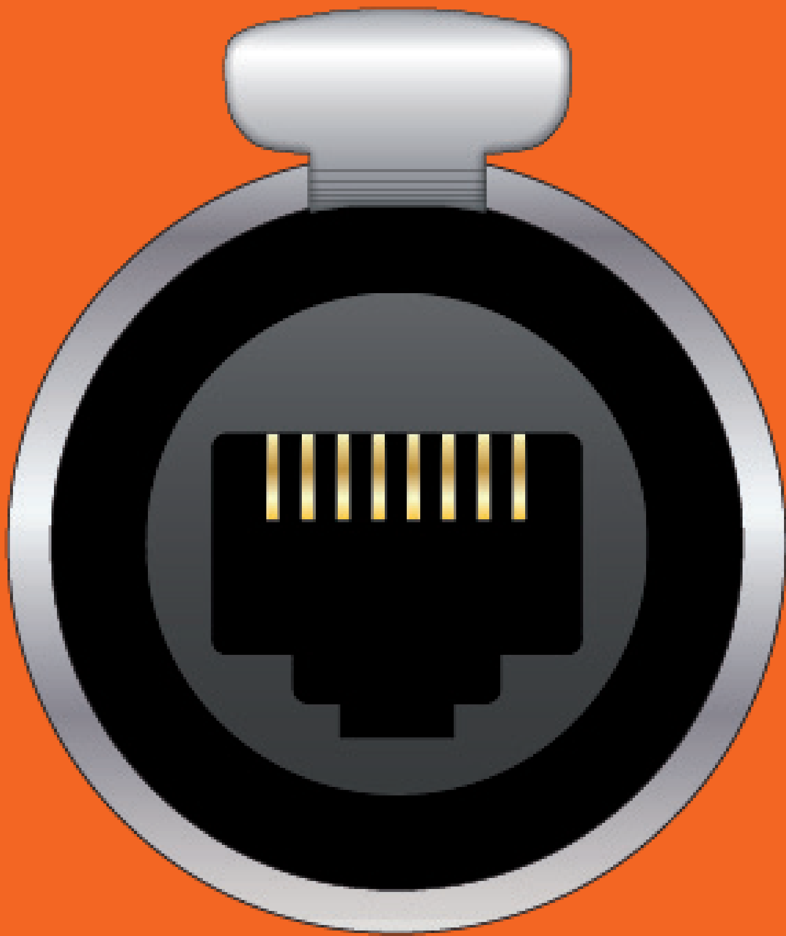
AoIP and Networking:
Network Primer V2

CALREC



Introduction to professional audio networks - 2017 edition

Calrec Audio Ltd
Nutclough Mill
Hebden Bridge
West Yorkshire
England UK
HX7 8EZ

Tel: +44 (0)1422 842159
Fax: +44 (0)1422 845244
Email: enquiries@calrec.com

calrec.com

Whilst the Company ensures that all details in this document are correct at the time of publication, we reserve the right to alter specifications and equipment without notice. Any changes we make will be reflected in subsequent issues of this document. The latest version will be available upon request. This publication is for International usage.

Calrec Audio Ltd reserve the right to change specifications without notice. E & O.E.

The established policy of Calrec Audio Ltd. is to seek improvements to the design, specifications and manufacture of all products. It is not always possible to provide notice outside the company of the alterations that take place continually.

Despite considerable effort to produce up to date information, no literature published by the company nor any other material that may be provided should be regarded as an infallible guide to the specifications available nor does it constitute an offer for sale of any particular product.

NETWORK PRIMER V2
CONTENTS

NETWORK PRIMER V2
FORWARD

This Primer is a new edition that updates the original 2013 Calrec Networking Primer to reflect the many changes that have occurred in the world of broadcast networking technology since then — four years is a long time in this field.

Nonetheless, our aim is still to explain the background and technology behind data and IP-based networks with specific reference to its application in broadcasting, such that a broadcast engineer with no formal training in the field of computer networking technology might gain a clearer understanding of the subject.

This Primer aims to explain the increasing use of data and IP-based networking technology in broadcast applications, examines the benefits that this technology can offer forward-thinking modern broadcasters, and considers where it may take the world of broadcast in the future.

Hebden Bridge, Spring 2017

NETWORK PRIMER V2
INTRODUCTION

calrec.com

CALREC

INTRODUCTION

The word ‘network’ has a long history; it was first recorded in the mid-sixteenth century, but its roots hint that it may well have been in use much earlier than that. Like much of modern English, it has gone through a variety of ever more abstracted meanings, from a description of a physical object to more figurative, intangible concepts. What began as a way of describing an invention for catching fish was later applied to similarly interconnected physical systems, including those of canals, railways, telephone wires – and eventually computers.

By the early 20th century, the word had come to refer to systems of related entities that were no longer physically connected, as with radio and TV transmitters belonging to a single broadcaster, or groups of business colleagues.

In the last few years, we’ve had ‘wireless networks’ (which medieval speakers of English might have regarded as an oxymoron) and ‘social networks’ composed of users that ‘connect’ solely via data transmissions over the Internet; itself a network of physically disconnected, but nonetheless linked computers.

A similar process of abstraction has been taking place in the world of broadcast technology, transforming the hard-wired, localised studios of the past into more flexible, networked systems. Forty years ago, television studios consisted of cameras and microphones hard-wired into discrete hardware vision mixers, patchbays and audio mixing consoles which routed to specific video tape machines in one location. In such systems, a separate physical connection is required for each audio channel, whether from a microphone, mixing console, or recording device.

Modern networked broadcast systems offer more flexibility; all of the hardware is permanently connected to a data network, and the precise nature of the interconnections between the equipment can be redefined and reassigned at any time under software control, remotely if required.

Such ideas are not new, and small-scale proprietary networks of this type have existed in broadcasting for many years. But the declining complexity and improving cost-to-benefit ratio of implementing large networked broadcast systems, coupled with the widening capabilities of the technology, has tempted more and more of the world's forward-thinking broadcasters to move to networked systems.

In the past few years, the spread of IP-based IT networking technology across the world has furthered the development of such networks. Increasingly, broadcast technology manufacturers and standards organisations are developing systems that utilise, at least in part, hardware, standards and data transmission protocols originally created for the world of IT, rather than developing proprietary systems.

This is bringing costs down further, as well as adding a more global dimension to the capabilities of the networks being developed for broadcast.

As IP networks make geography and the physical proximity of resources less and less important, the very concept of a studio itself is becoming more abstract. Already, it is possible to conceive of - and even to work 'in' - virtual broadcast studios, where the control surface is in one location, the DSP used for the mixing in a completely different building, and the audio inputs and outputs somewhere else altogether - perhaps even in a different country.

As long as they are all connected to the same network, they should be able to interoperate just as effectively as the days when all these components were part of a single mixer in a hard-wired studio, permanently located in one place.

These more flexible, abstracted implementations of once-tangible, hard-wired, and immovable resources, offer broadcasters many benefits, as we shall see. These include the ability to move projects swiftly from one studio to another by reassigning connections, or controlling aspects of the mixing process remotely, without having to be in the same geographical location as the event being mixed.

