss of a packet results in losing one symbol per channel coding block. Not only the use of this scheme increases the payload size of packets and reduces the packetization overhead, but also it mitigates the effects of packet erasures.

Once packet payloads are formed, the header of each packet is added. In order to protect the header of each packet against bit errors, additional parity bits are assigned to the header bits. For each packet, the amount of parity added to its header bits is calculated such that the packet is lost due to header bit errors with a small probability of ϵ. Without this protection, a single bit error on the header bits will render the entire packet useless. Packets are transmitted over a wireless fading channel which may be utilizing multiple transmit and/or receive antennas. We assume that the total number of blocks, all of individual block sizes, and the number of packets are transmitted in advance such that the receiving end of the link can re-establish the grid and attempt at reconstructing the blocks of the grid from the received packets. Automatic Repeat reQuest (ARQ) scheme is used to guarantee the delivery of these meta data packet.

Once all of the packets are received, the information grid can be reconstructed and channel coding can be applied to the columns of the grid in order to compensate against bit errors and symbol erasures. As the result of applying channel coding, every individual block is either fully recovered or completely discarded. In the case of discarding a block, error concealment is used to replace the discarded block with the content of the previously received blocks. The blocks are then passed to the audio decoder and play back stage.

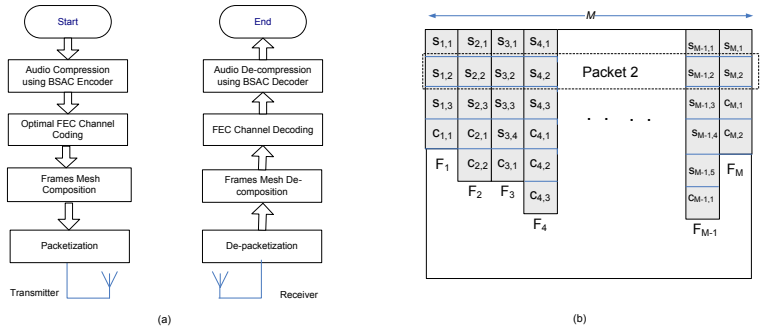


Fig. 1. (a) A block diagram of the proposed framework. (b) The grid alignment illustrating that packets are formed horizontally and coding blocks are formed vertically. $S_{x,y}$ corresponds to the frame source symbols and $C_{x,y}$ corresponds to the frame parity symbols.

In what follows, we briefly describe the wireless channel model, the calculation of the Symbol Error Rates (SERs), and the FEC scheme used to protect packets against random bit errors and packet erasures. A MIMO wireless fading channel is characterized by a temporally correlated pattern of bit loss [19]. In order to capture this loss behavior, we use the two-state Gilbert-Elliott (GE) model. In the GE model, the random corruption pattern of an audio bitstream is described by a two-state Markov chain introducing a good state (G) and a bad state (B). State G represents a bit error rate of ε_G while state B represents a bit error rate of ε_B, where $\varepsilon_B >> \varepsilon_G$. Let $P(t, q, G)$ and $P(t, q, B)$ denote the probability of receiving q bits from t transmitted bits and ending up in state G and B of the GE model, respectively. Then the overall probability of receiving q bits from t transmitted bits under the GE model is calculated as [19]

$$P(t, q) \;=\; P(t, q, G) + P(t, q, B) \tag{1}$$

where the recursive probabilities $P(t, q, G)$ and $P(t, q, B)$ are given by

$$\begin{aligned} P(t, q, G) = \varepsilon_G \left[\gamma\, P(t-1, q, G) \;+\; (1-\beta) P(t-1, q, B)\right] \\ (1-\varepsilon_G) \left[\gamma\, P(t-1, q-1, G) \;+\; (1-\beta) P(t-1, q-1, B)\right] \end{aligned} \tag{2}$$

and

$$\begin{aligned} P(t,q,B) = \varepsilon_B \left[(1-\gamma)\, P(t-1,q,G) \;+\; \beta P(t-1,q,B)\right] \\ (1-\varepsilon_B)\left[(1-\gamma)\, P(t-1,q-1,G) + \;\beta P(t-1,q-1,B)\right] \end{aligned} \tag{3}$$

