Advances in AoIP Connectivity for Extending the Radio Operation Beyond the Studio

Andrew Calvanese
Wheatstone Corporation
New Bern, North Carolina USA

Abstract - *Audio-over-IP, or AoIP, technology is common in broadcast studio operations today as a way to route and control audio. The maturation of AoIP technology and techniques along with emerging new standards make it possible to extend the reach of studio operations remotely across a cluster or region. By hitching real-time audio onto the same IP network technology that has spread ubiquitously across our desktops, homes, workplaces, and even our phones and automobiles, broadcasters gain unprecedented flexibility in sourcing and producing audio content and distributing signals around the cluster and ultimately out to the world. This paper examines QoS, transport, protocols and other issues relevant to extending AoIP networks for sharing resources and controlling studios across a wider area network.*

AoIP Background

Few areas on the planet today lack cat5 cable, Ethernet jacks or cell phone coverage. The raw capability exists for extending AoIP connectivity beyond the studio. But in order to move an audio signal from one location to the next, we must address three separate elements of AoIP in order to be effective. These are: transport, discovery, and control.

To illustrate the role of these elements on the audio signal, consider a simple analog microphone in one room connected to a loudspeaker in another.

Transport. Analog voltage from the microphone is distributed over a shielded twisted pair cable. To be effective, the cable must be the correct type and must be run intelligently to avoid interference or the audio will suffer. A hardware device (preamp) is usually necessary to increase the signal voltage to a usable amount for the loudspeaker, which needs its own amplifier. The cable infrastructure, patch panels, and I/O device hardware constitute the transport mechanism for the mic signal.

Discovery. The microphone won't be useable to the loudspeaker unless it has correct labeling giving instructions on how to plug it in, patch it through, pot it up, and so forth.

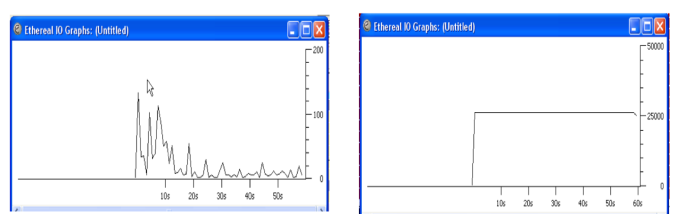
Control. Without some way to turn it on or off and adjust its level, the microphone isn't very useable. Control is the means by which users interact with the microphone to get the audio they want when and where they want it. So it is, too, with AoIP systems. Let's take a closer look at these three elements as they relate to the AoIP system.

Transport

AoIP systems fundamentally work by using IP networks (Ethernet cables, switches, and routers) to transport audio signals to and from audio I/O devices. It's important to note that the network, switches, and routers don't "know" that they are transporting audio. To the network, audio data is invisible. Audio is digitized at the I/O devices and converted to data bits, which are then added as data payload to IP packets; these the network understands. The various ways of embedding and extracting the audio data from the packet payload is one difference between the various AoIP systems available; we'll discuss this in more detail a little later in this paper.

The IP network treats audio packets like any other packet. It "looks" at the packet headers and pushes the packets to their destinations using the rules of Ethernet and IP, not audio. The rules of IP packet distribution are not at all friendly to real-time audio. Consequently, AoIP systems must work around these rules with tools like buffering and QoS to assure seamless audio transport. Complicating the issue is that even a few audio channels generate significantly more packet traffic than the network normally sees, as shown in our examples, below. [1]

Fig.1 Affects of audio channels on the network



Simple network, 18 PCs 144 packets/sec peak

Same network, add 3 AoIP channels, 25,000 packets/sec steady

For this reason, the AoIP transport mechanism must meet the audio requirement. IP networks come in a variety of sizes and capabilities, from modern high-performance layer 3 Gigabit switches at one end of the spectrum and the common internet at the other. A key variable is bandwidth, or how much data capacity the network has. This must be matched to the audio requirements, or AoIP networks simply won't work. It's easy to see the problem if you think of a

DSL modem connecting to the internet with its 1MB/sec upload speed (you hope) and a single uncompressed audio channel with 2MB/sec of data. This is a combination that won't work.