At the same time, we're getting closer to the goal of being able to transmit broadcast audio, video, sync information, control data and metadata together over agnostic, scalable IP-based data networks consisting simply of data cable and IT switches.

This promises further advantages which will transform how audio, video and control data is encoded, transported and managed, such as the prospect of being able to dispense with separate audio, video and data transports, expensive analogue audio and video cabling, and relatively inefficient embedded-audio video interfaces such as SDI.

In the medium term, broadcast workflows will evolve to take advantage of the greater flexibility and geographical freedom available to them.

In the longer term, the broadcast industry will further borrow from the IT industry by shifting away from bespoke hardware towards software processing running on commodity computing platforms. While not all broadcast processes will fit this model, many will, and in doing so, offer benefits in scalability and economy.

As the shift to an IP infrastructure continues, we will be encouraged to drop our conventional signal-based approach in favor of a services model, where content, both live and stored, may be discovered and accessed by anyone in possession of access rights and an appropriate IP connection, regardless of their location.

However, at the time of writing, much remains to be done. Standards enabling practitioners to define and realise all of these possibilities are still being written and ratified, and manufacturers are still developing networking protocols based on open standards that will allow their equipment to interface and work together.

There is no doubt that there are huge opportunities for forward-thinking broadcasters and technology manufacturers who are prepared to embrace and engage with the changes. It's time to look at some of these benefits in more detail.

NETWORK PRIMER V2

CHAPTER ONE: BENEFITS OF NETWORKING

CHAPTER ONE: BENEFITS OF NETWORKING

A networked broadcast studio, editing suite or transmission station isn't necessarily more efficient than a hard-wired one — indeed for smaller broadcasters, the costs associated with setting up a network can outweigh the benefits. But introducing networking into a large-scale broadcast environment can benefit the whole system; networked equipment is both more accessible and more flexible than its hard-wired counterpart.

To understand this, consider what the invention of audio patchbays did for arrays of hard-wired studio equipment. Studios function perfectly efficiently when equipment is connected directly, but time is always lost whenever new equipment is connected and its output needs to be made available to other parts of the system. Introducing patchbays to studios initially costs money and the time to wire everything up to the patchbay, and some studios chose to save themselves that time and expense.

However, once a patchbay has been integrated into a studio, any input can be routed to any output, producing significant long-term savings. Introducing new equipment to the system and giving it the same routing flexibility becomes a simple matter of connecting it to the patchbay.

When MIDI patchbays were invented, the routing of signals also became remotely controllable, and this greater flexibility and remote controllability is another benefit of networked studios.

But as we approach the 2020s, data networks have sufficient bandwidth to do much more than route audio around; today it's possible to send broadcast video, audio, meta and control data over a network. Standards to transmit all of these simultaneously are still under development, but progress towards that goal is moving fast.

You may well ask what difference it makes whether audio is being routed around a studio via CAT5 network cable, fibre-optic or co-axial links, or even analogue tie-lines? And while the idea of leveraging IP-based IT infrastructure promises lower-cost connectivity in the longer term, for broadcasters with existing facilities, there's still the cost of installing it all these network connections in the first place.

Routing & Cost Benefits

Nonetheless, there are a number of benefits associated with a networked approach, and the larger and more complex the facility, the more attractive these are - especially if the studio is a new-build project that requires no investment to replace or adapt existing infrastructure.

Firstly, modern mixing consoles can take control of all the audio routing in a studio if networked, allowing broadcasters to save themselves the expense of a separate audio router or patchbay, together with all the connections and wiring to it. The more studios you have, the lower the outlay on routers and patchbays.

The second advantage is flexibility, which is again a particular benefit in broadcast complexes with large numbers of mixing consoles, control rooms and studios. Using a network and an audio routing protocol/management system, it's possible to route microphone sources from a wallbox in one studio into the control room of another in seconds, or assign the mixer in one control room the task of mixing the combined output of two or more studios, and then return it to being dedicated to a single room again when the job is complete.

Even with patchbays that allow flexible interconnection between studio and control room, this would be a challenge in a traditional, hard-wired studio. What's more, achieving such 'super-studios' via traditional audio connections, whether analogue or digital, requires the running and temporary installation of a lot of extra, expensive cables and looms, and increases the risk of on-air faults.

But in a new-build studio designed around a network from the outset, the cost of the network interconnection cabling is minimal and all the hardware is already connected to the network. The routings are simply reassigned by control software. A few clicks of a mouse, and the work is done — or undone.

Such flexible workflows are increasingly apparent in modern broadcast and the genie is out of the bottle. The next logical step once audio is networked is to connect one console's router to another. It then becomes very simple to transfer all of the audio being received at one console, or even just at one I/O box, to a console in a completely different studio for mixing.

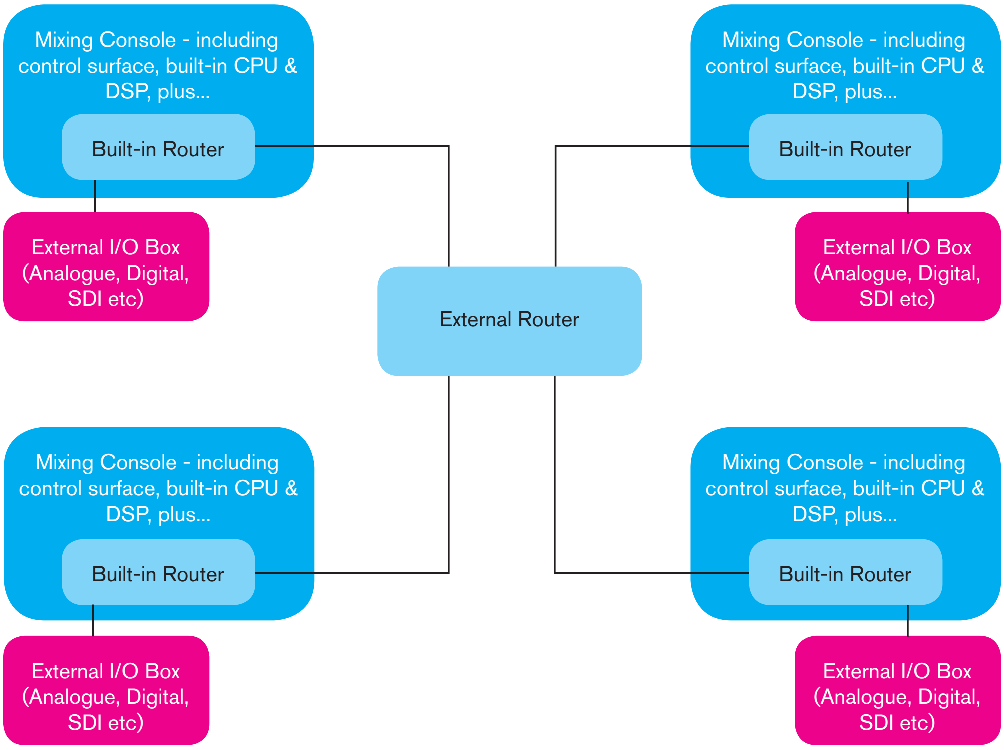


Figure 1 - STAR NETWORK

Given the bandwidth of today's networks, which typically allow many hundreds of audio channels to be passed down a single connection, there's no need to stop at interconnecting a pair of consoles and their associated I/O. Why not link many consoles together with a stand-alone audio network router, and thus allow several mixers to freely swap audio channels? This is the basis for a star network like the one above (figure 1), with several consoles connected to a stand-alone router.