for $t \geq q > 0$ and the initial conditions

$$\begin{aligned} P(0,0,G) = \; g_{ss} \; = \; \frac{1-\beta}{2-\gamma-\beta} \quad P(0,0,B) = \; b_{ss} \; = \; \frac{1-\gamma}{2-\gamma-\beta} \\ P(1,0,G) = \; \varepsilon_G \left[\gamma\, g_{ss} \; + \; (1-\beta)\, b_{ss}\right] \quad P(1,0,B) = \; \varepsilon_B \left[(1-\gamma)\, g_{ss} \; + \; \beta\, b_{ss}\right] \end{aligned} \tag{4}$$

In the above equations γ is the probability of self transitioning for state G and β is the probability of self transitioning for state B. Further, per state bit error rates ε_G and ε_B can be calculated in terms of the number of transmit/receive antennas, and the average received signal-to-noise ratios. In [19], closed-form expressions describing these per state error rates are identified assuming a flat fading Rayleigh channel. Based on that discussion, the generic modulation symbol error rate of a link associated with a single-transmit W-receive antenna link using Maximum Ratio Combining (MRC) and Z-PSK modulation is identified as

$$\begin{aligned} \varepsilon_G = \tfrac{Z-1}{Z} - \tfrac{1}{\pi}\sqrt{\tfrac{\vartheta_G}{1+\vartheta_G}}\Big\{(\tfrac{\pi}{2} + \tan^{-1}\chi_G)\textstyle\sum_{j=0}^{W-1}\binom{2j}{j}\frac{1}{[4(1+\vartheta_G)]^j} \\ + \sin(\tan^{-1}\chi_G)\textstyle\sum_{j=1}^{W-1}\sum_{i=1}^{j}\frac{\sigma_{ij}}{(1+\vartheta_G)^j}[\cos(\tan^{-1}\chi_G)]^{2(j-i)+1}\Big\} \end{aligned} \tag{5}$$

where $\vartheta_G = SNR_G \sin^2(\frac{\pi}{Z})$, $\chi_G = \sqrt{\frac{\vartheta_G}{1+\vartheta_G}}\cot\frac{\pi}{Z}$, and $\sigma_{ij} = \frac{\binom{2j}{j}}{\binom{2(j-i)}{j-i}4^i[2(j-i)+1]}$. According to the same discussion, Equation (5) can also be used to calculate the modulation symbol error rate of wireless links utilizing Space-Time Block Codes (STBCs) of [14] with the insertion of a proper SNR scaling factor. Relying on BPSK modulation, i.e., ($Z = 2$), Equation (5) can map modulation symbol error rates to bit error rates. In order to differentiate between per state bit error rates ε_G and ε_B, two different measures SNR_G and SNR_B are considered for state G and state B where $SNR_G >> SNR_B$.

For simulating packet erasures, we use the Gilbert (G) model, which can be obtained from the GE model by substituting ε_G and ε_B with 0 and 1, respectively. We apply the model to the packet symbols. If one symbol is erased, the entire packet is marked as erased packets. Next, we propose the use of an RS FEC scheme at the link layer to mitigate the effects of random bit errors and packet erasures. A channel coding symbol is to be differentiated from a modulation symbol and may itself consist of a number of modulation symbols. The maximum block size is determined by the channel coding symbol size s. An RS code operating on an s-bit symbol size can have up to $n = 2^s - 1$ symbols per block. An encoded block contains k data symbols and $C = n - k$ parity symbols. An RS code with C parity symbols can correct up to N_{err} symbol errors and N_{ers} symbol erasures for as long as $2N_{err} + N_{ers} \leq C$ [6]. Suppose the RS coder generates a set of channel coding symbols where each symbol consists of s bits. A channel coding symbol is received error free if all of its s bits are received free of errors. Thus, the probability of receiving a channel coding symbol free of errors under the GE model is described by Equation (1) with $t = q = s$ as $P(s,s)$. Referring to [18] and [19], we rely on a hybrid loss model to describe the probability of channel coding block loss. In our hybrid model, channel coding inter-symbol correlation is assumed not to be significant in comparison with channel coding intra-symbol correlation captured by the expression $P(s,s)$. As such, the probability of channel coding block loss $\Psi(L_m, C_m)$ is described as [6]

