GATESAIR

Methods for Mitigating IP Network Packet Loss in Real Time Audio Streaming Applications

Keyur Parikh and Junius Kim
Harris Broadcast
Mason Ohio

Abstract – *Increasingly, IP based networks are being used for transport of real-time broadcast quality audio. IP networks provide clear benefits of cost and flexibility over circuit switched networks. However IP networks can have impairments which must be mitigated in order to realize the quality of service that broadcasters are used to with circuit switch networks. Among these impairments, packet loss poses the toughest challenge. In many non real-time applications, transport protocols such as TCP or reliable UDP can be used to re-transmit lost packet. However for real-time media streaming applications that are delay sensitive and where multicasting is used, such protocols cannot be used. Therefore, techniques used for packet loss mitigation have to minimize delay and work in unidirectional network paths The challenges faced as we encounter different packet loss patterns in real networks is the effectiveness of the mitigation techniques depend upon the model these packet losses follow. In this paper we first show how packet losses in real networks can be modeled. We then show results of the effectiveness of different mitigation techniques using these models and how they can be cascaded to provide for a scalable method to mitigate varying levels of packet loss.*

NETWORK CHALLENGES

There are several challenges associated with transporting audio over IP networks, among these are: network jitter, duplicate and out-of-order packets, network failures, and packet loss. Let's examine each of these in detail along with mitigation techniques that can be deployed within the architecture of an audio over IP system.

Network Jitter

Network jitter is defined as a variance in end-to-end one way delay time of packets. It is also referred to as Packet Delay Variance (PDV). Network jitter can be caused by transmission system factors such as congestion on the router and switches. If not handled properly, it can cause missing packets to occur if the receiver's jitter buffer is unable to handle packets that arrive too late or too early. Proper sizing and configuration of a receive jitter buffer, either statically or dynamically based on the measured jitter is used to absorb the network PDV.

Duplicate and Out of Order Packets

Duplicate packets at the receiver can be caused by inappropriate link level retransmission or switching problems, while out of order packet generally point to a layer 3 routing event. In either case, if these are not handled properly at the receiver, audio distortion will occur. By using Real-time Transport Protocol (RTP) which provides for packet sequence numbers, duplicate packets can be discarded and the out of order packets can be re-sequenced prior to playback.

Network Failure

The involvement of layer 2 and 3 packet forwarding protocols makes a packet switched network more complex than a circuit switched connection. The added protocol complexity means that, besides a physical network failure, problems in layer 2 and layer 3 protocols such as a switching loop or route flap can also break a network path.

To avoid network service outages, broadcasters are increasingly employing multiple IP network connections. When these connections are used concurrently, identical audio stream packets may be sent over different network connections for diversity. On the receive side, the system needs to provides means to correlate and assimilate the packets across multiple streams such that packets from any one of these streams at any given time can be used in a "seamless" or "hitless" manner. The network diversity inclusion within the architecture not only provides for protection against a single network failure, but as shown in an upcoming section, it is one of the most effective tools to reduce the effective packet loss rate of a system.

Packet Loss

IP packet losses can occur without a complete failure of a network. These losses occur for many reasons and vary based on type of the network connection. For a well managed, guaranteed bandwidth type connection, the packet loss rate increase can occur due to routing changes as a result of an individual link failure or a link (possibly terrestrial or satellite) maybe degraded. For a "best effort" type service, packet loss may be a result of congestion in the network. Furthermore the packet loss rate and pattern on a network can vary over time and therefore the mitigation techniques need to dynamically adapt to these conditions in order to maintain effectiveness.

Although audio packet loss concealment techniques such as replaying previous packet or energy substitution can be applied, the listener has a degraded experience when the packet loss rate increases beyond a "soft" threshold or packet loss occurs in bursts. This can be especially so with compressed audio. Compressed audio carries longer duration of audio per packet as opposed to uncompressed audio. Therefore as the packet loss rate increases or loss pattern changes, the effectiveness of the concealment techniques can also start to deteriorate and packet loss mitigation techniques must be used along with concealment to maintain a high quality user experience.