In a broadcast complex structured like this, sound stages or studios are (quite literally) no longer tied to a single control room. It's very simple to take multi-channel audio being received from one studio and mix it in another, or to route that audio to another studio to create different mixes for (say) international versioning or commentary. On a rolling news show, the production team in one control room can be mixing the live audio from the news studio and can hand the audio from that studio over to the incoming team in a different control room and go off shift.

Or the team in one control room can switch from mixing the output of one studio or sound stage to working on the output from a different one in seconds. Such things can be done with analogue or digital tie-lines, but a vast amount of expensive wiring is required. This is not a concern with networked audio given that you can route several hundred channels of high-resolution audio down a single inexpensive Ethernet-style network cable.

But even a star network is only the beginning once multiple consoles are networked. In Figure 1, each console still has its own dedicated I/O interface (or more usually in this day and age, several interfaces, each handling different audio output formats). This is still quite a traditional structure that owes much to the days when the I/O was a fixed part of individual consoles. A network permits something more like the modified star structure shown in Figure 2. Here an assortment of I/O interfacing boxes in a central location is shared commonly by all the consoles in a broadcast complex, and is connected to them via the central stand-alone router.

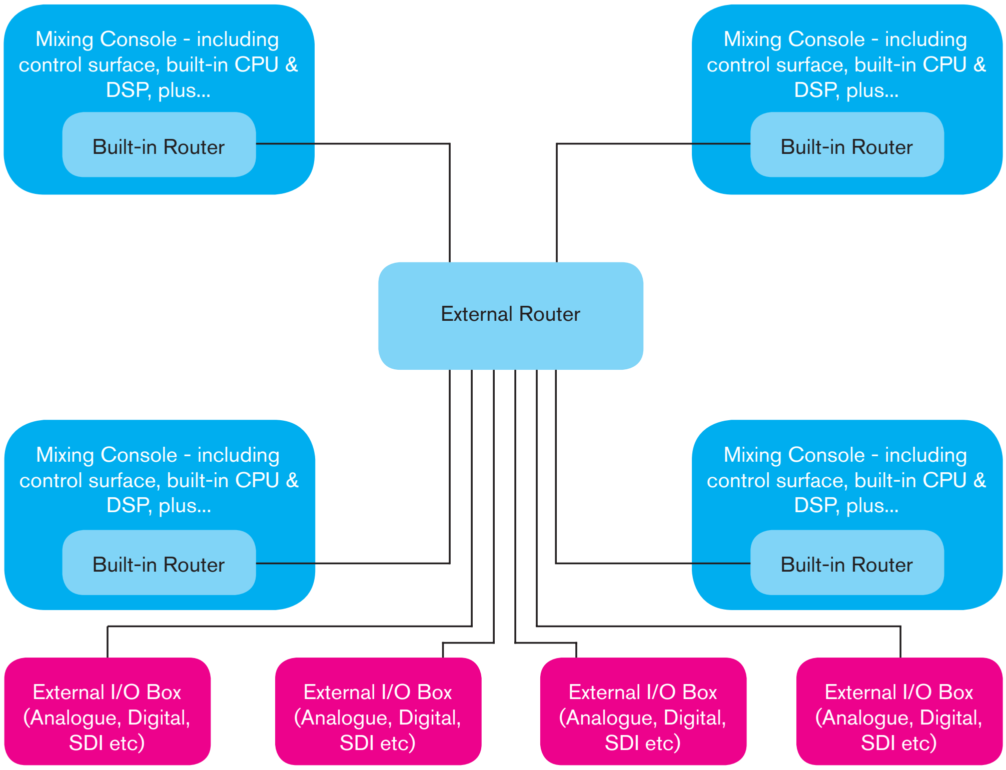


Figure 2 - MODIFIED STAR NETWORK

Packing hundreds of channels of audio into a single high-bandwidth networked data connection, where connections can be easily made and reassigned, encourages the construction of complex workflows and network topologies that would be difficult to achieve with standard audio connections.

When one considers the costs of wiring a national broadcaster's transmission control centre, it becomes clear that audio networks can offer significant savings.

Networks and Resilience

For the same reason, it's considerably less costly to design failsafe systems if your broadcast audio is part of a network. In broadcast, the failure of mission-critical connections live on-air is unacceptable. Modern broadcast control centres like to factor redundancy into their designs, so that every connection has a backup which can easily be pressed into service in the event of a failure.

Doing this with traditional analogue or digital connections requires twice the amount of cabling and a lot of complicated cable splits. With networked audio, the entire output of a master control room or transmission centre - which may include thousands of channels of audio - can be duplicated on a few Ethernet-style IT cables.

Many broadcasters choose to mirror their resources across a network in this way, assembling identical equipment in physically or geographically separate buildings which may be networked together, but may also continue to function independently in the event of one half of the network becoming unusable, as in figure 3, which shows one such 'linked star' network.

In this way broadcasters can be said to achieve 'resilience' in their systems more easily and affordably than those employing traditional designs. Several leading broadcast audio mixing console manufacturers also take advantage of this technology to offer highly resilient network structures with redundant hardware as well as routing.

In such systems, the fundamental hardware components of the audio console can be duplicated, and the duplicates placed in separate physical locations (figure 3). In the event that one of the locations is rendered inoperable by power failure, fire, flooding or some other unforeseeable natural or man-made catastrophe, the components in the other location can take over and be switched into operation seamlessly over the network, ensuring continuity of operation.

Many audio consoles, including those from Calrec, offer this way of working as part of a standard setup; the fundamental hardware components that drive the latest generation of consoles (such as the main processor, router, and DSP cards) may be supplied in pairs as standard. Placing one set of the pairs in a different location and linking them via the audio network is a simple matter. Once again, it stands repeating that offering such resilient workflows without networked audio would be incredibly complex and expensive.

Networks & Contingency Planning

Networks also make contingency planning easier and more flexible. For example, what do studio owners do if routine maintenance needs to be performed on one studio or control room, and the program that usually transmits from that studio is due to be broadcast? Or, in a commercial broadcast complex, what happens when a last-minute booking is received from an important client and needs to be accommodated without disrupting the usual assignment of studios and control rooms to other clients?

In a networked broadcast complex, the output of any studio can easily be assigned to a different control room, or conversely a familiar control room can be used to mix the output from a different studio.