A modern broadcast facility can have hundreds of different audio sources and destinations with anywhere from a few dozen to hundreds of them actually playing at any one time. The bandwidth requirements of each section of an AoIP system and the audio requirements of the facility must be in sync for a system to work successfully. At the current state-of-the-art, reasonably priced Ethernet switches are available that will support dozens of audio channels for inter-studio transport. In addition, higher end switches are now available that can link all the studios together and high-speed WAN connections are available for a price to link campus facilities together. With some audio compression and the common internet, IP links can be used to inexpensively link a few channels of STL and remotes. So, too, is it with RF IP links.

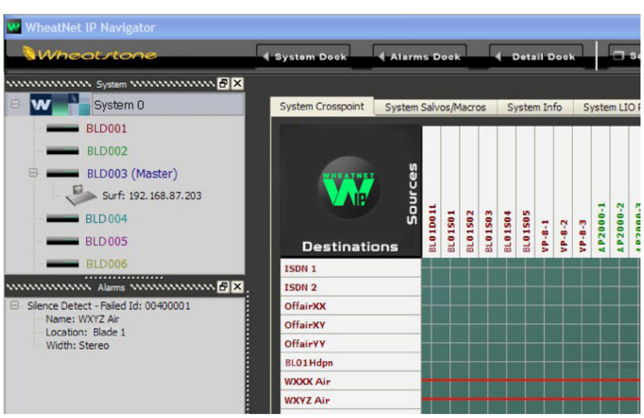
Discovery

AoIP networks need to be able to handle hundreds of audio channels, but they also need some way of identifying and announcing channels to be useful. This is the discovery element. Wheatstone, Axia and other AoIP system manufacturers have taken a common broadcast approach to discovery and label components in the AoIP network similarly. AoIP manufacturers designate extra packets on the network to communicate discovery data and display it seamlessly to all users with signal names and other information easily created and recognizable to broadcasters.

Control

Gaining access to hundreds of channels of audio on a network is useless if you can't route them, turn them on or off, fire their playback, or turn an ON AIR light on when needed. To accomplish this, broadcast AoIP manufacturers take of the network and use packets to communicate command and control. Sometimes an ancillary PC is used for this and sometimes the intelligence is built right into the network devices.

FIG. 2 AOIP NAVIGATION AND CONTROL



BANDWIDTH AFFECTS ON REAL-TIME AUDIO

The biggest problem with transporting real-time audio over IP networks has to do with timing and synchronization. As previously mentioned, IP networks distribute packets by the rules of Ethernet and IP, which, by their very definition are non-deterministic. Packets are routed based upon the moment-by-moment condition of the network traffic and its switches and routers, and not necessarily which packets were created first. While not a significant issue for a very small system, this can be detrimental as the number of packets goes up and traffic increases, causing the packets to get jumbled and delayed. As we've seen in Fig. 1, a few channels of streaming audio generate a huge amount of network traffic. Compounding this problem can be the method of packet distribution chosen in the first place.

Packet Distribution Protocols

There are three packet distribution choices available in the IP protocol: point -to-point (TCP/unicast), broadcast (UDP) and point-to-multipoint (UDP/multicast). TCP/unicast offers a direct source-to-destination path with built-in acknowledgment and retry mechanisms to assure the packets get to their destinations. While this sounds good on paper and, in fact, works very well for a simple source-to-destination system like a CODEC, TCP/unicast breaks down in real-world systems with multiple devices. There are several factors why this is so. If, for instance, you want to send the same audio to multiple places, TCP transmitters will duplicate packets for each destination and process duplicate acknowledgements. Furthermore, if a packet does not get through to the destination and isn't acknowledged in time, the transmitter will send the packet again. If this happens too often, the network will slow down transfer speed so the packet can make it through the network. This effect is like thermal runaway in a transistor and wrecks havoc on streaming audio. As a network becomes more congested, TCP sends duplicate packets of the same data, further clogging up the network and ultimately slowing it to a halt. Fundamentally, TCP emphasizes reliability over timeliness.

UDP broadcast takes the opposite approach and indiscriminately sends packets to every destination. UDP is very efficient from a transmission perspective because it sends only one copy of the packet to the switch. However, the network switch passes on duplicates of the packets to every connected device, even those that couldn't possibly use them, which makes for a lot of unnecessary traffic on the network.

