Abstract

For the past 60 years, most professional broadcast signals were transported using dedicated coax infrastructures. These infrastructures are simple to set up and they “just work.” System designers never had to worry about factors such as traffic and Quality of Service (QoS). To reduce cost and increase flexibility, more and more of these real-time audio and video signals are now being transported over IP networks. These infrastructures are not simple to set up, and they do not “just work.” System designers have several options, and there are gaps between what theoretically should work and what practically does work. This paper will examine design options including Layer 2 with VLAN vs. Layer 3, MPLS, managed networks such as SONET vs. the Internet cloud, dark fiber vs. managed optical networks, transport protocols, FEC, QoS, network evaluation parameters, standards vs. propriety protocols and failure recovery.

Introduction

Dedicated coaxial infrastructures for professional broadcast signals have been around for the last 60 years. These are fairly simple to set up and use. Each signal has a separate piece of coaxial, and routing is simply selecting a source and destination pair on a crosspoint router. For the most part, one signal does not impact another signal, and signal performance does not vary over time. Video bits arrive in the order they were sent. When errors occur, they occur as a small number of bits at a time. The propagation delay in getting from the source to the destination does not vary over time. From a technical point of view, dedicated coax infrastructures are good. From an operational point of view, they are not very flexible. From a business point of view, they are becoming less cost effective compared to using common IT equipment.

IP networks are converged networks that allow different types of data and users to share the network. This is both their advantage and disadvantage. From a business point of view, using IP networks for professional broadcast signals has the potential for cost savings. From an operational point of view, the IP networks are more flexible. Unfortunately, from a technical point of view, using IP networks creates several challenges that must be overcome. These challenges arise from using a packet-switching network with blocking switches and include the following: one signal impacting another; data being received out of order; switches getting congested and packets or bursts of packets being lost; and, lastly, because other signals impact your signal, a system may be working for weeks and then just stop working or develop high jitter.

The high-level technical challenges with moving realtime professional broadcast signals over IP networks can be grouped into the following: how to encapsulate the video data, how to deal with data corruption and loss, how to get video data across the network, how to protect the data from unwanted eyes and, finally, how or if the signals should be compressed.

The particulars of your application will impact the solutions. Some questions to consider are: Is your system point-to-point or it is multicast? Is there a back channel from the receiver to the transmitter? Are the network connections relatively static or do they change often? What level of control over the network do you have? Can you directly manage the network, or is it done with Service Level Agreements (SLAs)? What is the error rate of the network? And, what are the bandwidth costs for the data and overhead required?

This paper examines solutions to the above technical challenges with consideration to the items that are specific to different applications. The focus is on realtime professional broadcast or contribution signals over IP and Ethernet networks. These signals are characterized as having high-quality video requirements and high bit rate because of the quality requirements. These signals are high-value assets. The networks may be point-to-point or point-to-few destinations. The term “video” will be used throughout the paper to include video, audio and metadata signals.

Encapsulating The Video Data

For video signals to pass through an IP network, they must be divided into IP packets and have additional information added. This information is used to route the packet, detect corrupted packets, regenerate corrupted packets, and reorder the packets into the original order. This information is added in layers so different lowerlevel protocol and networks can be used.
For compressed signals, MPEG-2 Transport Streams (TS) are a popular way to transport the data. With this method, the data is already divided in 188 bytes or octets packets. For uncompressed video signals, the continuous stream of data needs to be divided into chunks. A typical size is 1376 bytes.

One standard way to encapsulate compressed video data is with the use of SMPTE 2022-2 “Unidirectional Transport of Constant Bit Rate MPEG-2 Transport Streams on IP Networks.” SMPTE is working on additional standards such as SMPTE 2022-6 “High Bit Rate Media Transport over IP Networks” for uncompressed video. SMPTE 2022-2 encapsulates the video data in layers. Up to 7 TS packets are combined into a Real-Time-Protocol (RTP) packet. Using 1316 (7x188) bytes packets improves the transmission efficiency over single 188-byte packets. The RTP packet is encapsulated into a User Datagram Protocol (UDP) packet. The UDP packet is encapsulated into an IP packet. The IP packet is encapsulated into an Ethernet packet.