$$\Psi(L_m, C_m) = 1 - \sum_{q=0}^{L_m} p\left(N_{err} \leq \lfloor\frac{C_m - q}{2}\rfloor | N_{ers} = q\right) P_{ers}(L_m, q) u(C_m - q), \tag{6}$$

where N_{err} is the number of non-overlapping symbol errors, L_m is the size of the block m in symbols, $P_{ers}(L_m, q)$ is the probability of q symbol erasures out of L_m transmitted symbols. $u(C_m - q)$ is the unit step function defined as

$$u(C_m - q) = \begin{cases} 1 & \text{if } q \leq C_m \\ 0 & \text{if } q > C_m. \end{cases} \tag{7}$$

It follows that

$$p\left(N_{err} \leq \lfloor \frac{C_m - q}{2} \rfloor | N_{ers} = q\right) = \sum_{j=0}^{\lfloor \frac{C_m - q}{2} \rfloor} p(N_{err} = j | N_{ers} = q). \tag{8}$$

If the packets are sufficiently large, the symbols errors can be considered independent and as a result

$$p(N_{err} = j | N_{ers} = q) = \binom{L_m - q}{j} (1 - P(s,s))^j (P(s,s))^{L_m - q - j}. \tag{9}$$

$P_{ers}(L_m, q)$ is the probability of having q symbols erased out of L_m transmitted symbols calculated using Equation (1) by substituting ε_G and ε_B with 0 and 1, respectively. Notice that due to the formation of the grid of blocks, the average probability of packet erasures is approximately the same as the average probability of symbol erasures. We have investigated different low and high values of P_{ers} to study the impact of packet erasure rates on the transmission performance, and as we see on Section IV, the proposed framework can tolerate high packet erasure rates without scarifying the transmission quality.

III. Optimization Formulation and Solution

The main objective of the optimization problem of this section is to find the optimal parity assignment for each frame maximizing the quality of received audio sequence. Each frame obtains different number of parity bits according to its perceptual importance. We use Segmented Signal to Noise Ratio ($SSNR$) [2], as one of the best time domain objective metrics used to evaluate the quality of audio and voice streams for performance evaluation, our proposed optimization technique can be applied to other performance metrics as well. We note that a higher measure of $SSNR$ metric indicates a better quality. The $SSNR$ is defined as

$$SSNR = \frac{10}{M} \sum_{m=0}^{M-1} \log \left\{ 1 + \frac{\sum_{n=1}^{N} x^2(mN+n)}{\sum_{n=1}^{N} [y(mN+n) - x(mN+n)]^2 + \delta,} \right\}, \tag{10}$$

where $x(.)$ is the set of normalized samples of the transmitted audio sequence and $y(.)$ is the set of normalized samples of the received audio sequence. N is the frame length in samples, M is the number of frames of the audio sequence, and δ is a small number used to prevent dividing by zero. Defining Frame Segmented SNR ($FSSNR$) as the Segmented SNR for one frame defined as

$$FSSNR(m) = \log \left\{ 1 + \frac{\sum_{n=1}^{N} x^2(mN+n)}{\sum_{n=1}^{N} [y(mN+n) - x(mN+n)]^2 + \delta.} \right\}. \tag{11}$$

We note that the summation in the denominator of (11) represents the distortion D between the received and the transmitted frames measured in terms of Mean Square Error (MSE). Thus, the $SSNR$ of an audio stream can also be represented in terms of $FSSNR$ as

$$SSNR = \frac{10}{M} \sum_{m=0}^{M-1} FSSNR(m). \tag{12}$$

If frame m is received successfully, $FSSNR(m)$ is expressed as

$$FSSNR(m) = \log\left\{1 + \frac{\sum_{n=1}^{N} x^2(mN+n)}{\delta}\right\}. \tag{13}$$