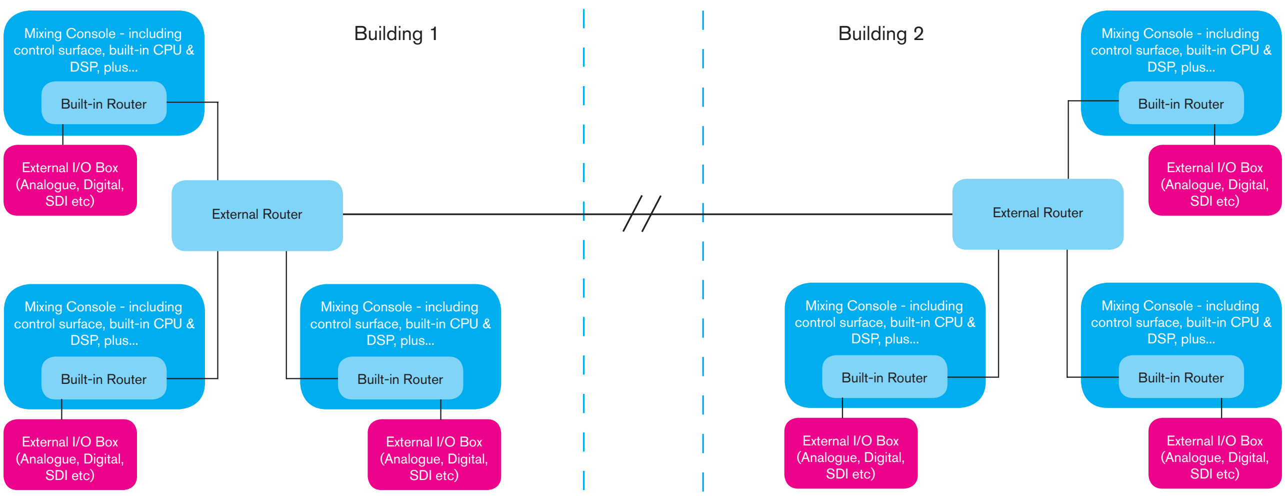


Figure 3 - SPLIT ROUTER CORE

Remote Production

A more recent application for networked studios has only become possible now that IP technology is making data-intensive global intercommunication a reality, such that high-resolution audio and video can be moved around the world reliably and affordably.

This is the step that promises to render geography irrelevant, and allow the distribution of the components formerly held in a single studio across several locations, internationally if required.

Again, it's possible to read this and think 'But why introduce network technology? Remote Production has been possible for decades already — that's what Outside Broadcast is.' This is true, but IP technology is making it possible to undertake Outside Broadcast at a fraction of what it has hitherto cost, opening up new possibilities for broadcasters.

Until recently, remote production — that is, covering a sports, news or entertainment event a significant distance from a traditional broadcast studio (figure 4) — has required the deployment of at least one broadcast truck, together with its associated staff: camera operators, vision and audio mix engineers, and support engineers. Such a team usually comprises several highly skilled and experienced (and therefore equally expensive) personnel.

With IP technology, it is possible to scale back the equipment and personnel required at the site of the event to some audio and video capture devices (ie. cameras and microphones), some digital signal processing hardware of the kind found at the heart of modern broadcast

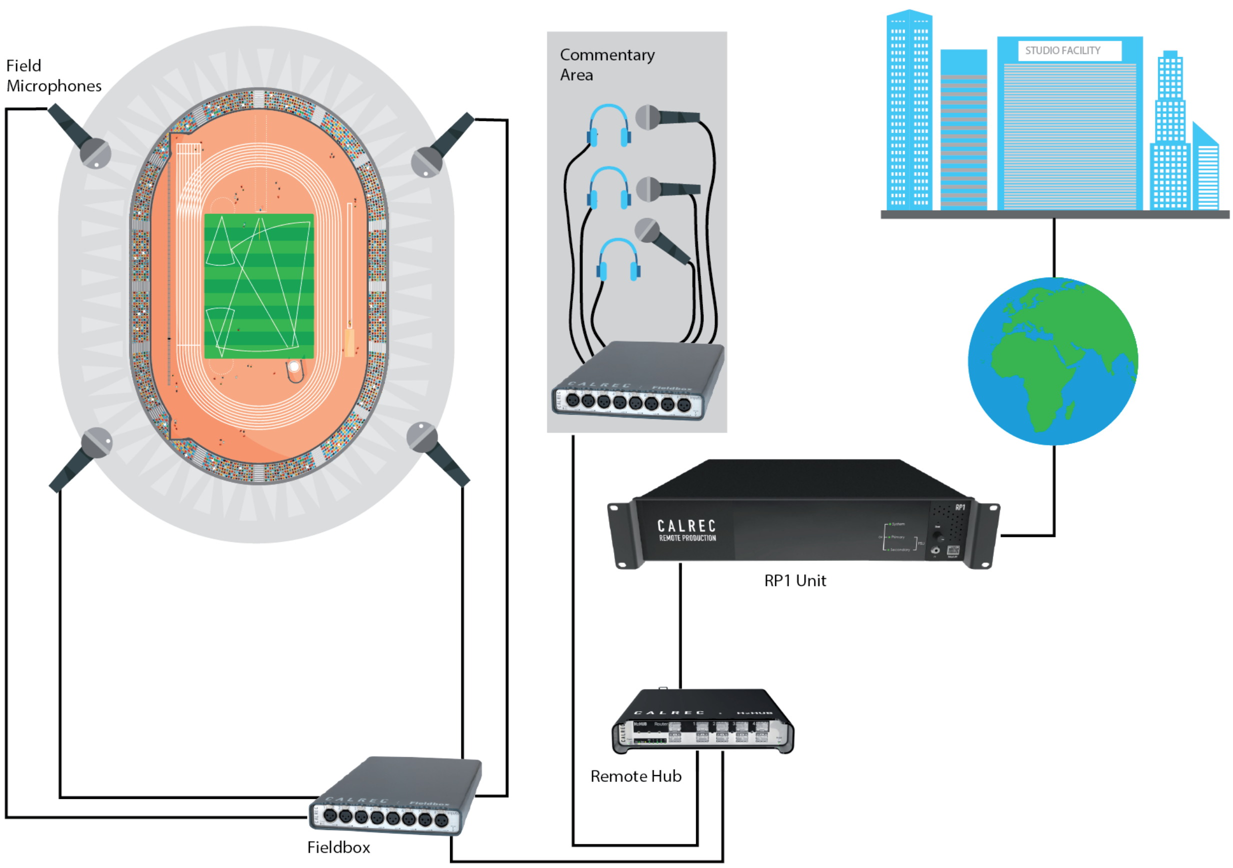


Figure 4 - REMOTE PRODUCTION

consoles, and some interface units such that the captured audio and video can be uploaded to a high-resolution broadcast network. Far fewer staff are required to set up this equipment on site and keep it operational, reducing the financial outlay of a dedicated on-site team.

Once this equipment is at the event venue and connected to a secure IP network, the clever part is that the audio can be readied for broadcast centrally, at a traditional broadcast facility. To a networked mixing control surface, the audio interface boxes and mixing DSP at the venue are connected just as dependably as if they were in the same room as the control surface.