RTP defines a standard packet for transporting video and audio across an IP network. RTP has sequence numbers so out-of-order packets can be reordered. UDP is a lightweight protocol, and unlike Transmission Control Protocol (TCP), it does not guarantee packet order, or prevent duplicated packets received or missing packets. UDP includes a port number, as well as a simple check field so corrupted packets can be discarded. IP includes the source and destination address so the packets can be routed from the transmitter to the receiver. It also includes information to Layer 3 QoS. Ethernet includes source and destination information so the packet can be routed at the physical layer. It also includes a CRC check field so corrupted packets can be discarded. Each of these layers adds overhead to the packet. In total, they add up to 82 bytes. For a 1316-byte payload, this is 6.2 percent overhead. This overhead percentage could be up to 44 percent (seven times larger) if a single 188-byte TS packet is used as the payload. To maximize bandwidth efficiency, the largest data payload should be used.

Dealing With Packet Loss

As the data passes from the source to the destination across the network, data can be corrupted or lost. This can be caused by many factors including random noise on the links, packet loss from congestion in the network switches, cuts to the links and equipment failure. At the receiver, most of the lost data can be recovered; however, not all of it can always be recovered. In recovering the lost data, you need ways to detect corrupted data, detect lost packets and recover the lost data.

For detecting corrupted data, the Ethernet and UDP layers have check fields, and the packets will be discarded if they are corrupted. Therefore, corrupted packets become missing packets. For detecting missing packets, the RTP layer has sequence numbers that can be used. For recovering lost packets, three options are available: 1) do nothing, 2) add Forward Error Correction (FEC) information so the missing information can be regenerated, or 3) retransmit the lost packet.

Doing nothing will cause a medium-sized disruption in the video. Unlike baseband video, where the minimum error is a few bits, the minimum error size with IP networks is one packet. Losing a packet will cause around 1300 bytes of video data to be lost. The impact is larger if the video data is compressed and can cause a few seconds of disruption. Because not all errors can always be recovered by the receiver, the receiver needs to handle this situation.

A second option is to utilize FEC. One generic FEC method is SMPTE 2022-1 “Forward Error Correction for Real-Time Video/Audio Transport Over IP Networks.” SMPTE 2022-1 structures the packets into a matrix and adds one FEC packet per column and, optionally, another FEC packet per row. Using this additional data, a receiver can regenerate missing packets. The number of rows and columns chosen changes the amount of overhead bandwidth and the amount of missing packets that can be recovered. Some downsides to using FEC are additional overhead, additional complexity to the transmitters and receivers and more latency. Typical overheads are 5-20 percent. Using FEC is good for multicast and “send and forget” applications when a reverse channel is not available, or for lower latency applications. FEC adds a fixed and deterministic overhead, whereas retransmission is not deterministic.

Figure 2 illustrates how the SMPTE 2022-1 FEC matrix is constructed and the FEC packets are created. The FEC packets are created by XORing the data in the columns and rows. This is creating a parity packet. The FEC packets are encapsulated in RTP and then UDP. The FEC packets use a different port number than the video data packets. As long as a row or column has only one missing packet, the missing packet can be regenerated by XORing the received packets and FEC packets. Depending on the processing capabilities of the receiver, this process can be applied iteratively so that additional missing packets can be received. By using a matrix approach, SMPTE 2022-1 can recover both single errors, as well as bursts of lost packets. In the
example in Figure 2, packet #2 can be regenerated by XORing the received packets in Row #0 with the Row #0 FEC. Packets (L+1) to (2L-1) can be regenerated by repeating the same process on columns 1 to L. The number of packets that can be recovered, percentage overhead, added latency and receiver buffer size all depend on the number of rows and columns in the matrix. Signals that are encapsulated with SMPTE 2022-2 can have SMPTE 2022-1 FEC added. SMPTE 2022-2 limits the matrix size to 1-20 columns and 4-20 rows and the # rows × # columns ≤ 100. This matrix size is relatively small, so the maximum burst length that can be recovered is also small. In Figure 2, if packet #1 was also missing, then packets 1, 2, (L+1) and (L+2) could not be recovered since there are two packets missing in each of the affected rows and columns. SMPTE 2022-1 is good at recovering random errors and small burst errors. Reed–Solomon (RS) codes are another popular method for creating FEC information. These are much more mathematically complex than the XOR used in SMPTE 2022-1.

