HOW AES-67, THE NEW AUDIO-OVER-IP STANDARD, WILL BRING THE CONVERGENCE OF TELECOMMUNICATIONS, STUDIO AUDIO, AND INTERCOM

G.F. Shay and M.J. Dyster

The Telos Alliance, USA

ABSTRACT

AES67 is a new industry standard for interoperability of high quality audio over IP networks from the Audio Engineering Society, published just under two years ago (September 2013). This standard was quickly embraced by all of the main broadcast audio equipment vendors, and compatibility modes announced by all of the major competing networking audio vendors: Livewire, Q-LAN, Ravenna and Dante. Outside of broadcast, there has also been a high level of audio industry acceptance.

AES67 specifies the method for carrying uncompressed 24-bit linear audio over layer 3 IP networks. There are options and choices of sample rates, packets sizes, number of channels and bit depths, but a strict interoperability requirement is made so that all vendors must implement at least the one common set of parameter choices. This requirement is what produces the interoperability between all vendors labelling their equipment AES67.

The technical details of AES67 are readily available. This paper examines the features of the design of AES67 that enable it to be the platform for the convergence of working with audio, telecom, studio and intercom.

INTRODUCTION

At the present time, audio technology is leveraging audio over IP technology at the basic network transport level, but is not taking advantage of all the benefits that are possible. Today, the audio in telephony, studio audio and intercom use network technology rather like computers were using networking technology in the 1990s. We have managed to get the number of cables down to one, but audio applications are using many different protocols on that same one wire.

The workflows, user interfaces, and mental models of how we use voice, sound, music, effects, communication, and languages for telephony, for studio and for intercom, are separate. We think of these as somehow different, separate, and requiring unique equipment and different user interfaces and operating sequences. But these all have that one thing in common: they are *audio*. They are *sound*. There is a fundamental commonality.

LAYING OUT THE PROBLEM:

The Audio Interconnection Situation Today

Where radio has already adopted standards-based Audio-over-IP, making it a *de-facto* part of everyday life for literally thousands of users throughout the world, television has been much less willing to step up and embrace what is already an established technology, favoring the development of baseband signal transport and even proprietary Ethernet protocols. AES67 is a natural evolutionary step for TV to take in order to leverage the many advantages that network-based infrastructure provides. To gain wide acceptance for AES67, the industry has to take a look at the inefficiencies of the systems in use today and begin to understand how AoIP can replace what has gone before, and how it will enhance operational practices.

The audio system at the heart of any television broadcast facility has often been considered a necessary, yet less well regarded relative of video. Conversely, the complexity of the audio and communication systems are generally acknowledged as being inversely proportional to that of the arguably more simple vision infrastructure. In part, this is down to the sheer numbers of connections used to make audio work compared to video, but in reality, the true complexity of audio and communication has evolved in sympathy with gradually changing production workflows and the increased expectations of the program makers, engineers and technicians who create television.

For many years audio and video signals were kept apart, with separate systems used to acquire sound and vision to mix, edit and ultimately transmit. The birth of recordable video tape provided a means to store, transport and broadcast pictures and sound via a single unifying format, but linear emission of signals still relied on individual analog video and audio routing and distribution infrastructures. It took until 1989 for the first standardized version of Serial Digital Interface (SDI), in the form of SMPTE 259M, to introduce the concept of embedded audio which could be transported with the video signal along a single piece of cable. 26 years later, the majority of broadcast television ecosystems are still built around an embedded audio infrastructure.

For the many instances of audio within the system where no video signal is present, such as microphone circuits, monitoring, tie-lines, audio playback and recording devices, cues and communications, many broadcasters still rely on baseband connectivity using analog, AES3 or MADI, with the associated overhead of copper cable, fibers, connectors and patching that these forms of signal entail.

So that it can be viewed in the context of this paper, Figure 1 is a condensed representation of the typical baseband cabling that exists in a production TV studio where discrete de-embedders & embedders are used to connect to and from the audio routing and console system. Other devices and circuits shown include distribution amplifiers, local audio playback and processing devices, studio tie-lines, studio mic circuits, intercom trunks and monitoring. Also note that almost every circuit entering or leaving a hardware device does so via a patchbay, often perceived as a necessity but rarely used in a well-designed, assignable system.