Mitigation of packet loss is a difficult challenge to overcome as the effectiveness of the techniques depends upon the pattern of packet loss and therefore understanding the packet loss model of a network is essential. Another critical constraint, in real time audio broadcasting, is keeping the end-to-end delay low. While delay on the order of tens of seconds maybe acceptable for internet audio streaming, real-time radio broadcasting applications typically require delays that are orders of magnitude less. In addition to the delay constraint, in many cases, network paths are either unidirectional or multicasting is deployed. This makes usage of re-transmission protocols such as Transmission Control Protocol (TCP) unsuitable. For these reasons, the usage of RTP over User Datagram Protocol (UDP) as the transport layer has been standardized for transport of real-time media over IP networks. This began with Voice over Internet Protocol (VoIP) applications and it is now standardized by European Broadcasting Union (EBU) with the N/ACIP interoperability standard for audio and by Society of Motion Picture and Television Engineers (SMPTE) with the 2022 standard for video.

Packet loss mitigation techniques associated with RTP over UDP transport require the usage of additional network bandwidth. However, advancements in audio compression algorithms has achieved near uncompressed audio quality with fraction of the bandwidth required for uncompressed audio transport. This reduction in bandwidth using audio compression along with falling prices for bandwidth and coupled with increasing availability of Internet Service Providers (ISP)s for subscribing multiple network connections, allows for schemes such as redundant streams and Forward Error Correction (FEC) to be implemented. This combined with packet interleaving and network path diversity can create a scalable set of system tools for combating different packet loss patterns. This allows users to save operational expense by streaming reliable broadcast quality audio using unmanaged or best effort type IP networks.

UNDERSTANDING PACKET LOSS MODELS

Packet loss patterns differ from network to network and over time. These patterns fall into one of two major model categories: random or burst loss. In the random loss model, each packet has an equal probability of getting lost. In other words consecutive packet to packet loss probabilities are uncorrelated. In the burst loss model consecutive packet to packet loss probabilities are correlated and losses tend to occur in bursts.

Most real world networks losses can be modeled with burst loss model which can be simulated using a 4-state Markov model as shown in Figure 1.

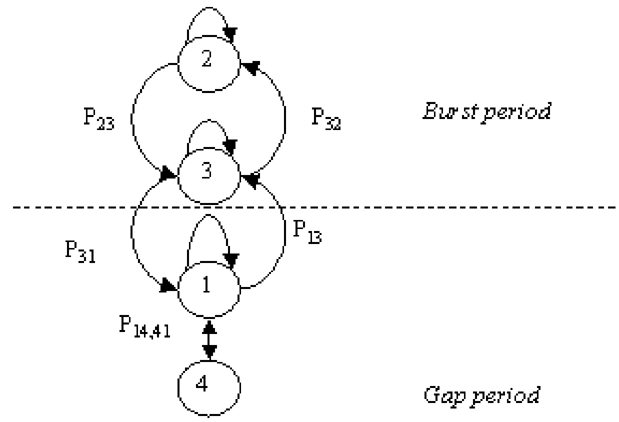

FIG 1 FOUR-STATE MARKOV MODE

The four-state Markov model is a combination of two 2-state Markov sub-models that represent a burst period in which packets are received and lost according to a first 2-state model and gap periods during which packets are received and lost according to a second 2-state model [2]. Where:

State 1 - Packet is received successfully in gap period
State 2 - Packet is received within a burst period
State 3 - Packet is lost within a burst period
State 4 - Isolated packet lost within a gap period

Generally, in the burst loss model, losses are divided between two periods: burst period and gap period, as shown in Figure 1. The burst period is where majority of the losses occur, while the gap period is when isolated losses occur. The parameters used to practically characterize a burst loss model are: burst density, gap density, burst length, and gap length. The algorithm for calculating these parameters is explained in more detail in RFC 3611 [4]. The burst density indicates the probability of losing a packet within the burst loss period, while gap density is an indication of the loss probability within the gap period – or quiet period. Burst length and the gap length indicate the duration of each period in terms of packets or time. Of these four parameters, the burst density is the most important as it indicates the degree of randomness of the overall loss. Lower burst density implies that the losses are spread out and appear more random. As the burst density increases, the chances of multiple consecutive packet losses increase. For example, the effort and technique to mitigate an average packet loss rate of 1% with burst density of 80% compared to same average packet loss rate with same burst length, but with burst density of 40% are vastly different. The 40% burst density loss appears more random and techniques such FEC can be very effective, while a burst density of 80% requires more than just FEC to bring the effective packet loss rate to an acceptable level. Therefore it is important to calculate the burst loss parameters so the most effective and efficient techniques can be utilized.

Packet Loss Mitigation Techniques

In this section we will review several packet loss techniques with different network topologies along with simulation results.

Single RTP Stream with FEC

In this model, a single RTP stream with FEC is sent from an audio encoder to a decoder over the Wide Area Network (WAN) as shown below.