A third option for recovering missing packets is to retransmit them. Retransmission could be handled by the TCP layer, however, this is not well-controlled and is not suitable for real-time professional broadcast signals. If retransmission is used, it needs to be tightly controlled. Retransmission requires a reverse channel from the receiver to the transmitter to request retransmission of the lost packets. For this method, the receiver needs large buffers. These buffers are typically one to two seconds. The receiver buffer needs to be large enough to delay the video while the retransmission request is getting to the transmitter and while the retransmitted data is getting to the receiver. Since the retransmitted packet can also get lost, the receiver buffer should be large enough to handle a few round trips. When the retransmitted packet is received, it is inserted into the correct location in the buffer. A retransmission method is best suited for point-to-point links and has higher latency over an FEC method. Additional bandwidth needs to be allocated for the retransmitted packets. There are no industry standards for the retransmission methods, so proprietary vendor equipment must be used for both the transmitter and receiver.

**Getting Video Across The Network**

An advantage (and disadvantage) of using an IP network is that it is a converged network, and so it is shared with different types of data and users. Without any network management, large file transfers could cause video traffic to be delayed to the point where packets are dropped in the network switches or the packet delay variation is greater than what the buffers in the receiver can handle. Video has higher QoS requirements compared to other network traffic. Video traffic is also more time sensitive, requiring more timely delivery, less packet loss, less packet jitter and less packet reordering compared to file transfer traffic. File transfers are not time sensitive, and as long as lost packets are eventually retransmitted, the transfer will succeed. How to get the video data across the network is the most challenging problem in using IP networks for real-time professional broadcast signals.
To get the video data across the network, the network needs ways to identify the video packets as special packets and rules on their special processing. The process for this depends on the type of network. Networks can be grouped into:

- Smaller Local Area Networks (LAN) and Campus Area Networks (CAN)
- Connected islands of LAN/CAN
- The Internet cloud

Equipment in the LAN and CAN environments is physically located in the same building, such as a TV station or small collection of nearby buildings. For the LAN/CAN case, the user has control and management of the network. In the connected LAN environment, equipment is grouped into LANs; however, the LANs are separated by a larger distance (e.g., New York station and LA station). For connected LANs, the individual LANs are controlled by the user, but the connections across the Wide Area Network (WAN) are handled by third parties. The user’s control of the connection is through SLA. The last type of network is the Internet cloud. This is often called the Wild Wild West.

For the LAN/CAN networks, IEEE 802.1Q or Virtual LAN (VLAN) tagging can be used to identify the video packets. VLAN works at Layer 2, is below the IP layer and creates virtual networks on the same physical switch. The Ethernet packet has a 4 byte field added to the Ethernet header. This field has 12 bits of VLAN Identifier (VID) or tag and a 3-bit Priority Code Point (PCP). The transmitting device adds the VLAN header. The switches in the network are configured to give special priority to these packets. In addition, VLANs are broadcast domains that can span multiple physical switches. VLANs allow cheaper Layer 2 switches to be used instead of expensive Layer 3 routers. A network engineer is required to configure the network switches to use VLANs.

Instead of using VLANs, a different approach is to over provision the network. Network congestion and latency increase with network load, so this method keeps the load on the switches low. This method does not require network engineers. From a protection switching point of view on LANs, Spanning Tree Protocol (STP) needs to be used with caution. STP can take tens of seconds to run if there has been a change to the network topology.

Some options for connecting islands are dedicated fiber, SONET/SDH using Generic Frame Procedure (GFP) or ATM, Carrier Ethernet and the Internet cloud. A dedicated fiber allows complete control of the signals. Therefore, QoS is easy to achieve using the same methods as the LAN/CAN. Protection against damage to the fiber could be done with a second fiber link following a separate physical path. This option is very expensive, not flexible and takes time to set up.