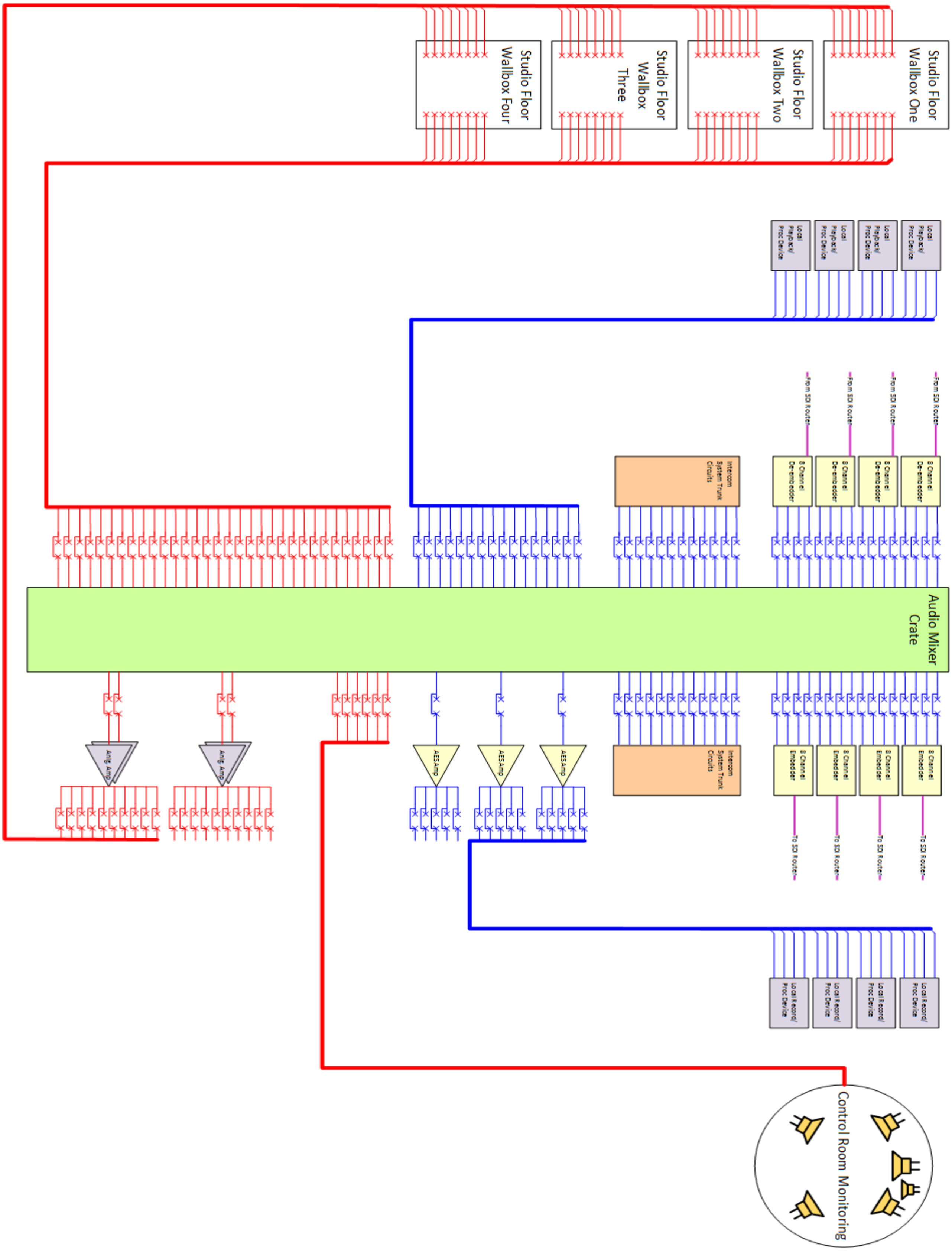
Studio Floor Wallbox One
Studio Floor Wallbox Two
Studio Floor Wallbox Three
Studio Floor Wallbox Four
Local Playback/ Proc Device
Local Playback/ Proc Device
Local Playback/ Proc Device
Local Playback/ Proc Device
From SDI Router
From SDI Router
From SDI Router
From SDI Router
8 Channel De-embedder
8 Channel De-embedder
8 Channel De-embedder
8 Channel De-embedder
Intercom System Trunk Circuits
Audio Mixer Crate
Intercom System Trunk Circuits
8 Channel Embedder
8 Channel Embedder
8 Channel Embedder
8 Channel Embedder
To SDI Router
To SDI Router
To SDI Router
To SDI Router
AES Amp
AES Amp
AES Amp
Anlg Amp
Anlg Amp
Local Record/ Proc Device
Local Record/ Proc Device
Local Record/ Proc Device
Local Record/ Proc Device
Control Room Monitoring