The third choice, UDP /multicast, is much more efficient in terms of distribution. The transmitting device simply streams its packets to the Ethernet switches, which then pass them on to only those elements registered for specific packets. With UDP/multicast, there's no duplication of packets in transmission, no needless duplication by the switches, no acknowledgement packets needed, and no retries or throttling required. What you give up is absolute acknowledgment that the packet got through to the destination. UDP emphasizes efficiency over reliability.

So while TCP gives us secure packet delivery at the expense of greatly reduced network efficiency, UDP/multicast gives us network efficiency but the possibility of missing packets.

In real-world multichannel AoIP systems, both protocols are typically used: TCP for system, command, and control packets in which a missing packet would be disastrous; and UDP for streaming audio data packets where a missed packet is less significant and for which it can be corrected. The problem then becomes how to manage the indeterminate nature of packet distribution in IP networks with limits on bandwidth while still keeping our audio streams flowing.

The solutions vary among AoIP manufacturers but are really variations on two themes:

1) Specifically identify the individual packets so that their playback order can be maintained.

2) Provide a way to synchronize all of the devices on the network so that the audio playout sample rate and the audio sample creation rate are precisely equal, therefore eliminating dropouts and clicks due to over- or under-sampling.

Packet Order Protocols

Engineers in different applications long ago realized that some mechanism for recreating the proper packet order in an IP network would be necessary; hence they created additional protocols to add more information to IP packet headers. Of these are RTP (Real-time Transport Protocol) and RTCP (Real-time Transport Control Protocol), which together provide sequence numbers and time stamping and prioritization (QoS) to the packets at a small increase in packet overhead. Wheatstone, Axia, the AES X192 group, Ravenna, and AVB all use RTP [2]. In this respect, most of the popular AoIP systems are similar. However, they differ in the specific packet loading, timing and synchronization mechanisms within the protocols. RTP provides identification in the packets about their creation time and order but it is up to the AoIP system manufacturer to extract this information and to recreate the audio data and timing. Axia and Wheatstone have proprietary solutions in use for years; AES X192 is adapting the new PTv2P (Precision Time Protocol - IEEE 1588-2008) standard as the time reference.

Latency and Transport

Latency is the amount of time that elapses from when an audio signal is first created to when the digitized, packetized and transported signal is recreated. Too much latency is undesirable, as anyone who has talked on a phone circuit with excessive delay can attest. There is intrinsic latency involved in any digital system, even non-networked ones that we won't get into here. This is due to the built-in latency of A/D and D/A converters and sample rate converters, for example. Beyond that, extra latency is introduced by the packetizing, routing, and reassembly process used in AoIP systems. Here again, the trade off between network transport efficiency and performance comes into play. From our earlier discussion you will recall that audio data exists in the AoIP system as a data payload enclosed in standard IP packets. There is actually quite a lot of data bits used by the packet protocols themselves, normally 20 bytes per packet. [3]. With this fixed amount of overhead, if we minimize the number of packets on our network by making the audio payload in each packet large, network efficiency goes up because there are fewer packets to route and the majority of the data in a packet is payload. On the other hand, if we put a smaller audio payload in each packet, then it will take more packets to send the equivalent amount of audio. Transport efficiency goes down because the majority of the data in a packet is now protocol overhead.

Let's look at some extremes. Audio at 48kHz sample rate is equivalent to 1/48,000 second of audio. If our packets have one sample of stereo audio in them (26 bytes total; 20 bytes overhead, 6 bytes data), we would have to send them 48,000 times a second or one every .000020 seconds to maintain audio without interruption. Ignoring other factors, we would get our first audio sample .000020 seconds after we started. If on the other hand we put 100 audio samples in each packet, we would have to send them only 480 times a second or one every .0020 seconds. We would get our first audio sample .002 seconds after we started. Add to this digitizing latency and the buffering required because of the indeterminate nature of network traffic (you need enough audio samples buffered up to continue to play out good samples while you wait for the erratic arrival of new packets), and latency can become a serious issue. So audio payload size becomes a compromise between network efficiency and latency. This is especially so with 100baseT networks and the common internet because of the restricted bandwidth available, which is one reason why some manufactures require you to make that choice on latency when setting up the AoIP system. Audio payload size is one of the issues being addressed by the various standards.