Notice that in this case, distortion D in the denominator of (11) is equal to zero, i.e., $\sum_{n=1}^{N}[y(mN+n) - x(mN+n)]^2 = 0$. In the event of a frame loss, we use the Insertion-Base Repair (IBR) algorithm of [11] to represent a lost frame. In IBR, a lost frame is replaced by the last accurately received frame or if there is no previously received frames, the next received frame is used. More specifically, $FSSNR$ is calculated using Equation (11) after calculating distortion by computing the MSE between the original reference sample values of the frames $x(n)$ and the sample values of the frames used in the error concealment process. Hence, the value of $E[FSSNR(m)]$ for a frame m is expressed as

$$\begin{aligned}\mathcal{E}[FSSNR(m)] \;=\; & (1-\Psi_m)\log\left\{1 + \frac{\sum_{n=1}^{N} x^2(mN+n)}{\delta}\right\} \\ & + \Psi_m \log\left\{1 + \frac{\sum_{n=1}^{N} x^2(mN+n)}{\sum_{n=1}^{N}[y(mN+n)-x(mN+n)]^2+\delta}\right\}\end{aligned} \tag{14}$$

Equally, we can express

$$\mathcal{E}[SSNR] \;=\; \frac{10}{M}\sum_{m=0}^{M-1}\mathcal{E}[FSSNR(m)]. \tag{15}$$

Consequently, the optimization problem is given by

$$\max_{(C_0,\cdots,C_{M-1})} \quad \mathcal{E}[SSNR] \tag{16}$$

$$\text{Subject To :} \quad \sum_{m=0}^{M-1} C_m \le B_C \tag{17}$$

$$0 \le C_m + R_m < 2^{s_m} - 1, \quad \forall m, \tag{18}$$

s_m is the symbol size of the block m bits chosen such that the block size (L_m) consists of the frame payload symbols R_m, and the parity symbols C_m assigned to that frame does not exceed the maximum RS block size of $(2^{s_m} - 1)$ symbols [13]. Further, B_C is the parity budget allocated to transmit the audio sequence which equals to $B_C = B_T - B_R - B_H$, where B_T is the total budget allocated to transmit the audio sequence, B_R is the payload budget(the size of audio frames), and B_H is the packetization overhead. We calculate $B_H = (2^s - 1) * H$ where $2^s - 1$ is the maximum number of packets , which corresponds to the maximum block size a frame can have, H is the sum of the UDP/RTP/IP compressed header size and the header parity symbols added to protect the header against bit errors.

Then the values of $\mathcal{E}[FSSNR]$ for each frame, corresponding to all possible parity symbol assignments that each frame can have, which is determined by the maximum allowable size of the RS block size, are calculated and inserted into a so called $SSNR$ matrix. Denote the values of this matrix as $V(r, w)$ where r is the row index and w is the column index. Fig. 2(a) demonstrates how this matrix is calculated. Consider an audio sequence of M frames where the number of rows of the $SSNR$ matrix is equal to M and each row corresponds to one frame . To find the elements of the row associated with frame m, we simulate a loss event for frame m and conceal the frame using the IBR concealment algorithm. Then, we calculate the distortion measured in terms of MSE between the original frames and the frames used in IBR concealment algorithm. Next, we use the distortion measure to calculate $\mathcal{E}[FSSNR(m)]$ and evaluate the perceptual sensitivity of the frame m. To calculate $\mathcal{E}[FSSNR(m)]$, we calculate the probability of losing frame m for all possible