Figure 1

Communication within broadcast facilities has evolved quite independently of the changes in program audio technology. Borrowing more heavily from telephony, the earliest forms of dedicated intercom & talkback used the Partyline concept and many in service still do. (Partyline is a phrase that dates back to the earliest days of commercial telephony and describes a means of connecting multiple subscribers to a single common communication circuit.) Communication beyond the boundaries of the physical broadcast facility has also relied heavily on telephony for years, with many broadcasters still connected to the outside world via an armory of telephone hybrids.

IP connectivity in the form of VoIP is not new to broadcast intercom and all of the familiar manufacturers of intercom systems promote a means to trunk to remote locations using dedicated codecs.

Figure 2 is also a condensed view (for the purpose of this document) illustrating the level of complex connectivity that exists in a fairly typical TV intercom system. In this instance the system includes an Intercom Matrix shown in two parts, showing to the left the traditional ‘star’ connectivity of the panels and then the baseband connections to the remainder of the studio and remote systems. The matrix itself is used to switch bi-directional communication circuits between the intercom matrices, audio console and the bank of codecs used to connect to the outside world. A bank of POTS and ISDN legacy hybrids are used for communication to contributors (such as journalists) who call into the show and a VoIP interface that connects to affiliated studios.

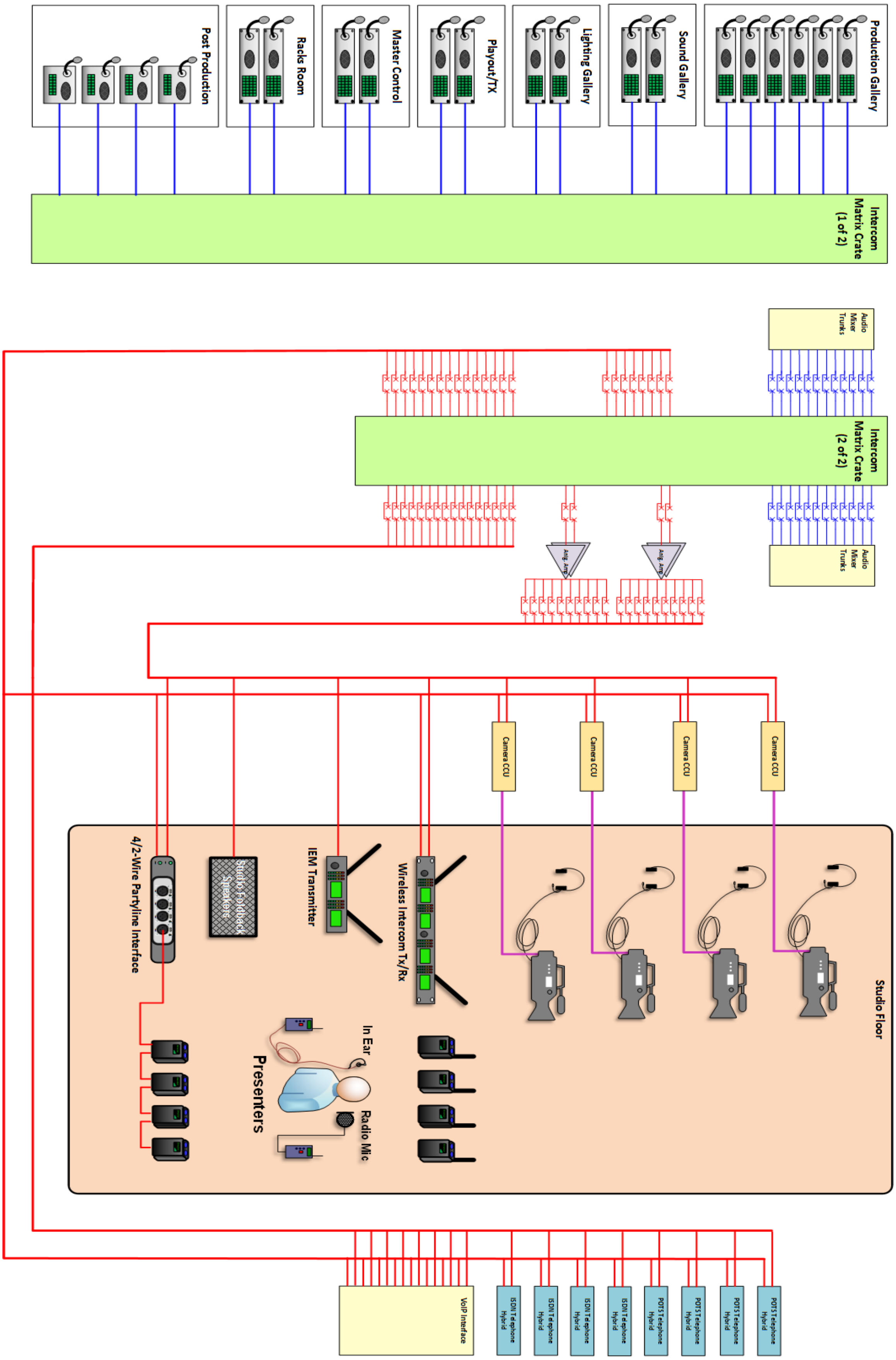
Production Gallery
Sound Gallery
Lighting Gallery
Playout/TX
Master Control
Racks Room
Post Production
Intercom Matrix Crate (1 of 2)
Audio Mixer Trunks
Intercom Matrix Crate (2 of 2)
Audio Mixer Trunks
Anlg. Amp
Anlg. Amp
Camera CCU
Camera CCU
Camera CCU
Camera CCU
Studio Floor
Wireless Intercom Tx/Rx
IEM Transmitter
Studio Foldback Speakers
4/2-Wire Partyline Interface
In Ear
Radio Mic
Presenters
VoIP Interface
ISDN Telephone Hybrid
ISDN Telephone Hybrid
ISDN Telephone Hybrid
ISDN Telephone Hybrid
POTS Telephone Hybrid
POTS Telephone Hybrid
POTS Telephone Hybrid
POTS Telephone Hybrid