EXTENDING THE AOIP NETWORK

Extending the reach of the AoIP network takes into consideration transport, discovery and control issues

Following is a diagram of what a network might look like:

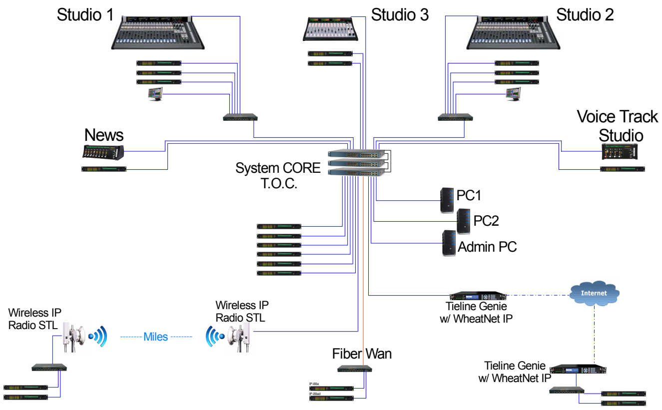


FIG. 3 DIAGRAM OF A TYPICAL EXTENDED AOIP NETWORK

Transport considerations

The extended AoIP network can be thought of as rings of diminishing bandwidth. At the center is the high bandwidth core of the system. This is where the highest traffic is located, being the central hub of your campus or cluster through which all traffic destined for other parts of the system must flow. Here is where the largest, highest capacity Ethernet switches are used and the maximum system capability (as limited by those switches) is determined. Notice the high-capacity switches at the core in Fig. 3. (Cisco 3750, 160Gbps, 1000 IGMP, $5000 or Cisco 6500, 2080Gbps, 256,000 IGMP, $40,000+) [4], [5].

The next level out has smaller switches, one for each studio or area. These handle traffic within the studio and by so doing, reduce the traffic requirements on the central core. Note the examples in Fig. 3. (Cisco 2960, 32Gbps, 256 IGMP, $750) [6].

The outer ring is the low bandwidth ring. Here is where you are reaching out with your network beyond the walls of your cluster. You might have a dedicated WAN connection to another facility, or an RF STL, a satellite connection, T1's, ISDNs, 4G, or the common internet. Typically in this area, your connections cannot support enough bandwidth for even a single channel of audio so you must resort to data compression and CODECs. Because of low bandwidth, this is where you will likely experience the most latency. And here is where the transport mechanism of the AoIP network must change to work within the constricted bandwidth. The available bandwidth is so limited and traffic can be so congested that mechanisms for packet timing and audio reconstruction that work so well within a restricted LAN break down and won't work. Similarly, security concerns come into play, as you won't want to directly expose your network to the outside world. Unless the extended network connections are entirely under your control (such as in a protected WAN), the transport mechanisms shift to TCP, large packet size, and audio compression. Fig.3 shows these changes.

Regarding discovery, connections between the inner core and the studio ring are easily identified, as long as you are using one of the AoIP systems from a broadcast manufacturer. These systems are designed as complete interoperable systems in themselves. Moving to the low latency ring is where discovery can break down. Here is where your AoIP system can lose its discovery abilities because the various CODECs, modems and interfaces may not be manufactured by your AoIP system provider and thus the system doesn't know how to communicate discovery with these elements. The current state-of-the-art is for manufacturers of these third ring devices (CODECs, modems, and so forth) to partner with the AoIP system via drivers or interfaces, or you can provide this discovery element yourself by wiring analog or digital audio signals to AoIP I/O devices to get them into the network. Here is where more work on interoperability needs to be done.

In a similar way, those system partners that have installed drivers or interfaces for discovery will usually include control functions. As a last resort AoIP logic control devices (from the AoIP system provider) can be wired to non-partnered devices, and, clearly, this is a prime area for interoperability work.

AOIP INTEROPERABILITY STANDARDS

This brings us to a discussion on interoperability. The transport, discovery, and control aspects of a system must all work together to truly meet the needs of a broadcast facility. Broadcast manufacturers design products for an ever-changing environment, and the more flexibly and transparently we can do this, the more efficient the process becomes. In this regard, the broadcast audio industry has gone much further than the rest of the audio industry. Nowhere else in the audio industry do you find IP networks with hundreds or even thousands of audio channels routing with dozens of mixing consoles working concurrently.