The audio captured by the on-site microphones and uploaded to the network by the audio interface boxes at the venue can be mixed from afar on the control surfaces at the broadcast facility, by skilled operators who are spared the need to travel to the venue. In doing so, they can make use of the monitoring facilities at the broadcast centre, which are usually superior to anything outside broadcast vehicles can offer. Importantly, because the mixed audio is output locally by the DSP at the venue, IFB mixes for on-site commentators or presenters on site are created without latency, and the mixed audio is also returned to the central facility over the IP network virtually simultaneously for broadcast.

The cost savings broadcasters can make using such network-based workflows are considerable, and make it theoretically possible to broadcast many new kinds of specialised events, the coverage of which would not be economically viable using traditional OB infrastructure. For example, the interest in regional or college sports, or smaller entertainment events has not hitherto justified the expense of sending the broadcast trucks and staff needed to make broadcast coverage a reality. In today's diversifying industry, broadcasters have an ever-growing need for more content, but have fewer resources with which to capture it. Networked infrastructures can deliver a possible solution to this difficult problem, and broadcast audio manufacturers including Calrec are now delivering the technology to make it happen.

Interoperability — The Holy Grail

The advantage conferred by the network-based infrastructure described so far is undeniable, but the development of the technology that allows broadcasters to implement these new workflows has been slow and piecemeal.

Although data networks have been interconnected across the world for several years and it has been possible to adapt IT networking technology to carry audio and video data, it doesn't necessarily follow that you can simply feed a live broadcast stream into an Ethernet router in Salford, UK and expect it to emerge unscathed in an edit suite in Shenzhen, China.

IP networking technology originally designed to handle office-based data transport initially proved far from optimal for routing, mixing, processing and controlling real-time multi-channel audio and video, and although means of reliably networking audio were eventually established over the past few years, they were all proprietary standards at first, which made it impossible to use a mixture of equipment from different manufacturers.

Furthermore, only very recently has there been any progress made towards establishing agnostic standards that allow audio and video to be transmitted together across an IP network, other than by inefficiently embedding the audio with the video, de-embedding it for processing or mixing, and re-embedding it afterwards, which adds processing time at every stage.

Throughout this time of transition, the Holy Grail has been a set of networking standards that allow the use of a single high-capacity IP network for all of a broadcaster's infrastructure: IT, phones, intercoms, broadcast audio and video.

The goal has been that these standards, properly written, will allow overall system description and management, equipment monitoring, and describe control protocols to allow one set of devices to control others on the network, irrespective of their original manufacturers.

This shining goal is known as 'interoperability.'

The good news is that much progress has been made towards this end, and the realisation of all the benefits it will confer; we will look at the cross-manufacturer audio standards that exist so far in chapter three of this primer, and consider how video and audio should work together via the medium of IP networking in the not-too-distant future in chapter four.

Before we do that, however, it's time to move to chapter two and look at a few networked audio basics in more detail. These will enable us to better understand:

a) the proprietary audio networking standards developed over the past few years, and;

b) the more recent cross-manufacturer standards that are the focus of chapters three and four.

NETWORK PRIMER V2
CHAPTER TWO:
SOME TECHNICAL BACKGROUND

CHAPTER TWO: SOME TECHNICAL BACKGROUND

As we've seen, proprietary standards for transmitting audio over IT networks have been independently developed by individual manufacturers over the past few years.

Some do a more complete job than others, but all were designed to deal with the basic fact that the requirements for delivering broadcast audio over a network are considerably more stringent than those for IT-related data.

IT networking protocols for data transfer (such as the ubiquitous Ethernet) are asynchronous, meaning that the order in which data arrives is not held to be so important as long it arrives eventually.

However, a broadcast audio feed usually consists of many channels of high-resolution audio, all of which must be kept in sync with respect to each other and which have to be delivered in real time to avoid dropouts.

Looking at the technicalities of data transport over a network, we can categorise audio networking protocols in terms of how closely (or not) they resemble IT networking data standards.

Modern electronic networks are often described in terms of a notional model of up to seven layers of increasing complexity that can be used to integrate communications protocols into real-world applications. For audio networking, the most important of these are the first four layers (which are also the most fundamental).

Layer 1 describes the basic electrical standards and voltages used to transmit data over a wired or wireless network, such as an Ethernet network.

Layer 3 Protocols
Encapsulate audio data in standard IP packets.
Layer 3 audio networking products include:

Audinate's Dante
QSC's Q-LAN
Wheatstone's WheatNet-IP

Layer 2 Protocols
Encapsulate audio data in standard Ethernet frames.
Layer 2 audio networking products include:

AES51
CobraNet
Digigram's EtherSound

Layer 1 Protocols
Ethernet wiring and signalling componants .but do not use Ethernet frame structure.
Layer 1 audio networking products include:

Aviom's A-Net
Riedel's RockNet
Calrec's Hydra2

Layer 2 describes the most basic unit of data used on the network; in an Ethernet network, this is the 'frame' containing the electronic data. Ethernet Frames include source and destination MAC addresses to identify the source and destination device for data being transmitted.

Layer 3 adds the IP subnet structure used by all Ethernet networks (and Internet servers) to uniquely identify network devices across the globe, and packages the data being transferred in IP packets. These are all numbered to ensure that all of the data arrives in the right order and can be accounted for.

Layer 4 adds the ability to check the arrival of these packets has occurred in the correct order, without losses or duplication.

In practical terms, all of the audio networking technologies currently on the market are either Layer 1, 2 or 3 protocols (and all of the Layer 3 protocols contain Layer 4-style data verification capabilities). However, it would be misleading to suggest that Layer 1 protocols are the most basic and Layer 3 the most feature-rich.

Certainly, Layer 3 protocols conform more closely to the defined standards of Gigabit Ethernet (the most common network standard) than others, including Layer 4-style packet checking spliced into Layer 3-style IP packets.

These packets sit in turn within an overall Layer 2 Ethernet Frame structure and adhere to the basic Layer 1 electrical definitions of Gigabit Ethernet.

Layer 3 Protocols

Thus the structure of the data in Layer 3 protocols, which include Audinate's Dante, Ravenna from ALC Networx, Axia's Livewire and QSC's Q-LAN, most closely resemble that passing over a standard Gigabit Ethernet network. As a result, they can multicast data to multiple IP addresses simultaneously, as on an office network, and they can pass data via connected Ethernet bridges and routers. This potentially allows the data to be passed over a wide geographical area and not to remain locked within one Local Area Network (or LAN).

Layer 2 Protocols

The data structure of Layer 2 protocols, which include the IEEE's Audio Video Bridging standard (AVB), Calrec's original Hydra protocol, Peak Audio's Cobranet and Digigram's EtherSound, less closely resemble standard Ethernet data. These protocols dispense with the IP packet structure and thereby lose the ability to be routed to other standard LANs. In this way, Layer 2 protocols cannot traverse the internet, although they still use the Ethernet Frame structure and can therefore still be routed within their network via off-the-shelf Ethernet hubs and switches.

Layer 1 Protocols

Layer 1 protocols, which include Riedel's RockNet, Aviom's A-Net, Gibson's MaGIC, and Calrec's Hydra2, have the least in common with IT-style network data. They are geographically limited and also have to use proprietary routing hardware. However, many Layer 1 audio protocols offer similar routing and real-time verification capabilities as Layer 3 protocols — but they do it by means of self-developed, proprietary means.