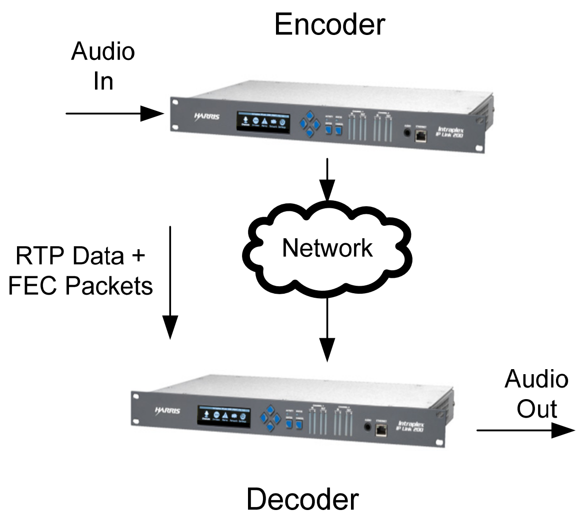

Fig 2 FEC Protected RTP Stream

RTP FEC has been around for several years and its use has been specified in several RFCs such as RFC 2733 [3] and RFC 5109 [5]. In Figure 2, the data flow starts at the encoder, which ingests PCM audio samples and generates an encoded frame. The encoded frame is then packetized with RTP to generate a stream of audio packets. Concurrently, FEC packets are then generated from a matrix of audio packets. Both RTP and FEC packets are streamed to a receiver or audio decoder, where they are de-jittered using a receive jitter buffer. The decoder will periodically pull the next packet to be decoded from the receive jitter buffer for play-out. If a packet is missing, then the corresponding FEC and audio packets are used to re-create the missing packet. If a missing packet cannot be created, then the decoder's concealment technique will fill in the time gap for the missing packet. FEC uses additional network bandwidth to reduce the packet loss rate. However, the effectiveness of how well the FEC works depends on several factors such as the type of FEC being utilized as well as the packet loss model.

RTP Level FEC and its Effectiveness

FEC packets are generated by arranging the RTP data packets into a two dimensional matrix of N rows and M columns and then XORing the RTP packets (including RTP header) in each row or column. Single dimension FEC generally creates only column FEC packets, while two dimensional FEC creates both column and row FEC packets. Table 1 shows the rows and columns with the RTP packets represented sequentially as 1, 2, 3, all the way to 16 for the 4x4 matrix. On the recovery side, a lost packet can be recovered by XORing the FEC packet with the rest of the column or row data packet. The recovery algorithm works over the full matrix of data and FEC packets to recover packet in an iterative manner. The bandwidth overhead for FEC packets is the ratio of the FEC packets to data packets in the matrix. As an example, Table 1 shows a 4x4 two-dimensional matrix which has 8 FEC packets to every 16 data packets; hence 50% additional bandwidth is required for the stream.

	Col 1	**Col 2**	**Col 3**	**Col 4**	**F(x)**
Row 1	1	2	3	4	XOR(1,2, 3,4)
Row 2	5	6	7	8	XOR(5,6, 7,8)
Row 3	9	10	11	12	XOR(9, 10,11,12)
Row 4	13	14	15	16	XOR(13, 14,15,16)
F(x)	XOR(1, 5,9,13)	XOR(2, 6,10,14)	XOR(3, 7,11,15)	XOR(4, 8,12,16)	

Table 1 4x4 Two-Dimensional FEC Matrix

The correction capability of FEC is dependent on a number of factors such as the amount of FEC packets, the size of the matrix, and matrix dimensions. A larger matrix size provides better protection for burst loss. However, the delay at the receiver is also higher since N x M data packets need to be buffered at the FEC generator. The column FEC packets provide burst packet loss protection up to the number of columns in the matrix. The row FEC packets provide random packet loss protection. In theory FEC can be used to effectively recover most types of packet losses. In practice, due to the delay requirements of real-time audio streaming, there are constraints to the sizes of the matrices used.

Effectiveness of FEC Matrices for Random Packet Loss

Figure 3 and Table 2 illustrates the correction capability for some combinations of the 2-dimensional FEC matrix when subjected to random packet losses - as illustrated on X-axis on Figure 3. The Effective Packet Loss (EPL) after recovery is shown on Y-axis.