Nowhere else in the audio industry are processors crunching audio or playback systems streaming and CODECs compressing, all under the management and control of one system user interface. Broadcast AoIP manufacturers have developed extensive interoperability over the years, while the rest of the audio industry is waking up to the advantages and just now talking about interoperability.

Transport and Interoperability

For transport within the LAN environment, the IP protocol is the hands-down favorite of the broadcast industry because of its ease of routing within standard networks. Axia, Wheatstone, and Ravenna, as well as others in commercial audio, all use the IP protocol as the basic transport mechanism and use RTP/RTCP or similar protocols for packet sequence control. Various timing mechanisms are used for synchronization. The Audio Engineering Society has formed the X192 standards working group to come up with a common set of specifications for the use of these protocols (packet sizes, sample rates, identification, and timing) that will allow various audio devices to transport audio packets between them and reconstruct the audio signal correctly. Broadcast AoIP system providers are all members of the group and are working to insure that the standard is defined in ways suitable to broadcasters (specifically, performance versus efficiency compromises). This group is close to finalizing their standard.

The IEEE 802.1 standards committee has produced the AVB (Audio Video Bridging) standard for 802.3 and 802.11 links and includes both audio and video. This standard is targeted at home AV and commercial AV applications and requires specific AVB compliant hardware devices at every point in the system, including Ethernet switches and routers and PC NIC cards. As such, it is therefore not compatible with existing network infrastructure and must be built from the ground up. Very few devices are currently AVB compliant. Comparatively, X192 is compatible with current installed network protocols and closest to what current AoIP systems employ.

Discovery and Interoperability

Earlier work on the interoperability aspects of AoIP system discovery was done by the X192 group, although subsequent work has been tabled.

In contrast, AVB requires a specific discovery mechanism in order for the transport mechanism to work (transmitters must announce their requirements to the AVB specific network hardware and be authorized by the hardware before transport can happen); therefore discovery at the most basic level is built into AVB. However, the standard does not provide for discovery and identification at the user level; that is up to the individual manufacturers.

Control and Interoperability

AES X192 makes no mention of control. Interoperability in the control domain is left out of the picture. Another task group at the AES has just released a standard for control, known as AES64; its application to broadcast AoIP systems is yet to be determined. Like x192, AVB is a transport specific standard and does not deal with control.

In Conclusion

Modern IP network infrastructure and AoIP technology have evolved to the point where whole studio clusters, including STL and multiple remote connections, can all be efficiently integrated into one seamless interoperable system with instant routing and control flexibility. From playout systems with AoIP drivers to Ethernet switches, with IP codecs and the common internet replacing expensive leased lines, the options are greater and costs have never been lower.

It's important to remember that the broadcast audio industry is not the engine driving the standards organizations; we're a very small voice in a large group of related audio industries. Of the 30-plus "members" of the AVnu Alliance (the AVB promotion group) not one is a broadcast manufacturer [7]. Broadcast AoIP system providers are participating in standards discussions in support of the broadcast industry while standards continue to evolve. When standards have advanced to the point where true interoperability adds benefit to what we are providing in our integrated AoIP systems today, and truly becomes a step forward, compliance for the broadcast industry will become more useful.

In the meantime, we will continue to upgrade our products, develop new ones and encourage our broadcast equipment partners to connect with us to provide the highest level of AoIP interoperability deployed today.

References

[1] Calvanese, A., "IP Networks: Planning for Reliability," *Wheatstone*, May 2010, pp 2,4

[2] Hildebrand, A., "High-Performance Streaming Audio-Over-IP Interoperability", Sept. 10, 2011 AES standards task force, pp 2

[3] Internet Engineering Task Force, RFC1122, October 1989, pp 84

[4] Cisco Catalyst 3750 data sheet

[5] Cisco Catalyst 6500 data sheet

[6] Cisco Catalyst 2960 data sheet

[7] AVnu Alliance, IEEE 802.1/Audio Video Bridging (AVB)