parity assignments. More specifically, $\mathcal{E}[FSSNR(m)]$ is calculated using Equation (14) for each parity assignment in the set $\{0, 1, \cdots, B_C\}$. The first column element of row m of $SSNR$ matrix is set as the value of $\mathcal{E}[FSSNR(m)]$ with a parity assignment of zero. Then the number of parity symbols is incremented by one symbol, the block size associated with this assignment is calculated, and compared against Constraint (18). If the constraint is satisfied, Ψ_m and $\mathcal{E}[FSSNR(m)]$ corresponding to this assignment are calculated, and inserted into the second column element of row m of $SSNR$ matrix. The process of incrementing the parity by one symbol, calculating the block size, checking the block size constraint, calculating Ψ_m and $\mathcal{E}[FSSNR](m)$ is repeated till the maximum allowable block size determined by Constraint (18) is reached. The rows associated with each block m where $m \in \{0, \cdots, M-1\}$ are filled the same way as described above. After that, using dynamic programming, we find the optimal parity assignments for each block maximizing the $\mathcal{E}[SSNR]$ of the overall sequence for a given budget B_C. We solve the optimization problem using dynamic programming [4]. We divide the original problem into sub-problems and solve the sub-problems optimally in order to construct the optimal solution of the original problem. We refer the interested reader to [8] for a detailed description of our utilized dynamic programming algorithm and its associated complexity analysis.

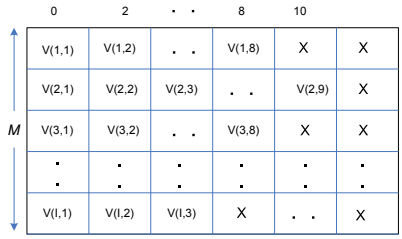
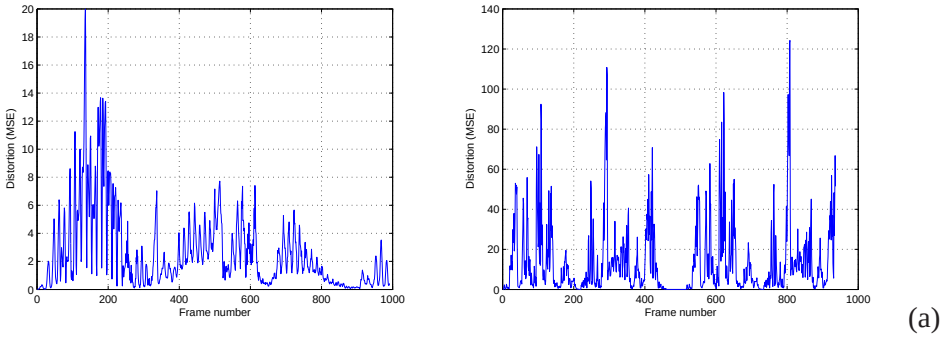


Fig. 2. The $SSNR$ matrix. Symbol X represents points at which the number of parity symbols for a given frame exceeds the maximum possible block size.

IV. Performance Evaluation

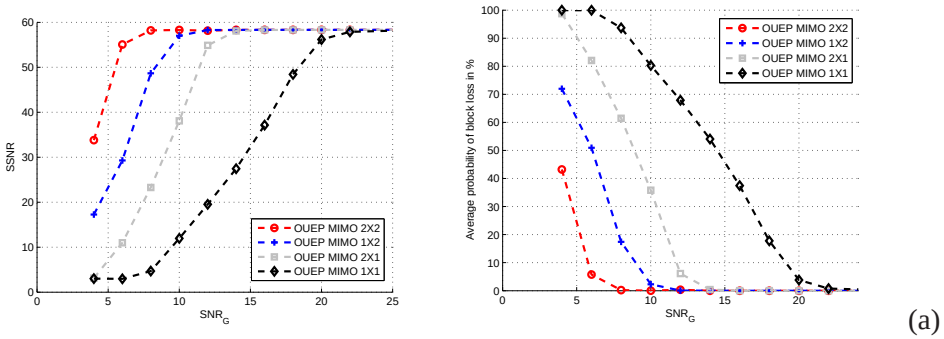
In this section, we present our performance evaluation results based on our proposed framework. We consider the transmission of the MPEG-4 encoded sequences over wireless links. As indicators of sequences with various characteristics, our reported results relate to the sopr44-1 opera sequence and the vioo10-2 music sequence audio clips. For both clips, the sampling rate is $48k$ sample/second with a sample size of 16 bits. Each frame includes 1024 samples [10]. Our protocol stack model utilizes header compression technique to compress the headers of Internet Protocol (IP), User Datagram Protocol (UDP), and Real-Time Protocol (RTP) resulting in a header size of 5 bytes [5]. We generically emulate the effects of PHY and MAC layers through the two-state GE model for simulating bit errors, and the two-state G model for simulating packet erasures. We utilize RS codes to protect the audio blocks using a symbol size of s bits. The symbol size s is chosen such that the combined size of payload, and parity in symbols does not exceed the maximum allowable block size determined by Equation (18). We choose $s = 8$ bits allowing a maximum block size up to 255 bytes. In our experiments, we utilize a play back buffer at the receiving end. We choose the buffer size according to transmission delay, jitter, and play back consumption rate such that the buffer is neither in the state of overflow nor in the state of underflow during play back. Packets are modulated using BPSK modulation and transmitted over the wireless channel. For the wireless channel model, the transition probabilities of the GE model are set as $\gamma = 0.99875$ and $\beta = 0.875$ representing average burst lengths of 800 and 8 bits for state G and B, respectively. For G model, we choose

