Emerging Technology Trends Report

AES Technical Committee on Network Audio Systems
November 2011 Editor, Tim Shuttleworth; Tim@Renkus-Heinz.com

This document is a compilation of contributions from numerous members of the Technical Committee on Networked Audio Systems. The committee has identified the following important topics related to emerging audio networking technologies. Technologies which have emerged since the last published Emerging Trends Report from the committee in 2007 are included. To provide structure to the report items are discussed in order of their maturity; commercialized technologies implemented in products available for purchase being discussed first and embryonic concepts in early development come up last. Other categorizations referred to in this document are consumer market orientation versus professional market focus, as well as media transport methods versus command and control protocols.

Dante

Dante is a media networking solution developed by Audinate. In addition to providing basic synchronization and transport protocols Dante provides simple plug and play operation, PC sound card interfacing via software or hardware, glitch free redundancy, support for AVB and support for routed IP networks. The first Dante product arrived in 2008 via a firmware upgrade for the Dolby Lake Processor and since then many professional audio and broadcast manufacturers have adopted Dante.

From the beginning Dante implementations have been fully IP based, using the IEEE 1588-2002 standard for synchronization, UDP/IP for audio transport and are designed to exploit standard gigabit Ethernet switches and VoIP-style QoS technology (e.g. Diffserv). Dante is evolving with new networking standards. Audinate has produced versions of Dante that use the new Ethernet Audio Video Bridging (AVB) protocols, including IEEE 802.1AS for synchronization and RTP transport protocols. It is committed to supporting both IEEE 1733 and IEEE 1722. Existing Dante hardware devices can be firmware upgraded as Dante evolves, providing a migration path from existing equipment to new AVB capable Ethernet equipment.

Recent developments include announced support for routing audio signals between IP subnets and the demonstration of low latency video. Audinate is a member of the AVnu Alliance and the AES X192 working group. More information is available on the Audinate web site – www.audinate.com.

Q-LAN

Q-LAN is a third-generation networked media distribution technology providing high quality, low latency and ample scalability aimed primarily at commercial and professional audio systems. Q-LAN operates over gigabit and higher rate IP networks. Q-LAN is a central component of QSC's Q-Sys integrated system platform. Q-Sys was introduced by QSC Audio Products in June 2009. Q-LAN carries up to 512 channels of uncompressed digital audio in floating point format with a latency of 1 millisecond. Source: Q-LAN whitepaper - http://tinyurl.com/qlanwp

EBU N/ACIP

The European Broadcasting Union (EBU) together with many equipment manufacturers has defined a common framework for Audio Contribution over IP in order to achieve interoperability between products. The framework defines RTP as a common protocol and media payload type formats according to IETF definitions.

SIP is used as signaling for call setup and control, along with SDP for the session description. The recommendation is currently published as document EBU Tech 3326-2008 and can be downloaded from the ACIP website - http://www.ebu-acip.org/.

Audio Video Bridging

The Audio Video Bridging initiative is an effort by the IEEE 802.1 task group working within the IEEE standards organization which brings media-ready real-time performance to Ethernet networks. The IEEE is the organization that maintains Ethernet standards including wired and wireless Ethernet (principally 802.3 and 802.11 respectively). AVB adds several new services to Ethernet switches to bring this about. The new switches interoperate with existing Ethernet gear but AVB-compliant media equipment interconnected through these switches enjoy performance currently only available from proprietary network systems.

AVB consists of a number of interacting standards:

- 802.1AS - Timing and Synchronization
- 802.1Qat - Stream Reservation Protocol
- 802.1Qav - Forwarding and Queuing
- 802.1BA - AVB System
- IEEE 1722 - Layer 2 Transport Protocol
- IEEE P1722.1 - Discovery, enumeration, connection management and control
- IEEE 1733 - Layer 3 Transport Protocol

AVB standardization efforts began in earnest in late 2006. As of November 2011, all but the P1722.1 work have been ratified by the IEEE.

A LinkedIn group exists to discuss the technology - http://www.linkedin.com/groups?gid=2056907

AVnu, an alliance of interested companies promoting adoption of AVB technology. Additional information is available at the AVnu website – www.avnu.org

RAVENNA

A consortium of European audio companies has announced an initiative called RAVENNA for real-time distribution of audio and other media content in IP-based network environments. RAVENNA uses protocols from the IETF’s RTP suite for media transport. IEEE 1588-2008 is used for clock distribution. Performance and capacity scale with the capabilities of the underlying network architecture. RAVENNA emphasizes data transparency, tight synchronization, low latency and reliability. It is aimed at applications in professional environments, where networks are planned and managed.