Figure 2

THE VISION OF WHAT IS POSSIBLE:

A look at those same subsystems using AES67 equipped technology

We have seen how complex audio can be in today’s installations where hundreds, if not thousands of baseband connections join together a collection of subsystems to create a larger integrated system. In a world of AoIP, and in particular AES67, many of those connections and the associated physical hardware disappear completely, replaced with managed Ethernet switches and, when necessary, edge devices that convert baseband audio into IP packets. Gone too are the many audio patch bays that are still deemed a necessary part of many broadcast systems, even those where the flexibility of the routing infrastructure renders them redundant. Using IP, however, there is no demarcation point between systems and, unless blocked by deliberate configuration rules, all sources are truly available to every device or destination and patching becomes irrelevant except perhaps for ad-hoc tie-lines.

Connectivity beyond LAN becomes more easily achieved using codecs, when needed, that bridge between AoIP and VoIP with dynamic bitrates and bandwidth management enabling the user to define the protocol that maximizes the quality of the audio without risking the integrity of the signal.

The benefits of AES67 are manifold. The reduction in infrastructure costs can be measured easily by calculating the reduction in physical hardware as well as cable and connectors. The traditional router is replaced by Ethernet switches, and the same is true of the distribution equipment commonly used in abundance within production, communications and also reference systems. The question often asked during the early specification of facility routers, where the designers calculates the matrix port count, is obsolete. The network and therefore the equivalent routing infrastructure is limited by bandwidth and not physical port numbers; when you need more capacity you can add another switch and manage your network accordingly.

Tie-line routes between different areas of the facility that are used for ad-hoc connectivity can be served using inexpensive end-point devices plugged into local network points as and when necessary, scaled to suit the requirement.

Figure 3 combines the facilities represented in the previous baseband dominated systems and replaces almost all of the connectivity with either AES67 or VoIP network connectivity. The remote elements connect to the local network via a managed gateway which might take the form of a PBX, dedicated VoIP communication interface or in the future a direct high quality WAN (wide area network) connection. Patchbays have disappeared completely, there are no matrices or traditional routers, all replaced by Ethernet switches on which all available sources and destinations share connection. In this scenario, any end-point, contributing or receiving device can be routed to any point in the local or extended network.