$\gamma = 0.99875$ and β is used as a parameter to vary the average probability of packet erasure rate (Pers) between 1% and 15%. Further, we consider $SNR_G = 10SNR_B$ to differentiate between the qualities of a link in state G and B. We consider four different MIMO configurations representing improved SER characteristics of a link in an ascending order. They are namely (1) single-transmit single-receive (1×1) utilizing MRC, (2) double-transmit single-receive (2×1) utilizing STBC, (3) single-transmit double-receive (1×2) utilizing MRC, and (4) double-transmit double-receive (2×2) utilizing STBC. At the receiving side, the grid of frames shown in Fig 1 is reconstructed after receiving the entire sequence. Then using parity symbols, the blocks are attempted to be compensated against bit errors and symbol erasures. In the case of discarding a block, the last received block is used to conceal it. With the exception of Fig.3, Fig. 4.b, and 5.b, our generated performance evaluation curves indicate $\mathcal{E}[SSNR]$ measured in dB scale for the entire received audio stream on the vertical axis. Fig. 3 shows the MSE distortion and Fig. 4.b as well as 5.b show the average probability of block loss in $\%$. All figures show SNR_G in db scale on the horizontal axis. Every point on each curve indicates an average value taken over at least 50 experiments. We conduct our experiments by varying the MIMO configuration, the budget, and the average packet loss rate.



(a) (b)

Fig. 3. The distortion in terms of MSE for each frame for (a) sopr44-1 audio clip (b) vioo10-2 audio clip



(a) (b)

Fig. 4. A performance comparison of different MIMO configurations utilized in conjunction with our Optimal UEP (OUEP) algorithm. The vioo10-2 audio clip with an average packet loss ratio of 5% with total budget $B_T = 142KB$ are used to depict (a) SSNR, and (b) the average probability of block loss.

In order to illustrate the importance of deploying UEP for the audio frames, we show how the distortion varies from one frame to another. Fig. 3 depicts the MSE distortion associated with losing each frame of the sequence. Our optimization technique assigns the parity bit such that the most sen-