All protocols and mechanisms used within RAVENNA are based on widely deployed and established methods from the IT and audio industry or comply with standards as defined and maintained by international standardization organizations like IEEE, IETF, AES and others. RAVENNA can be viewed as a collection of recommendations on how to combine existing standards to build a media streaming system with the designated features.

RAVENNA is an open technology standard without a proprietary licensing policy. The technology is defined and specified within the RAVENNA partner community, which is led by ALC NetworX and supported by numerous well-known companies from the pro audio market. More information is available at the RAVENNA web site – http://ravenna.alcnetworx.com/

AES X192

Audio Engineering Society Standards Committee Task Group SC-02-12-H is developing an interoperability standard for high-performance media networking. The project has been designated “X192”.

High-performance media networks support professional quality audio (16 bit, 48 kHz and higher) with low latencies (less than 10 ms) compatible with live sound reinforcement. The level of network performance required to meet these requirements is achievable on enterprise-scale networks but generally not on wide-area networks or the public internet.

The most recent generation of these media networks use a diversity of proprietary and standard protocols. Despite a common basis in Internet Protocol, the systems do not interoperate. This latest crop of technologies has not yet reached a level of maturity that precludes changes to improve interoperability.

Technology	Purveyor	Date introduced	Synchronization	Transport
RAVENNA	ALC NetworX	In development	IEEE 1588-2008	RTP
AVB	IEEE, AVnu	In development	IEEE 1588-2008 advanced profile (IEEE 802.1AS)	Ethernet, RTP
Q-LAN	QSC Audio Products	2009	IEEE 1588-2002	UDP
Dante	Audinate	2006	IEEE 1588-2002	UDP
LiveWire	Telos/Axia	2004	Proprietary (native) , IEEE 1588 (in development)	RTP

The X192 project endeavors to identify the region of intersection between these technologies and to define an interoperability standard within that region. The initiative will focus on defining how existing protocols are used to create an interoperable system. It is believed that no new protocols need be developed to achieve this. Developing interoperability is therefore a relatively small investment with potentially huge return for users, audio equipment manufacturers and network equipment providers.

While the immediate X192 objective is to define a common interoperability mode the different technologies may use to communicate to one another, it is believed that the mode will have the potential to eventually become the default mode for all systems.

The X192 interoperability standard will be compatible with and receive performance benefits from an AVB infrastructure. Use of the standard will allow AVB implementations to reach beyond Ethernet into wider area applications.

While the initial X192 target application is audio distribution, it is assumed that the framework developed by X192 will be substantially applicable to video and other types of media data.

WAN Based Telematic/Distributed Performance and Post Production

Telematic or distributed performances are events in which musicians perform together synchronously over wide area networks, often separated by thousands of miles. The main technical challenge associated with these events is maintaining sufficiently low latencies for the musicians to be able to play together, given the distances involved. Emerging enabling technologies such as the low latency codecs CELT which stands for "Constrained Energy Lapped Transform", Opus, a merging of CELT and SILK (a Skype codec) as well as ULD which refers to “Ultra-Low-Delay” allow streaming over DSL or cable end point connections rather than high-bandwidth managed networks, such as Internet2, which are recently more commonly used. For more information on these codecs visit http://www.celt-codec.org/docs/ and http://opus-codec.org/

Another wide area networked emerging use case is streaming audio for cinema post production, in which studios and post-production facilities are connected with one another via high-bandwidth managed fiber networks. This allows studios to see and hear the latest version of a film in post-production without having to physically move the assets to the studio or use a file-transfer system. Real-time streaming of uncompressed audio and video also allows greater collaboration between directors and post-production facilities and between different departments in the post-production process.

Networked post-production uses two methods (at present) for streaming audio: when audio is streamed independent of video, hardware Layer-3 uncompressed audio-over-IP devices are used. When audio is streamed along with video, it is embedded in an HD-SDI video stream, and the stream is networked using a video codec. The former case is primarily used for audio post-production, in which the audio engineers are mixing to a poor-quality version of the video; the video is then sourced locally at all locations, and the audio synced to it. Control information is streamed between all nodes using high-definition KVM-over-IP devices, along with MIDI-based control surfaces connected via Apple's MIDI Network Setup. KVM over IP is a server management technology. See http://en.wikipedia.org/wiki/KVM_switch (Streaming of Ethernet-based control surfaces is forthcoming.) Videoconferencing to allow collaboration uses either H.323 devices or the same codec used to stream content video. Clock synchronization between nodes can be accomplished either with the hardware audio-over-IP devices, which usually stream clock information, or with GPS-based sync generators at each node.