Although compatibility with off-the-shelf networking hardware is lost in a Layer 1 protocol, dispensing with the 'higher-layer' data structures allows the development of very efficient, robust, high-performance, low-latency protocols. When coupled with the hardware required to use them, they are arguably better suited to professional broadcast applications (albeit usually more expensive to implement).

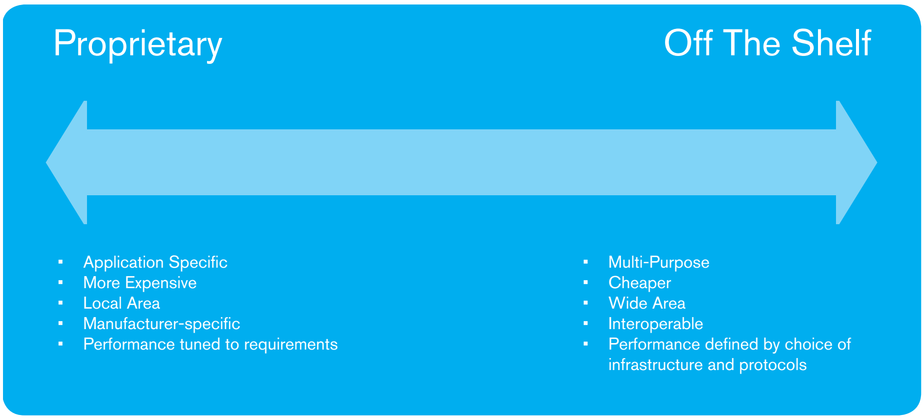
To summarise: 'higher level' protocols offer far greater compatibility with standard networking formats, and allow the use of standard, affordable networking hardware. This can make installation more cost-effective and usable over a wider area, but it can also mean that these protocols are less efficient and higher in latency.

Moreover, because the data in 'higher-layer' networks is usually passed via non-proprietary hardware which is not specifically designed to carry audio data, the reliability of these networks can be lower, and therefore less attractive to broadcasters who need their infrastructure to be as robust as possible.

The 'spectrum' diagram below is a reasonable summary in graphical form, with Layer 1 protocols at one end (more expensive to implement, more application- specific and geographically limited) and Layer 3 protocols at the other (cheaper, with the potential for use over a wider geographical area, and more interoperable, but with a performance that is entirely dependent on the quality of the hardware and infrastructure being used).

Most Audio over IP protocols or Internet Audio streaming standards would fall on the right side of this diagram, being cheap to implement, but may fall below the standard of reliability required by professional broadcasters.

Perhaps unsurprisingly, it has been attempts to marry the wider compatibility and greater interoperability of Layer 2 and 3 protocols with further standards designed to improve reliability that have formed the basis for the cross-manufacturer protocols that are now emerging. It is to these that we now turn in chapter three.



NETWORK PRIMER V2

CHAPTER THREE:
ROUTES TO INTEROPERABILLTY

CHAPTER THREE: ROUTES TO INTEROPERABILLTY

At the end of Chapter One, we introduced the idea of interoperability - the concept of data being shared freely between video and audio equipment over an IP-based network.

Central to the use of off-the-shelf IT components is conformance to a set of standards which, together, define IP networking. This includes protocols such as RTP (the Real-time Transport Protocol), IGMP (the Internet Group Management Protocol) and PTP (the Precision Time Protocol), all of which are used in audio and video over IP streaming, but would also be familiar to IT specialists outside the broadcast industry.

Over the next few years we can expect to see broadcast equipment manufacturers producing equipment which will interface with common transports, although there is still some work to do in this respect (more on this in Chapter four).

The past few years have seen some of the leading vendors in the audio for broadcast market working together towards interoperability, but as is often the way when an industry first tries to establish standards, various approaches are currently in existence, having been developed by different company groupings.

▪ Ravenna was proposed by ALC Networx at IBC 2010 as “a technology for real-time transport of audio and other media data in IP-based network environments”, Ravenna was, from the first, an open technology standard without a proprietary licensing policy, which encouraged partners to participate in development. As such, it adapted standard network protocols like RTP for use primarily in the professional broadcast market.

Ravenna is a Layer 3 protocol. Although intended for an Ethernet infrastructure, its use of IP packets abstracts it from the underlying network fabric, extending its reach beyond LANs to public networks. In other words, there are no geographical limits to this technology — and the use of standard protocols makes it possible to make use of existing IT-style IP infrastructure, a clear and obvious benefit.

A diverse range of companies is signed up to the Ravenna ecosystem; vendors currently producing Ravenna-compatible equipment include Calrec, Merging, Sonifex, AEQ and Digigram.

▪ Dante (Digital Audio Network Through Ethernet), also a Layer 3 protocol, was developed in 2006 by the Australian company Audinate, and is the most established of the audio protocols covered here. The biggest difference is that Dante is proprietary, rather than an open standard. Nonetheless, hundreds of Dante-enabled products are available, and many technology companies, particularly in the broadcast, installation, live and pro audio industries work with Audinate to provide compatible equipment, including Calrec and Shure to name but two.

▪ AVB (Audio Video Bridging), also known as TNS (Time Sensitive Networking), is another open standard promoted by the AVNU Alliance, a group of companies including Avid, Cisco, Intel, Dolby, Meyer Sound, Sennheiser and Yamaha. AVB is a Layer 2 protocol that uses etherframes rather than IP packets to transport data. Consequently, AVB networks cannot extend across routers or bridges, and are geographically limited to LAN segments. A further limitation is that AVB networks require specially manufactured, AVB-enabled switches
and hubs.

However, the positive trade-off is the guarantee that what you put in is exactly what you get out, and this reliability and predictability is attractive to broadcasters.

▪ AES67 does not describe a full protocol; rather it defines a set of ‘ground rules’ that make interoperation between equipment from third parties possible. It is based on the common ground between a number of established IP-based audio networking systems including Ravenna, Livewire, Q-LAN and Wheatnet. Development of the Layer 3, Ethernet-compatible standard began in 2010 under the name AES-X192, and it was published in 2013. The collaboration led to the formation of the Media Networking Alliance (MNA) in 2014, which consists of various like-minded technology companies who all wish to promote AES67 as the common interchange of digital media between different IP networking platforms. The standard continues to evolve, but at the time of writing is the closest thing the broadcast industry has to a common networking solution.

Management

While there has been some industry success in agreeing a common transport mechanism for IP audio, there has been less success in agreeing how IP streams can be managed.

To fully realise the benefits of IP, it must be possible for a device (or end point, as IP jargon terms it) to join a network, and discover for itself all the streams (or services) available on the network. It must do this in order to allow a human operator to see the available streams and to make choices about which to connect to.

Firstly, this process requires that all devices participate in an agreed ‘discovery’ scheme. Secondly, as the list of connected devices changes, the discovery mechanism must allow for dynamic tracking of available streams.

Thirdly, once a stream has been discovered, detailed information must be provided by the transmitting device giving details of exactly how to listen to, and decode, that particular stream. These details are known as a session description, and include sample rate, encoding mechanism, sample depth, number of channels and multicast IP.