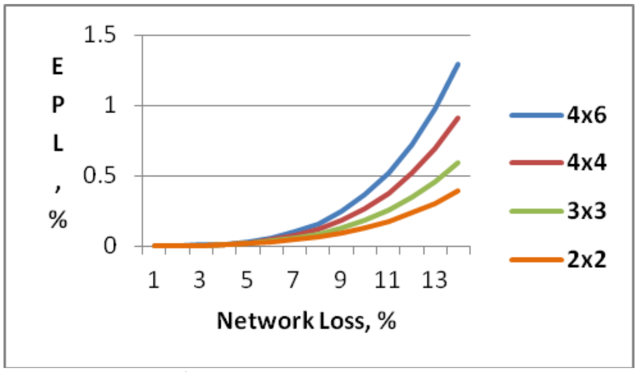

FIG 3 FEC RECOVERY FOR RANDOM PACKET LOSS

FEC Matrix	EPL for 1% Ntwk Loss, (%)	EPL for 2% Ntwk Loss, (%)	EPL for 5% Ntwk Loss, (%)	Ntwk Band-width, (%)
2x2	8×10^{-5}	7.8×10^{-4}	0.014	100
3x3	6×10^{-5}	9.5×10^{-4}	0.017	66
4x4	6×10^{-5}	9.6×10^{-4}	0.020	50

TABLE 2 FEC RECOVERY FOR RANDOM PACKET LOSS

As we can see from Figure 3 and Table 2, when it comes to random packet losses, FEC matrices are very effective. Even when the network packet loss rate approaches 5%, the EPL rate with all of the FEC schemes can provide good quality audio quality, especially when coupled with loss concealment. The 2x2 matrix gives the best EPL rate, but it also has the highest bandwidth overhead.

Effectiveness of FEC Matrices for Burst Packet Loss

Unfortunately, packet losses in real networks don't tend to exhibit total randomness, so let's examine the performance of FEC matrices when they are subjected to varying degree of burst packet loss.

In the Figure 4, the average packet loss rate is 1%, the burst length is 16 packets and the burst density is varied across the X-axis. The Y-axis provides the corresponding EPL for different schemes tested. The gap density which measures the probability of isolated packet losses was ignored from the simulation without loss of any appreciable resolution. Burst length of 16 was used with AAC-LC algorithm which made the burst duration approximately 340 milliseconds.

Looking first at Figure 4, you notice that as you vary the burst density along the X-axis, going from 0 to 80%, the effectiveness of all FEC scheme start to deteriorate. This is because as burst density increases, the packet loss model becomes less random and there is an increase instance of multiple consecutive packet losses or burst losses. This causes the FEC scheme to become less effective. Based on our subjective testing with energy substitution type error concealment and AAC-LC encoding, as long as the EPL was less than 0.1%, the audio quality was acceptable. By looking at the Table 3, when burst density is more than 40% none of the FEC schemes are effective.

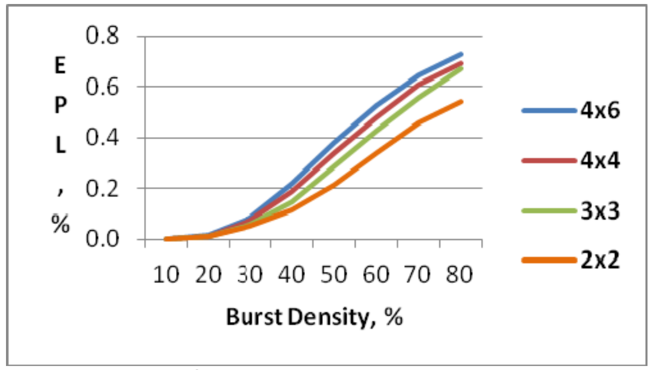

FIG 4 FEC RECOVERY FOR BURST PACKET LOSS

FEC Matrix	EPL for 1% Ntwk Loss (80% Burst Density), (%)	EPL for 2% Ntwk Loss (40% Burst Density), (%)	Ntwk Band-width, (%)
2x2	0.54	0.22	100
4x4	0.69	0.38	50
4x6	0.61	0.40	42

TABLE 3 FEC RECOVERY FOR BURST PACKET LOSS

Table 3 shows an interesting point, in real networks, just knowing the average packet loss rate does not tell the entire story, burst density along with duration are also critical. For instance, looking at 2% average loss rate column, we see that the FEC schemes are performing better than the 1% average loss rate column. This is because the 2% average loss rate column has 40% burst density where the losses occur frequently but are more disperse while the 1% average packet loss rate column has 80% burst density, where losses occur less frequently, but when they occur they wipe out most packets in the burst.