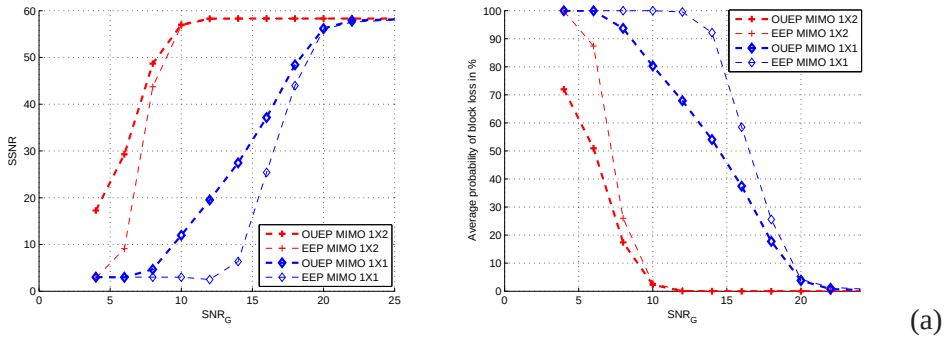


Fig. 5. Performance analysis of vioo10-2 audio clip with an average packet loss ratio of 5% with total budget of $B_T = 142KB$ utilizing 1×1, and 1×2 MIMO configuration to depict (a) SSNR, and (b) the average probability of block loss.

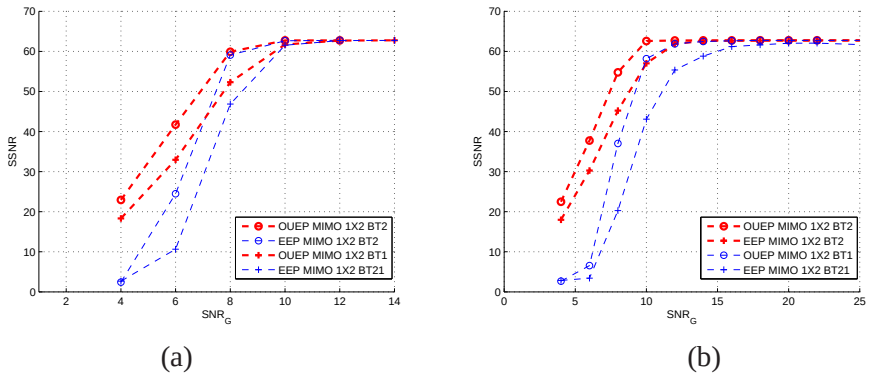


Fig. 6. Performance analysis of sopr44-1 audio clip with an average packet loss ratio of (a) 5% and (b) 10%, with two different budgets ; $B_{T1} = 139KB$ and $B_{T2} = 144KB$, utilizing 1×2 MIMO configuration.

sitive frames are protected most increasing the probability of receiving these frames correctly. Next, we provide a performance comparison of applying the Optimal UEP (OUEP) algorithm of Section III in conjunction with using different MIMO configurations. Fig. 4.a compares the performance of the four MIMO scenarios indicated above for vioo10-2 audio clip associated with a low packet loss ratio of 5%. The curves show a hysteresis pattern of improvement as the quality of the channel improves. However, the transitioning segment of a curve shifts to the left as a MIMO configuration with a better SER characteristic is used. Fig.4.b shows that utilizing a higher quality MIMO link results in achieving a lower average probability of block loss. The results are consistent with our other experiments performed using a variety of audio clips. Next, we provide a performance comparison of our OUEP algorithm against Equal Error Protection (EEP) scheme serving the role of our baseline. Given B_C, EEP assigns the parity budget of a block m as $B_m = \frac{R_m}{\sum_{m=1}^{M} R_m} B_C$.

Fig.5 compare the performance results of the two schemes. The most important observation based on the results of figures shown here is that our OUEP scheme significantly outperforms EEP independent of the choice of audio clip with different MIMO configurations, different FEC rates, and different packet loss rates. Aside from the observation above, the following observations are of importance. Fig.6 compares the two schemes under different packet loss rates and using different budgets. The results show that when the packet loss rate is high, the performance curves transition-

ing segments shift to the right. In the case of deploying EEP scheme, the curves may not reach to the saturation level even on high values of SNR_G, while the curves always reach to the saturation level using OUEP.

V. Conclusion

In this paper, we proposed an optimization framework protecting an audio sequence jointly against random bit errors and packet erasures while reducing the packetization overhead of small audio frames. Forming a two-dimensional grid of symbols, our framework formed horizontal packets compensated vertically against both types of errors. The utilized one-dimensional error correction coding scheme of our framework assigned parity bits according to the perceptual importance of frames such that the Segmented SNR of a received audio sequence could be maximized. Our simulation results revealed the effectiveness of our proposed framework under a variety of link qualities. As a part of our ongoing work, we are in incorporating the effects of delay as it pertains to the receiver buffer size and play back deadlines in our optimization framework.