A popular service discovery mechanism, originally developed by Apple Corp, is Bonjour, with other mechanisms including SAP (Session Announcement Protocol) and SIP (Session Initiation Protocol). More recently, development has begun on more industry-specific methods, in AMWA's NMOS IS-04 protocol, and in discovery extensions to the AES70 OCA framework standard. For a network of devices to work together, it is necessary for them all to support a common discovery mechanism.

Unfortunately, the AES67 standard does not mandate the use of a particular discovery mechanism. This is because each has its own strengths and weaknesses, and the authors of AES67 felt that manufacturers ought to have the freedom to choose a mechanism that best matches a particular application. While this position is supportable, its unfortunate consequence is that products with non-matching discovery mechanisms will not be able to interoperate.

This already raises issues; Ravenna uses RTP streaming and Bonjour for service discovery, Audinate's AES67-compliant profile uses SAP, and AES67 itself mandates only the use of SIP. In other words, devices running different protocols will remain unaware of each other's presence on the network. By contrast, for a network implementation to be successful, network designers must be careful to select devices that have compatible discovery mechanisms.

In the face of such continuing incompatibilities, broadcast equipment vendors have their work cut out to ensure that all products will be capable of mutual communication. Calrec's approach is to design end points that support multiple protocols simultaneously via networkable interfaces, making them network-agnostic, and capable of working with a wide range of devices from different vendors.

Other Control-related Matters

Connecting devices on an IP network is not done merely so that audio can be routed from one to another; it also offers the prospect of more sophisticated control integration.

For some years, proprietary manufacturer protocols, including Calrec's Hydra2, have offered differing levels of control over the hardware connected to their networks, depending on the level of support in the connected device and its relative sophistication. As broadcast hardware moves to its IP-based future, control protocols are also being developed to make networked hardware controllable remotely, across the network.

An example might be where a mixing console is connected to a stream from a third-party vendor's mic amp unit. In this case, the mixing console operator might want to be able to control the analogue gain of the mic amplifier in the third-party unit. Another example might be where someone wishes to take control of the AES67 connection management of a device from a remote location. In both these cases, a control mechanism is required that allows one device to control aspects of another device over an IP connection.

There are currently several control mechanisms in development that may be utilised, including EMBER+ and the AES70 open standard, which in addition to control and monitoring of equipment, also handle connection management and self-discovery. These are sophisticated protocols that use a server-client model and do not generally require vendor-specific extensions to allow control of functions.

A device may advertise its AES70 server using Bonjour, allowing potential controllers to discover it. The protocol allows user-defined controls to be discovered, and contains pre-defined objects for complex controls like routing, clock management and stream management.

At the time of writing, it is early days for AES70, and it remains to be seen how widely it is adopted. EMBER+, a fully featured, open source protocol, is also in constant development, aided by a long list of broadcast vendors whose hardware is compatible (including Calrec, Evertz, GrassValley, Riedel and SSL).

Following the development of proprietary, mutually non-interoperable networked audio systems in the 2000s, and the fragmentary development of partially interoperable audio standards in the first half of the 2010s, in recent years the global broadcast industry has committed itself to ensuring proper standards were created for the transmission of audio and video over IP, together with sync, control and metadata. This is the subject of our final chapter, which also looks at where all these new developments might be taking us.

NETWORK PRIMER V2
CHAPTER FOUR:
CONTROL, SYNC & METADATA OVER IP

calrec.com

CALREC

CHAPTER FOUR: CONTROL, SYNC & METADATA OVER IP

The prospect of fully networked, interoperable broadcast technology seemed distant in 2012. Back then, work on interoperable audio standards was already developing in multiple different directions, and no progress had been made on any standards for transporting audio and video over IP.

To avoid a repeat of the chaos of the previous decade, the Joint Task Force on Networked Media (JT-NM) was formed by three key players in the global broadcast industry: the US's Society of Motion Picture & Television Engineers (SMPTE), the European Broadcasting Union (EBU) and the international Video Services Forum (VSF). They were later joined by the Advanced Media Workflow Association (AMWA).

All feared a costly future in which broadcasters would feel forced into upgrading to mutually incompatible IP-based systems, many of which would need to be retrofitted or completely replaced as standards gradually emerged.

Better, they reasoned, to work together so that systems could be interoperable to widely accepted standards from the very beginning.

The taskforce conducted detailed research amongst broadcast companies, manufacturers and practitioners of best practice to find out exactly what they might want from future IP-based video and audio standards, what their concerns were, and how they saw such systems working in practice.

They pooled the results and used them to derive what they called the Reference Architecture: a detailed model of how an as-yet-undefined IP-based broadcast future might operate.

They invited other standards organisations to draw on existing protocols and technologies to create new standards that would help realise the ideal. Following publication of the JT-NM's Reference Architecture in 2015, various broadcast industry bodies have come forward with proposed standards that meet the requirements in the Reference Architecture.

▪ AMWA created the Networked Media Open Specifications (NMOS), developing specifications that provide for the discovery and registration of audio and video hardware over IP networks. NMOS has been met with great acclaim and enthusiastically adopted by manufacturers in the broadcast industry.

▪ SMPTE continued to add to its existing SMPTE 2022 standard, which it had begun publishing in 2007 to define protocols for transport of (initially heavily data-compressed) video signals over IP. In terms of realising an IP-based future, the most significant parts of the specification are published as 2022-6, which allows for the transport of high bit-rate and even uncompressed video over IP with embedded audio, and 2022-7, which provides standards for redundant transmission of the same video data over IP. This helps to safeguard against network-related signal interruptions.

▪ Following the publication of SMPTE 2022-6, the Video Services Forum released two further standards for high-resolution video and audio over IP which fit the requirements of the JT-NM's Reference Architecture. TR-04 recommends using the SMPTE 2022-6 standard for high-resolution video either with embedded audio, as in SDI signals, or with a separate stream of audio in AES67 format, along with sync and metadata streams.