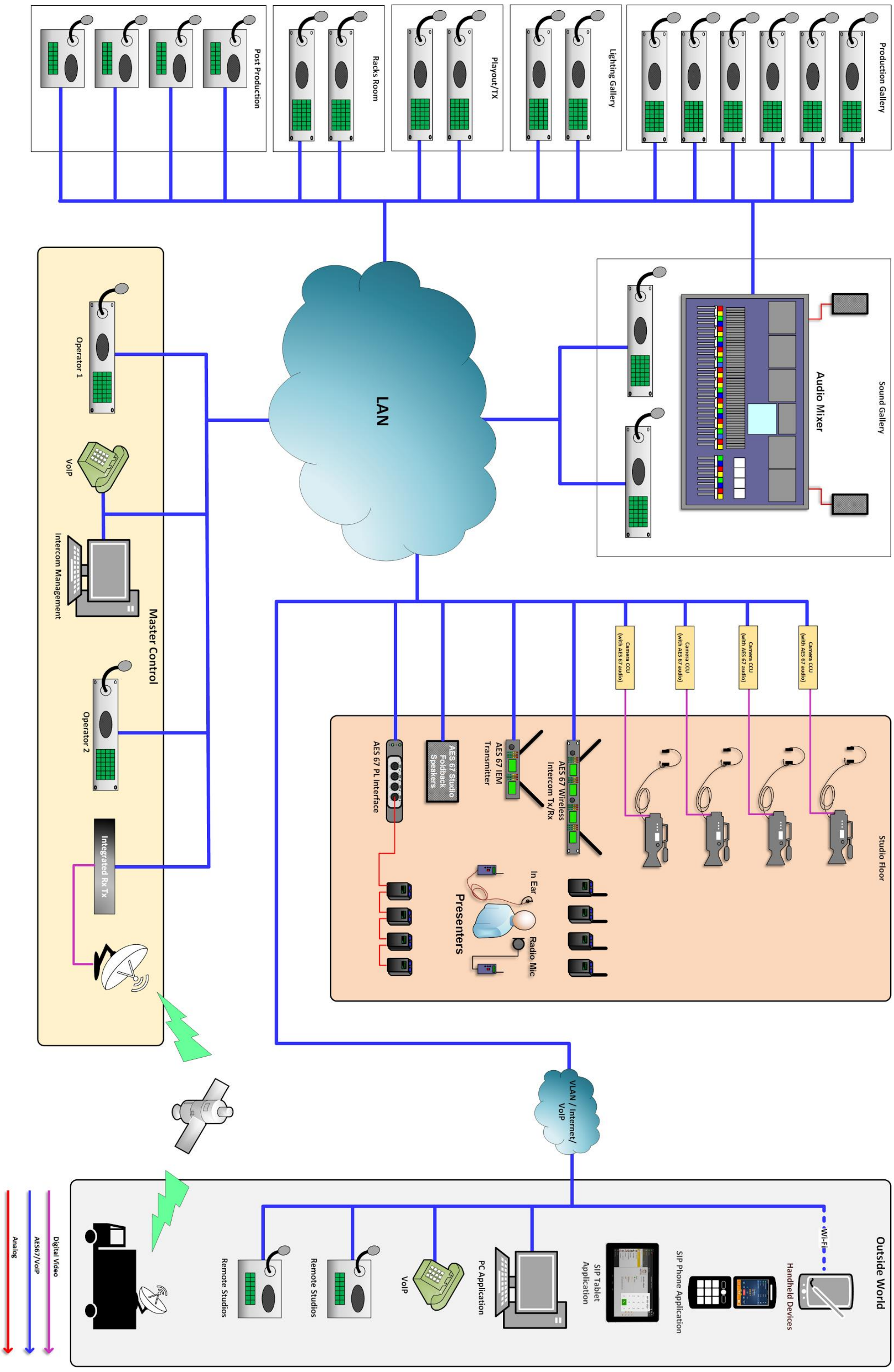
Post Production
Racks Room
Playout/TX
Lighting Gallery
Production Gallery
LAN
Sound Gallery
Audio Mixer
Master Control
Operator 1
VoIP
Intercom Management
Operator 2
Integrated Rx Tx
Camera CCU (with AES 67 audio)
Camera CCU (with AES 67 audio)
Camera CCU (with AES 67 audio)
Camera CCU (with AES 67 audio)
Studio Floor
AES 67 PL Interface
AES 67 Studio Foldback Speakers
AES 67 IEM Transmitter
AES 67 Wireless Intercom Tx/Rx
In Ear
Presenters
Radio Mic
VLAN / Internet/ VoIP
Outside World
Wi-Fi
Handheld Devices
SIP Phone Application
SIP Tablet Application
PC Application
VoIP
Remote Studios
Remote Studios
Digital Video
AES67/VoIP
Analog