References

[1] ISO Standard Website, http://isotc.iso.org/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm.

[2] A. Bayya and M. Vis. Objective Measures for Speech Quality Assessment in Wireless Communications. *In Proc. IEEE ICASSP*, 1996.

[3] P. Chou, A. Mohr, A. Wang, and S. Mehrotra. FEC and pseudo-ARQ for receiver-driven layered multicast of audio and video. *In Proc. IEEE DCC*, 2000.

[4] T.H. Cormen, Charles E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT press, second edition, 2003.

[5] M. Degermark, M., B. Nordgren, and S. Pink. Low-loss TCP/IP header compression for wireless networks. *Mobile Computing and Networking:*, 1997.

[6] F. Etemadi and H. Jafarkhani. An Efficient Progressive Bitstream Transmission System for Hybrid Channels With Memory. *IEEE Transactions on Multimedia*, 2006.

[7] E. Hellerud, J.E. Voldhaug, and U.P. Svensson. Perceptually Controlled Error Protection for Audio Streaming over IP Networks. *In Proc. IEEE ICDT*, pages 30–30, 2006.

[8] A. Khalifeh and H. Yousefi'zadeh. An Optimal Scheme to Protect Audio Against Wireless Channel Impairtments. *Technical Report*, 2007.

[9] S.K. Marks and R. Gonzalez. Object-Based Audio Streaming Over Error-Prone Channels. *In Proc. IEEE ICME*, pages 261–264, July 2005.

[10] M. Olausson, A. Ehliar, J. Eilert, and D. Liu. Reduced Floating Point for MPEG1/2 Layer III Decoding. *In Proc. IEEE ICASSP*, pages V– 209–12, May 2004.

[11] C. Perkins, O. Hodson, and V. Hardman. A Survey of Packet Loss Recovery Techniques for Streaming Audio. *IEEE Network*, pages 40–48, Sept. 1998.

[12] Matthew Podolsky, Cynthia Romer, and Steven McCanne. Simulation of FEC-based error control for packet audio on the internet. *INFOCOM (2)*, pages 505–515, Apr.

[13] B. Sklar. *Digital Communications: Fundamentals and Applications*. Prentice Hall, second edition, 2001.

[14] V. Tarokh, H. Jafarkhani, and A.R. Calderbank. Space-Time Block Coding from Orthogonal Designs. *IEEE Trans. Information Theory*, July 1999.

[15] J. Tourrilhes. Packet Frame Grouping: Improving IP Multimedia Performance over CSMA/CA. *In Proc. IEEE ICUPC*, pages 1345–1349, Oct. 1998.

[16] W. Wang, S.C Liew, and V.O.K. Li. Solutions to Performance Problems in VoIP over 802.11 Wireless LAN. *IEEE Trans. Vehicular Technology*, pages 366–384, June 2005.

[17] Y. Wang, A. Ahmaniemi, D. Isherwood, W. Cheng, and D.Huang. Content-based UEP: a new scheme for packet loss recovery in music streaming. *Proceedings of the eleventh ACM international conference on Multimedia*, 2003.

[18] H. Yousefi'zadeh, H. Jafarkhani, and F. Etemadi. Distortion-Optimal Transmission of Progressive Images over Channels with Random Bit Errors and Packet Erasures. *In Proc. IEEE DCC*, 2004.

[19] H. Yousefi'zadeh, H. Jafarkhani, and M. Moshfeghi. Power Optimization of Wireless Media Systems with Space-Time Block Codes. *IEEE Trans. Image Processing*, July 2004.

[20] C.W. Yung, H.F. Fu, C.Y. Tsui, R.S. Cheng, and D. George. Unequal Error Protection for Wireless Transmission of MPEG Audio. *In Proc. IEEE ISCAS*, pages 342–345, July 1999.