So clearly, as the burst density increases, simply turning on FEC with a reasonable size matrix is not going to be good enough, additional measures need to be taken to keep the audio quality at an acceptable level.

Time Diversity vs. Interleaving Matrix for Burst Packet Loss

Let's take a look at two options to mitigate burst packet losses for audio streams using a single WAN network. The first option is to use multiple redundant streams. The redundant streams are composed by duplicating the original RTP stream packets and sending each stream of the redundant group in a time diverse manner with respect to one another. For example, if we are sending two streams, the first stream would be sent with a relative time offset of zero and the duplicate stream would be delayed with respect to the first by some number of packets as determined from the burst density and burst length of the network loss model. This method will neutralize the burst loss at an expense of higher bandwidth and delay. The second method involves usage of interleaving on the encoder along with a FEC. The idea is that usage of interleaving randomizes the burst packet loss and to make it appear more random, which in turn makes the FEC scheme more effective at recovering lost packets. This method can be more bandwidth efficient than the first, however it incurs delay since interleaving increases delay. This is because the entire interleaved period must be received before the packets can be decoded. For example, consider the case of high burst density with burst length of 16 packets. With the first method your delay factor would be slightly higher than 16 packets, while the second method would requires a 16x16 interleaving matrix, implying a receive buffering of at least 256 packets. In general, to handle a burst length of n packets, the first method would incur buffering of around n packets, while the second method would need n^2 packets.

For real-time audio application where overall play-out delay is critical, a practical method would be usage of redundant streams with time diversity. Interleaving, when used should be limited to a smaller matrix and should be used in combination with redundant streams as explained in the next section.

PUTTING THE TOOLS TOGETHER

So far we have seen that using RTP level FEC can be effective as long as the losses appear random. For burst packet loss, the use of time diverse redundant streams is a practical method. However, as we encounter different levels of packet loss rates, these independent tools need to be combined to form a scalable method for packet loss mitigation. For example, combination of redundant time diverse streams with FEC and interleaving over a single network can provide not only burst packet loss protection, but interleaving can tend to randomize the remaining losses for the FEC scheme to be more effective. In some cases however, a single network may experience burst losses of long duration such that redundant time diverse streams with above combination may not be effective. For these cases, network diversity can be used, with each network having its own combination of streams with protection as shown in the Figure 5.

Different network situation require varied “Swiss army knife” type of networking tools to be used in combination and in a scalable manner. A robust IP audio codec system should allow redundant streams to be grouped together with independent control for network diversity, time diversity, FEC, and interleaving. At the receiver side, the system should utilize an intelligent buffering scheme to re-sequence out of order packets, discard duplicate packets, and deploy usage of FEC to recover lost packets.

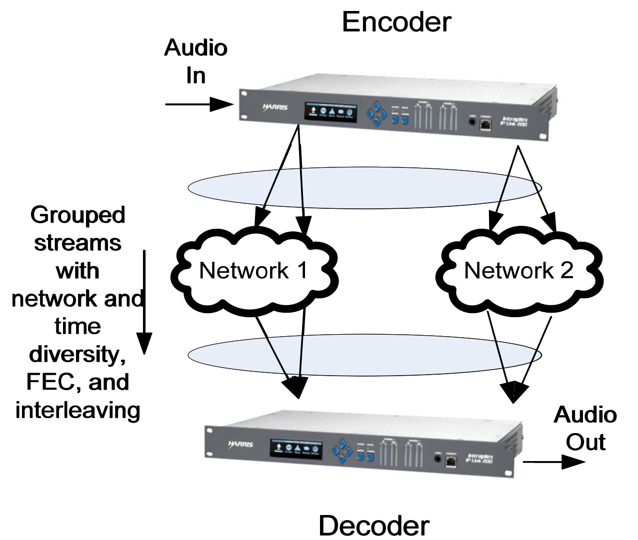

FIG 5 NETWORK AND STREAM DIVERSITY

Putting the Method to Test

We subjected an IP audio codec system using redundant streaming methodology to a burst packet loss model which generated 1% average packet loss with 80% burst density and burst length of 16 packets. A Linux router with NetEM package was used to generate burst losses. The codec algorithm used was AAC-LC with 21.3 millisecond packet interval. In the test cases where multiple networks were used, the router on each network was subjected to the same exact burst loss model.

The results for some test cases are summarized in Table 4.