Figure 3

LEAST COMMON DENOMINATOR TECHNOLOGIES:

Linear PCM Audio

You may agree that linear PCM audio is technically the least common denominator, as all use cases of audio can be derived from it. It is the 'superset' of audio. But what about the required bit rates and the limits of the network bandwidth? This, admittedly, is the first leap of faith the reader is asked to take. Consider, for instance, the bit rate of a typical stereo AES67 audio stream is around 3Mbps (3,000,000 bits per second). The cost to connect 3Mbps is highly dependent on which network you are connected to. On a local multi-Gigabit LAN, it is insignificantly small.

But what about the WAN? The fact is, network bandwidths grow exponentially, following Moore's Law of doubling roughly every two years. In an analysis of the consumption of bandwidth by media bitrates vs. the state of the industry in network bandwidth growth, Kevin Gross in his paper [2], points out that media bitrates have not increased that much over time. The exponential growth of network bandwidth has caught up and passed the slower growing media bit rates. In that analysis the crossover point of modern networks and modern media bit rates happened around the year 2010, and network rates are continuing to exponentially charge ahead.

What this means is that the economics which 15 years ago applied to the LAN, now also apply to the WAN. The cost of the bandwidth makes carrying high quality linear PCM audio data economical, and audio data compression economically unnecessary. Today, the high bandwidth WAN challenge is not technical. We right now are in a transition, waiting for the business models of the telecom carriers to catch up and even this is in the middle of being resolved. Private corporate broadcasters and national broadcasters (e.g. Swedish Radio) have built high speed IP network WANs of gigabits, spanning whole nations, for the simple reason of saving costs over the traditional telecom links of just a decade ago.

IP layer 3 network

In a similar sense, IP networking layer 3 is the backbone and everywhere present least common denominator of global network communications. In a way analogous to studio quality audio, IP layer 3 was for a long time at a premium in complexity and cost of interfaces. Therefore, in earlier times much more simplified choices were made for interconnecting digital audio equipment (for example the industry standards AES/EBU from 1985, and MADI AES-10 from 1991, which are simply specialized forms of high speed serial combined clock and data). But now the cost of IP layer 3 networking has become a commodity. When integrated with the core of modern digital equipment designs, the elements of the IP layer 3 interface essentially come for free.

In summary, because AES67 is leveraging the use of mainstream ubiquitously available technology, and carrying audio in a superset format, this combination is what makes AES67 the platform on which to converge**.**

Multicast

AES67 allows for both network routing modes, multicast and unicast. Each has particular strengths and limitations. Multicast is very valuable for creating an 'every audio stream available everywhere' facility, a virtual routing matrix, with one-to-many virtual distribution amplifiers on every channel,. This allows the ultimate of unrestricted real-time, on-demand use and monitoring of hundreds or even thousands of audio channels in an entire facility. It is ideal for the fast-paced, live broadcast plant and is the model on which thousands of audio over IP radio broadcast studios are built around the world.

However, multicast comes with the need to share and coordinate the use of network resources, which even though it can be done automatically, works best with unified authority over resources. Most WANs do not enable multicast traffic, even though it is technically feasible. The administration of who is using what part of a shared network resource, has proven to be problematic. The attempts to create a multicast backbone for the internet in the mid to late 1990s were stymied by the difficulties of getting competing corporations to cooperate [5]. Alas, widespread multicast based AES67 networks will have to await the development of some new economic force to motivate the required network cooperation. If and when that day comes, AES67 is ready.

Private wide area multicast networks certainly are possible and do exist. Some of these have impressive extents and reach. Multicasting simply takes the centralized policy decision for configuration and use.

Unicast

AES67 requires audio equipment to implement network unicast routing mode in addition to multicast mode. Unicast routing better fits the use model of on-demand use of resources, as point-to-point connections are established. This is the model that often better fits for connections between facilities and over WANs, when the WAN provider does not enable multicasting. Unicast networking mode also fits the immediate on-demand audio connection because it requires no coordination with, and causes minimal or no interaction with, any and all other audio connections in use.