XFN Command and Control Protocol

XFN is an IP-based peer to peer audio network control protocol, in which any device on the network can send or receive connection management, control, and monitoring messages. The size and capability of devices on the network will vary. Some devices will be large, and will incorporate extensive functionality, while other devices will be small with limited functionality. The XFN protocol is undergoing standardization within the AES, and AES project X170 has been assigned to structure the standardization process. A draft standards document has been written and presented to the SC-02-12 working group for approval.

Essential to the XFN protocol is the fact that each parameter in a device shall be addressable via a hierarchical structure that reflects the functional layout of the device. For example, there might be an input section on a mixing console that has a number of channel strips. Each channel strip might have a gain control, a block of equalizers, and a fader control. Each equalizer block will have its own structuring. Each of these groupings is considered to exist at a separate level. At the lowest level of any device structure are parameters. An XFN message accesses a parameter in a device by providing a hierarchical address that models the position of the parameter within the device, and thereby allows access to it. Each device implements an XFN stack that will parse such a message structure, and thus be able to locate the parameter.

There are situations where it is not efficient to have a control application request data values from a destination device. Such a situation exists, for example, when values such as amplitude and temperature need to be continuously displayed via meter bars. In order to make such a process efficient, and to more accurately model the requirements, a 'Push' mechanism has been created. A control application indicates to a destination device which parameters' values it is interested in receiving, and at what intervals.

A "wildcard" mechanism allows for control over multiple parameters. If the address of a message contains a wildcard parameter at any particular level, then the addressing applies to all possibilities at that level. In this way a single command may affect a large number of parameters. This wildcard mechanism also enables the discovery of devices and the enumeration of a device's parameters.

Apart from addressing a parameter via its hierarchical position in the device, it is also possible to address it via a unique identifier (ID). Each parameter has an ID that is unique for that parameter on the device that hosts it. This ID can be obtained by addressing the parameter and requesting it. Further messages can replace the hierarchical address, and simply provide the ID to access the parameter, thereby reducing message bandwidth.

The XFN protocol allows parameters across a network to be joined into groups. Each parameter can hold a list of other parameters on the same or different devices, to which it is joined. If the parameter is modified by a message, then that same message is directed at all parameters within its list. For example, a fader on a mixing console would typically have a parameter that represents its position. This parameter may be joined to fader parameters on other mixing consoles, and possibly to gain parameters on breakout boxes. When the single fader is moved, its group list would be scanned and messages sent to all joined parameters. In order to allow for the joining of disparate parameters, XFN incorporates the concept of 'global units'. Most XFN messages carry parameter values in terms of global units, and it is up to the device to convert these global units into units that are appropriate for the parameter being modified.

A situation that often plagues control applications is the transfer of large batches of data from networked devices, and the graphical display of this data on the controller. XFN has dealt with this problem by creating the Universal Snap Group (USG) mechanism, whereby a controller specifies the batch of parameters it is interested in from a device. The device keeps one or more USG buffers, and on request from the controller packs the parameter values in as efficient form as possibly for receipt by the controller.

Lastly, the XFN protocol incorporates the definition of "modifiers", whereby any message can have its hierarchical message structure modified at any one of the levels. The modifier may for example increment the channel number level of all messages entering a device, and thereby achieve immediate control over a second block of channels. Further modifiers allow for the modification of message values and can enable automation by storing time tagged event lists. More information can be found at http://www.aes.org/standards/meetings/init-projects/aes-x170-init.cfm

Home Broadband Audio-Over-IP and Home Wireless LAN

Home broadband connections are increasing in speed, up to a typical rate, worldwide, of about 2Mbps. This is sufficient for streaming audio services to produce a good performance, mostly using 256kbps WMA or AAC, which yields pretty good quality at a low bit rate.

Use of Wireless LANs in the homes, mostly WiFi, with some proprietary systems is increasing. IEEE802.11g routers and devices are realizing faster throughput rates, while IEEE802.11n achieves improved range, improved QoS, and speeds which exceed the needs of low bit rate compressed audio streaming. Two eco-systems co-exist at the moment. The first is the Digital Living Network Alliance (DLNA) http://www.dlna.org/home, which focusses on interoperability between devices using UPnP (Universal Plug and Play) [http://www.upnp.org/] as the underlying connectivity technology. DLNA is becoming available in more and more devices, such as PC servers and players, digital televisions with network connectivity, network attached storage (NAS) drives and other consumer devices. The second eco-system is Apple AirPlay which allows iTunes running on a PC or MAC to stream audio to multiple playback devices. AirPlay also supports streaming directly from an iOS device (iPhone, iTouch, iPad) over WiFi to a networked audio playback device. Both ecosystems are driving the rapid acceptance of audio networking in the home.