Group Configuration	Number of Net-works	Effective Packet Loss (EPL), (%)	Ntwk Band-width, (%)
2 streams without time diversity	1	0.73	100
2 streams with time diversity of 16 packets	1	0.10	100
2 streams with one stream with time diversity and 50% FEC	1	0.04	150
1 stream on network 1 and 1 stream on network 2	2	0.009	100
1 stream on network 1 with 50% FEC and 1 stream on network 2 with 50% FEC	2	0.003	200

TABLE 4 BURST LOSS RESULTS FOR “STREAM SPLICING”

Looking at the Table 4, the first column indicates the Group configuration of the streams, the 2nd column indicates number of networks that were involved with the test case, the 3rd column is the EPL as measured prior to decoding, and the last column indicates the additional bandwidth required for the stream.

Looking at the results, we see that in first configuration, where we are simply duplicating stream packets without time diversity, the EPL is still fairly high. In the second configuration, we turn on time diversity on one of the streams and we see the EPL drop significantly. In the third configuration we combine FEC with redundant stream and we see EPL drop down below 0.1%. The fourth configuration is where we see the power of network diversity and by using multiple unmanaged network connections one can achieve reliability of an expensive managed network. The last configuration is where we add FEC to network diverse streams to bring the EPL down further.

For real networks with burst type losses, the third configuration should be applied to each network interface to ensure that the EPL does not rise above 0.04% in the event of a complete single network failure.

ADAPTING TO CHANGING NETWORK CONDITIONS

Packet loss patterns of a network may vary over time, the change can be permanent or it can be temporary or periodic. For unmanaged networks, congestion can be experienced during certain hours of the day. For managed network, where the bandwidth is guaranteed, changes in the network path can cause an overall change in the packet loss rate. Therefore, based on the type of network, and with system monitoring and tracking of network statistics, certain automated actions can be taken to combat changes in the type of network packet loss. For example, for congestive type networks, one possible action would be to reduce the bandwidth load to the network by changing stream's encoding rate and reducing the egress packet rate by packing more encoded frames per packet. For guaranteed bandwidth type networks, a reduced encoding rate can be coupled with an increase level of FEC to mitigate packet loss. If redundant streams are used, then time diversity factor can be changed based on the calculated burst loss parameters.

For dynamic adaptation to provide benefit, the audio IP architecture needs to be able to apply configuration changes in a seamless manner similar to what is being done with HTTP based video streaming services such as HTTP Live Streaming (HLS) and Dynamic Adaptive Streaming over HTTP (DASH).

SUMMARY

Migration from fixed circuit based telecommunication services to IP based connections provides reduction in operational expenses as well provides flexibility in audio networking. However, the reliability and quality of IP connections can deter users from making this migration. A robust audio streaming over IP architecture includes elements such as FEC, interleaving, stream grouping, support for multiple IP networks, and dynamic and automatic network adaptation. These elements, if utilized in a systematic and intelligent manner can greatly improve the performance of audio streaming over impaired IP networks. Test results using both random and burst packet loss type IP networks show audio streaming performance improvements over a variety of impaired network environments.

REFERENCES

[1] S. Salsano, F. Ludovici, A. Ordone. 2012. Definition of a General and Intuitive Loss Model for Packet Networks and its Implementation in the Netem Module in the Linux Kernel.

[2] VoIP Troubleshooter http://www.voiptroubleshooter.com/indepth/burstloss.html

[3] RFC 2733, An RTP Payload Format for Generic Forward Error Correction", December 1999

[4] RFC 3611, RTP Control Protocol Extended Reports (RTCP XR)", November 2003

[5] RFC 5109, RTP Payload Format for Generic Forward Error Correction, December 2007

AUTHOR INFORMATION

Keyur Parikh is an Architect and Software Lead with Harris Broadcast in Mason, Ohio. Mr. Parikh has over 22 years of experience in design and development of communication systems for various applications. His current interests include architecture and design of system to reliably transport media over packet based networks. Mr. Parikh holds a BS in electrical engineering with a Master's in communication theory.

Junius Kim is a Hardware Engineer with Harris Broadcast in Mason, Ohio. Mr. Kim was a key member of the Harris design team responsible for creating the SynchroCast simulcasting system and IP Link, a next generation IP audio codec. His current interests include the architecture and design of robust packet switched based telecommunication systems. Mr. Kim holds a BS and MS in electrical engineering.

ACKNOWLEDGEMENT

The authors would like to thank Jacob Schreiver and Sudhesh Sudhakaran Nair of Harris Broadcast for their contribution to this research.

GATESAIR®