To be clear, all audio network connections, even within the facility, can be accomplished in principle with unicast addressing, in fact this might be the preference of the facility designer. The downside of unicasting when you have multiple listeners, is that the sender has to send a copy of the audio stream to each of the listeners individually, which could overload a simple source device with too many listeners. The virtual distribution amplifier on every audio channel that comes with multicasting is not present when unicasting.

AES67 allows the audio application to take advantage of all of the flexible and dynamic routing abilities that the IP network has to offer.

SIP

Continuing the analysis about the power that AES67 gains from using so-called least common denominator technologies, an additional protocol that AES67 takes advantage of to mesh with the rest of the communications universe is *SIP*, or Session Initiation Protocol.

SIP is the protocol that is used to ‘dial’ and connect one part to another in a familiar way, very similar to the traditional operation of a telephone. In fact, technically underneath, SIP bears a strong functional resemblance to “SS7 Signaling System 7”, the protocol at the foundation of the Bell System (later AT&T) digital switched telephone network for the past 40 years. SS7 begat ISDN which then, in marriage to IP, produced the functional stepchild, SIP, in 1996.

SIP for unicast - the difference between AES67 and proprietary AoIP protocols

“You’re engineering a session with a remote talent who is in another part of the building, or off-site, and you can’t directly see. You open the mic… and there’s silence... dead air“

SIP has 45 different responses for why the connection did not go through. Wouldn’t you like to have 45 different ways to indicate ‘why isn’t this working right now?’

SIP represents a big step forward in the technology of making connections. This can include sophisticated features like forwarding and following, parking a connection to be picked up somewhere else, allowing a proxy to redirect and, as mentioned, detailed system self-diagnostics. As systems get smarter and more powerful, some of that power needs to be used for this kind of self-checking, self-analysis, and self-diagnosis. This why AES67’s unicast mode uses the SIP protocol.

It is important to realize that the use of SIP is a genuine innovation in AES67. None of the previous proprietary professional audio over IP solutions used SIP. The prior AoIP systems were designed to more or less to solve the problem of getting audio around a facility or a venue from the ‘bottom up’ in a rather self-contained way. They certainly all took advantage of the flexible routing of the IP network, but they stopped short of leveraging higher level solutions for making connections. With the benefit of additional years of hindsight, the designers of AES67 realized the required technology was proven, and the time was right to add into AES67 a smart connection protocol.

Other examples using SIP

AES67 is not the only one to realize this power and usefulness of SIP. The N/ACIP organization adopted SIP for compressed audio codecs over IP [6].

The I3P, “Intercom Interoperability over IP” group also chose SIP for IP based intercom system [7].

Of course, already mentioned is the global telephony network and every VoIP PBX, including Asterisk, the industry-changing open source PBX [8].

MANAGING AES67 SIGNAL ROUTING USING SIP

Given SIP as the small to large scale spanning protocol for establishing audio sessions (connections), not all work can be done using a simple point-to-point connection.

On top of the simple audio connections are different types of more complex audio interactions between multiple parties, groups of people, or sets of equipment. Traditionally the workflow of what needs to be accomplished is quite different and separate between telecom, intercom and studio audio; made unique by the connection intent. In fact, the workflow, and captive knowledge to make efficient workflows, is what traditionally differentiated the equipment designs in these three system categories. Just what happens when you press a given button in a fast paced live production situation is so valuable that in the past getting this just right justified what amounts to a completely redundant parallel (and expensive) system for carrying that audio.

Viewing the different ways that audio is used, it can be seen that in the traditional domains of telecom, intercom and studio audio, there are sets of unique and different *connection intents* for each of these. However, each does not necessarily require a redundant complete system for carrying the audio or making audio connections. The different connection intents of each, in reaction to user commands and actions, can be implemented *on top of* a general audio connection mechanism used in common by all. This is the disruptive potential of using the least common denominator audio technologies in AES67 audio over IP. For the first time it is more economically feasible to use the common audio connection fabric, rather than separate, independent, purpose built audio interconnection systems.