Circuit-switched technology such as SONET/SDH allows controlled point-to-point connections. SONET/SDH is a tried and true technology that provides a high QoS and is readily available. A SONET connection is leased from a Telco carrier. The user is responsible for getting the video signals into and out of the SONET frames, and the Telco is responsible for getting the SONET frame across the network. Some bandwidth options are OC-3 for 155 Mb/s, OC-12 for 622 Mb/s, OC-48 for 2.5 Gb/s and OC-192 for 10 Gb/s. There are a few options on how to transport the IP or Ethernet packets across the SONET network. These include IP and Ethernet over GFP, Packet over SONET (POS), and IP over ATM over SONET. Each of these methods has a different amount of overhead and might not be supported by all equipment and carriers. There are two versions of GFP: GFP-F for frame based and GFP-T for continuous transport. GFP is the more popular method. ITU-T G.7041/Y.1303 Section 7.1 defines how to map IP and Ethernet frames into GFP-F frames. SMPTE and the Video Services Forum (VSF) are working on a Recommended Practice (RP) for sending video over IP over SONET using GFP-F. POS is defined in RFC 2615. Using ATM over SONET has higher overhead and requires more processing in the equipment. The challenges ATM addressed are less a problem in today’s networks, and ATM is becoming a less popular option. SONET/SDH has very good failure recovery. The network will automatically detect and switch over to a backup path within 50 ms.

Multiprotocol Label Switching (MPLS) can be used to connect the LAN islands with QoS. MPLS is considered a leading connection-oriented packet protocol. In MPLS, labels are added to packets. This is similar to the VLAN ID; however, it is not tied to the link layer. These labels allow for the creation of end-to-end circuits across any type of transport medium. The packet-forwarding decisions are made based on a label, without the need to examine the packet. MPLS operates between Layer 2 (Data Link Layer) and Layer 3 (Network Layer) and is often called a Layer 2.5 protocol. The MPLS label is a 4-byte field added to the packet. 20 bits are for a label and 3 bits are a priority. For transport applications, an improvement to MPLS is MPLS Transport Profile or MPLS-TP. MPLS-TP simplifies MPLS and adds transport network features. These include Network Management Services and protection switching. This is combining the best of MPLS with the functionality of traditional transport networks such as SONET. In February 2008, the Internet Engineering Task Force (IETF) and Telecommunication Standardization Sector (ITU-T) agreed to jointly work on MPLS-TP. 50 ms protection switching to a backup path can be achieved using Bidirectional Forwarding Detection (BFD) and Fast Reroute (FRR). This is the same carrier grade performance achieved with SONET. MPLS and MPLSTP are active areas of research, development and equipment deployment.

Carrier Ethernet is Ethernet over a WAN. Carrier Ethernet services are available as a point-to-point service (E-line), multipoint-to-multipoint (E-LAN) and point-to-multipoint (E-tree). Carrier Ethernet can be implemented on the WAN as Ethernet over SONET/SDH, Ethernet over MPLS, and Ethernet over Carrier-Ethernet Transport (CET). In the Ethernet over SONET/SDH implementation — unlike the above case where the user was responsible for getting the video data into Ethernet and then into SONET frames — the user only has to get the video into Ethernet, and the carrier is responsible for getting the Ethernet into SONET. Carrier Ethernet can have a good QoS and is a newer technology that is being actively developed.

The last option is to connect through the Internet cloud. This is a very cost-effective option; however, it has the greatest QoS challenges. Two options are: work around the QoS at the transmitter and receiver, or try to manage the QoS in the network. Using very large buffers
and retransmitting lost packets, the QoS issue can be mostly worked around at the transmitter and receiver. As long as the buffers are large enough, QoS issues in the cloud can be mostly ignored, since the buffer smoothes out packet delay variation/jitter, and lost packets are resent. For this method, the user must determine or guess what the worst case scenario in the network would be. Receiver buffers are typically on the order of one to two seconds.