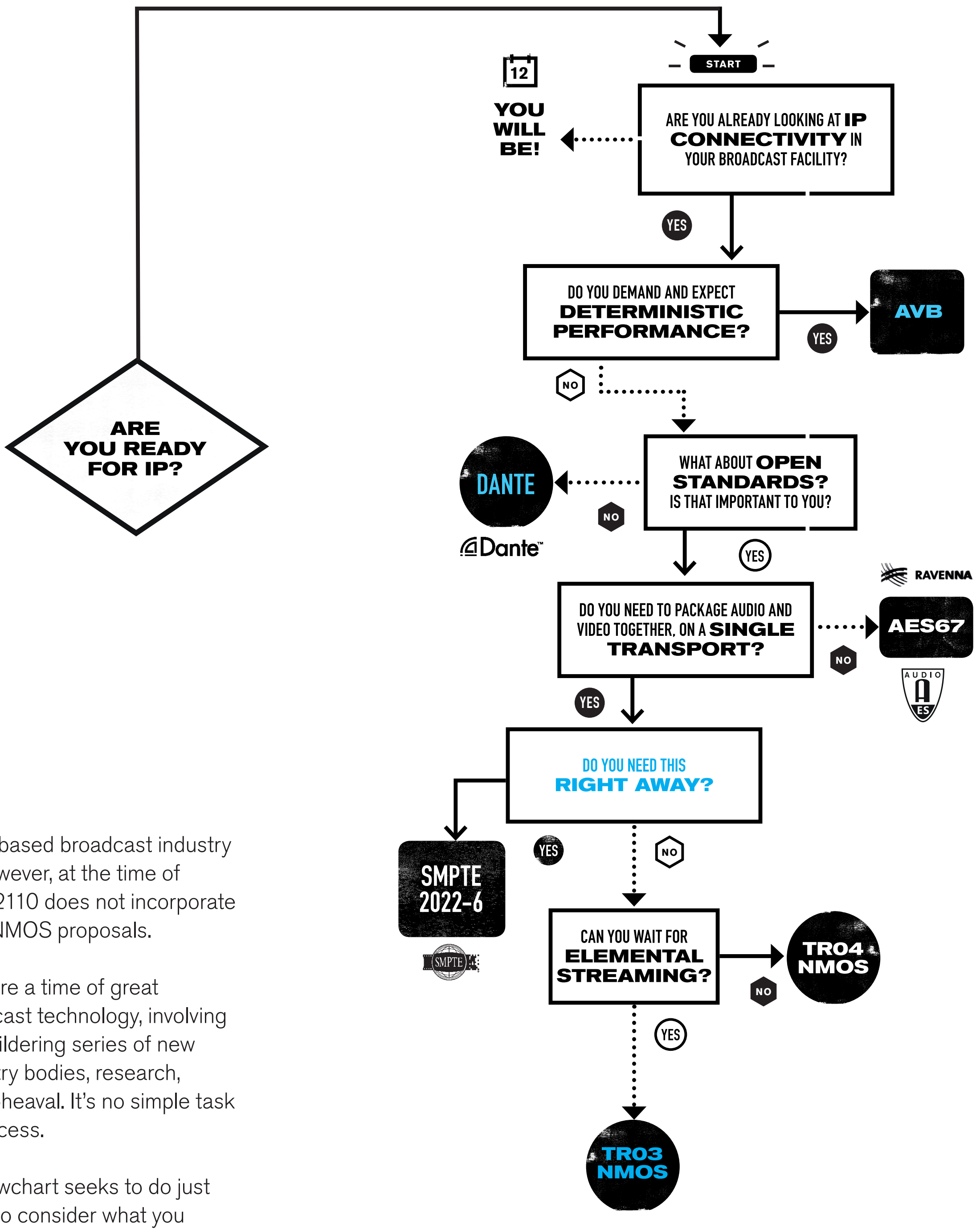
VSF's TR-03 describes the broadcast data being sent in completely segregated so-called 'elemental streams': HD or uncompressed video in RFC 4175 format, audio in AES67 format, and separate sync and metadata streams.

From an audio broadcast point of view, the TR-03 standard is more interesting as the audio is readily available as a discrete element; there is no requirement to de-embed audio from a video signal before it can be processed or mixed, or re-embed it afterwards, as is the case with SDI signals.

▪ AMWA has also supported regular meetings of broadcast technology providers at international plugfests. These events have allowed broadcasters, tech developers, design engineers and vendors to exchange information, collaborate and test their new prototype technologies, interfaces and protocols under condition of anonymity - so no-one manufacturer benefits too greatly from successful demonstrations, or is penalised by unsuccessful public demonstrations of their technologies.

▪ Following the creation of the VSF's TR-03 and TR-04 standards, another broadcast industry organisation, AIMS (the Alliance for IP Media Solutions) was formed to promote their adoption throughout the broadcast industry, together with the existing standards encapsulated in the VSF's Technical Recommendations: AES67 for audio and SMPTE 2022-6 for high-resolution video. As a result, it has been proposed that both VSF TR-03 and TR-04 should be adopted as a new SMPTE standard under the name SMPTE 2110.

This is expected to be ratified before the end of 2017, and the signs are that it may provide the basis for the high-resolution,



interoperable IP-based broadcast industry of the future. However, at the time of writing, SMPTE 2110 does not incorporate any of AMWA's NMOS proposals.

The late 2010s are a time of great change in broadcast technology, involving a potentially bewildering series of new standards, industry bodies, research, acronyms and upheaval. It's no simple task to clarify this process.

However, this flowchart seeks to do just that, asking you to consider what you might need from your broadcast facility over the next few years and leading you to a decision based on what is currently known about the standards currently under development.

CHAPTER FOUR: CONTROL, SYNC & METADATA OVER IP

With some of the standards required to take these changes forward still incomplete, forecasting the immediate future of IP-based broadcast systems is difficult — but it is possible to discern what might be possible in around a decade from now, once all the upheaval is past and these new technologies have bedded in, and to compare it to standard practices today.

Today, the broadcast of a major sporting event requires a large and extremely expensive truck (and sometimes multiple trucks) to arrive on-site at least 24 hours before the event, and for skilled staff to remain on site until the infrastructure is packed up afterwards. This workflow has unavoidably high attendant costs (travel, hotels and so on).

By contrast, here is how the workflow could be re-cast more affordably using IP-based infrastructure.

A few hours before the fixture starts, a single, small van could arrives on site, and one or two IP specialist engineers with some audio knowledge place some remotely controlled DSP on site and connect it to a robust IP connection.

At some venues, they'll also need to distribute cameras, microphones and interfacing for audio capture, but at many sites the mics and cameras may be pre-installed. There may be no cameramen or dedicated audio staff to place the mics as it will be possible to remotely control, steer and reposition them.

Two hours before the match, skilled cameramen, vision and sound mixing engineers drawn from personnel situated all over the globe, some at home, some at small broadcast facilities and some at large scale traditional studios, receive emails with access codes and privileges for the streams they need to access for video and audio monitoring, foldback mixes, and the raw audio and video from the event site. Shortly before the specified event time, they connect their respective camera, vision or audio mix control surfaces to their local network access point, and produce the broadcast of the event as though from an on-site OB vehicle today.

Afterwards, IP technicians disconnect the DSP hardware, pack it and any cameras or microphones back in the van, and leave. The virtual production 'team' disbands immediately; if required, personnel can move straight on to producing another event in a completely different part of the world, because they don't need to travel anywhere first.

The broadcaster fills its schedules with many more types of varied programming and live events produced in this way, because the associated costs are no longer only viable for the biggest, most high-profile events.

At the end of the first edition of the Calrec Audio Networking Primer, we wrote:

"If this Primer has a message, it's to emphasise that networking technology is on the brink of delivering unparalleled audio and video integration to broadcasters in many areas — even if some aspects still remain just out of reach."

Today, with internationally accepted technical standards close to ratification, and manufacturers nearly all pulling in the same direction, the future described above is closer than ever.

Calrec Audio Ltd
Nutclough Mill
Hebden Bridge
West Yorkshire
England UK
HX7 8EZ

Tel +44 (0)1422 842159
Fax +44 (0)1422 845244
Email Enquiries@calrec.com