Excess latency is the bugbear of digital packet based systems, and must be tightly controlled. By way of illustration, the goal of matching the low latency of traditional analog mixers was set from the very beginning of development of Telos' Livewire AoIP technology in 1999, and is in fact the genesis of the 'Live-' in the name. If the AoIP system did not reproduce the natural live feel of the talent hearing his or her own voice in headphones, and for in-studio conversations, the 'new' digital technology, for all of its economic benefits, would not have been accepted. The same continues to be true in the convergence from special purpose-built systems. AES67 contains the same ability to reach low latency, as live as analog, and the new systems built with it must function professionally and flawlessly in every way.

The electronic 'Little Black Book'

At the topmost level of working with audio is the question, 'Who?' *Who* do I want to connect to? The desired audio is a *person*. The *talent*. A *performer*. In the modern age, this person may be located in a number of different places and connected via a variety of equipment that is where they happen to be (or if mobile, on their person.)

This shift from location based connection to personal identity based connection is demonstrated of course by the shift from fixed telephones to cell phones. This shift has also trained us how to be gatherers, maintainers, and users of our own contact lists, the 'electronic little black book' of who is important to us. No single, so-called 'discovery' protocol can solve the many facets of automatically finding the SIP addresses of who the important contacts are, which is one of the reasons a protocol attempting to do just this was not included in the AES67 standard itself, but left to be handled at a larger scope.

It is a fundamental *human interaction* to gather and maintain the contact list important for getting the whole audio job done. It is straightforward to see the pattern of collecting contacts applying to collecting the important set of SIP addresses. And one can imagine the seamless transfer of important contacts from a personal smart phone device for use by professional audio equipment. Through the mechanism of SIP registration, the talent logs into the microphone where he or she is located, and the 'call' goes through from the console in perfect studio quality AES67 digital audio.

Of course, static audio connections are always possible as well. The majority of routing inside a plant has fixed function, and can be nailed up and tested ahead of time. By basing audio interconnection on a fundamentally addressing-based routing network, it is much simpler to pre-configure fixed connections within network routing, as compared to accommodating dynamic connections into a traditional fixed wiring facility. With network routing, all routes are under software command control, able to switch 'on the fly'. Fixed routing can be entered manually on configuration web pages of the audio equipment, or the routing can be put under the control of an audio routing software application.

The Axia Pathfinder software allows patterns of audio routes to be predefined and controlled in reaction to user button presses, audio level silence detection, pre-defined time schedules, and contact closure inputs (via contact closure to IP protocol, GPIO nodes). This layered architecture uses audio over IP for transport under the command and control of software using the same IP network for that control, but able to be separate yet integrated. This general paradigm of 'separate but integrated' using the IP network as the common platform applies to the convergence of control, monitoring and diagnostics as well.

CHALLENGES

One of the challenges for this vision is that we may not have universal access to the ubiquitous high speed WAN network… *yet*. So whatever we do in the meantime needs to be aligned with the coming future, even if we don't precisely know when the WAN network barriers will fall. But they will fall, with the certainty of Moore's Law.

The challenge is to identify and eliminate the architectural limitations, the hidden assumptions, the backwards looking conceptual models, and the built-in bottlenecks in our audio systems, so that as the network barriers fall, that speed and connectivity can be immediately used.

The way to design systems now is to plan for the ubiquitous network bandwidth future, base the system entirely on standard protocols and a network centric architecture, and then build specific bridge devices to span the gaps in available WAN network capability. But not to let these bridges and 'temporary fixes' come to be any part of the foundation or fundamentals of the new architecture.

Using appropriate codecs, if needed, to connect remote systems over WAN

Many choices exist for high quality, relatively low latency codecs, as needed for audio bitrate reduction, that can dynamically adapt to unpredictable network conditions. By using a codec device with an AES67 local audio interface allows the codec to be used as a bridge, but not create a future architecture barrier. When the WAN becomes capable of supporting the AES67 audio connection directly, the codec can be taken out of the path, and the end devices that used to address the codec using AES67 SIP can then connect to each other directly, with only a change of SIP address.

REFERENCES

[1] Audio Engineer's Reference Book, ed. Michael Talbot Smith, Focal Press, Jan 1, 1994

[2] http://www.aes.org/e-lib/browse.cfm?elib=16141

[3] http://fiber.google.com/about

[4] http://www.100gigcle.org/

[5] "MBONE: Multicasting Tomorrow's Internet", Kevin Savetz et al, John Wiley & Sons Inc. (Computers) (March 1996)

[6] https://tech.ebu.ch/groups/nacip

[7] https://tech.ebu.ch/groups/niiip

[8] http://www.asterisk.org/