If your LAN/CAN are connected using a third party, the QoS for the connections is controlled using a written contract called a Service Level Agreement (SLA). It is important to ensure the QoS terms meet your applications needs. Some items to include are:

- Packet Delay Variation
- Packet Loss Ratio
- Packet Reordering Ratio
- Packet Loss Period
- Failure Recovery
- Service Availability

Protecting The Data

Your professional broadcast signals are high-value assets. Unwanted parties may be able to copy your pristine digital video. Three points where encryption can be applied are to the source video, the TS and at the network level. At the source point, there are standards for encrypting compressed video. There are no complete standards for encrypting uncompressed video. This can cause interoperability issues. In the TS, the Basic Interoperable Scrambling System (BISS) can be used.

At the network level, Internet Protocol Security (IPsec) is a protocol suite for securing IP communication. This is not specific to video and can be used with any IP traffic. Because this is a suite, not all devices will implement all the features, and there can be interoperability issues. Most modern encryption blocks use the Advanced Encryption Standard (AES). This provides good security; however, it is computationally complex. Because of this, it can be difficult or expensive to implement at high network speeds. IPsec adds more overhead to the IP packet. The amount of overhead depends on which modes of the protocols are being used.

To Compress Or Not To Compress

The available bandwidth on IP networks has been growing by leaps and bounds. It is now possible to transport uncompressed video over the network. However, uncompressed video chews up a lot of bandwidth and can be costly. With Gigabit Ethernet, you can only transport three SD signals. Going up to 10 GE, you can only get six uncompressed HD signals. The decision to compress or not to compress comes down to a question of video quality versus bandwidth and latency. Depending on the image quality requirements of the application, some level of compression will reduce bandwidth requirements and cost. Three levels of compression are mathematically lossless, visually lossless, and visually lossy. Mathematically lossless such as some modes of JPEG-2000 have perfect image quality, but low compression ratios. The visually lossless is the level used by most applications. A design consideration with compression is where and when to compress and decompress. Each time a signal is compressed and decompressed, artifacts are added and there is processing latency. The artifacts can be magnified if different compression techniques are used on each compressed-decompressed pass. To reduce the artifacts from multiple compression-decompressed passes, 4:2:2 compression should be used over 4:2:0. Another design consideration is what to do with the HANC and VANC spaces. Some options are to pass these uncompressed, which is transparent but high bandwidth; strip out only the data of interest, such as embedded audio, and transport this
in a compressed or uncompressed format; use SMPTE 2038 to pass only the ancillary data packets; or not to transport the HANC and VANC data. A last design consideration is how the decoder behaves with errors. Even with FEC, errors will eventually occur. The amount of visual error concealment and if video audio delay changes (lip sync) happen depends on the decoder implementation.

<table>
<thead>
<tr>
<th>Bandwidth (Mb/s)</th>
<th>Compress Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>H264, 4:2:2, 10-bit</td>
</tr>
<tr>
<td>50</td>
<td>MPEG-2, 4:2:2, 8-bit</td>
</tr>
<tr>
<td>200</td>
<td>JPEG-2000, 4:2:2, 10-bit</td>
</tr>
<tr>
<td>1500</td>
<td>Uncompressed</td>
</tr>
</tbody>
</table>

**Conclusion**

IP networks are becoming a cost-effective way to transport real-time professional broadcast signals. In the future, they will become even more cost effective. IP networks do have several technical challenges that must be overcome to operate correctly. There are several different options for addressing these challenges, and the best solution depends on the particular details of your specific application. Using solutions based on open industry standards helps reduce interoperability issues and increase flexibility in vendor selection. The SMPTE 2022 family of standards provides ways to encapsulate both compressed and uncompressed video and add FEC. FEC allows you to recover data that is lost in the network. Real-time professional broadcast signals require special processing in the network to ensure they will be usable when they reach the receiver. VLANs allow you to get QoS on your LAN/CAN. SONET/SDH is one way to connect your LAN islands across a WAN and have QoS. Ethernet over GFP-F is one way to transport IP over SONET. MPLS-TP is becoming another good way to connect your LAN islands. IPsec allows you to protect your video assets. Compared to traditional, dedicated broadcast infrastructure, transporting real-time professional broadcast signals requires different kinds of equipment. In addition, different skills, such as network engineering, are required to set up and maintain these.

This paper was originally published in the 2011 NAB Broadcast Engineering Conference Proceedings.