Cloud computing, in particular cloud storage of audio content, is another emerging trend. The increasing popularity of premium audio services, for example Rhapsody, Pandora, Last.fm, and Napster, are driving a trend away from users the need to keep a copy of their favorite music in the home or on a portable device. Connection to the internet allows real time access to a large variety of content. Apple is also driving this trend with iCloud, released with iOS5. Consumer devices are becoming more complicated and connecting devices to the network has been difficult for users, resulting in many calls to tech support. The good news is that devices are becoming easier to set up. The WiFi Alliance has created an easy setup method call WiFi Protected Setup (WPS). This makes attaching a new device onto the home network as easy as pressing a button, or entering a simple numeric code.

Another trend driven by the adoption of home wireless LAN technologies is in the user interface (UI) of networked audio devices. More and more audio products are using the iPhone or iPad as the primary method of device control, via the home WiFi network. Some commentators are even announcing the death of the infrared remote control. Consumer Audio/Video Receiver manufacturers such as Denon and Pioneer offer free iPhone/iPad apps which allow complete, and intuitive control of their devices. This leads to another emerging trend, that of the display-less networked audio player. Once the player can be conveniently controlled from your smartphone, why should the device continue to include an expensive display and user controls? Display-less high end audio players are already selling well (for example B&W Zeppelin Air). Such display-less networked audio players will become ubiquitous and be available for under $100.

Open Control Architecture Alliance (OCA)

The Open Control Architecture Alliance has been formed by a group of professional audio companies who are working in different product markets and represent a diverse cross section of vertical market positions and application use-cases. Each of the companies realized that relying solely on proprietary solutions for media networking system controls made interoperability with other manufacturers' equipment or across application domains difficult.

The member companies agreed that an open standardized control architecture was not only possible, but should be created and made available as an open, public standard that could be available to any participant in the audio market in order to facilitate an improved environment for the entire AV industry. It is the stated mission of the OCA Alliance to secure the standardization of the Open Control Architecture (OCA), as a media networking system control standard for professional applications. OCA in its current form is a Layer 3 protocol that has been created by Bosch Communications based around the earlier (abandoned) command and control protocol AES-24.

The Alliance has been formed to complete the technical definition of OCA, then to transfer its development to an accredited public standards organization.

The founding group of OCA members is proceeding to complete the OCA specification and prepare it for transfer to a public standards organization without inviting new active members but welcomes any interested parties to join as an Observer Member.

OCA Alliance pre-released technical documentation in October 2011. Observer members have access to draft documents and pre-releases of technical documents. They are encouraged to give their feedback to the Alliance at an early stage of development. To enroll as an "Observer" contact info@oca-alliance.com

The OCA definition has three parts:

- An Architectural Framework, designated as OCF. OCF defines the set of structures and mechanisms upon which the rest of OCA rest.

- A Class Structure, designated as OCC. OCC is object-oriented. It is an expandable, evolvable hierarchical structure which defines OCA's repertoire of control functions.

- A suite of Protocol Definitions, which are designated as OCP.1, OCP.2, et cetera. Each protocol definition describes an implementation of OCA for a particular network type. At present, only OCP.1 exists. It describes the implementation of OCA for standard TCP/IP networks. Future protocol definitions will be created for USB and other interconnection methods.

These levels are not protocol layers, but simply sets of specifications upon which other specifications depend.

For more information visit http://oca-alliance.com/index.html

International Telecommunications Union: Future Networks

ITU-T Q21/13, Study Group SG13 is looking at "Future Networks" which are expected to be deployed during 2015-2020. So far an "objectives and design goals" document has been published (Y.3001, see <http://www.itu.int/rec/T-REC-Y.3001-201105-P>), and the study group is working on virtualisation and energy saving (soon to be published as Y.3011 and Y.3021 respectively) and on identifiers. These deliberations are at a very early stage and a clear direction is not yet apparent. The underlying technology could be a "clean slate" design, or it could be a small increment to NGN (Next Generation Network, which is based on IPv6). For details see http://www.itu.int/ITU-T/studygroups/com13/index.asp .

International Electrotechnical Commission/International Standards Organization: Future Network

ISO/IEC JTC1/SC6/WG7 is also working on Future Network, and also expects deployment during 2015-2020. Their system will be a "clean slate" design with a control protocol which is separate from the packet forwarding. It will support multiple networking technologies, both legacy technologies such as IPv4 and IPv6 and also new technologies able to provide a service suitable for the most demanding live audio applications.

It will carry two kinds of data, “synchronous” and “asynchronous”. For synchronous data there is a set-up process (part of the control protocol) during which resources can be reserved. The application requests QoS parameters (delay, throughput, etc) appropriate to the data to be sent, and the network reports the service the underlying technology is able to provide.

Asynchronous data can use a similar set-up process, or can be routed in a similar way to Internet Protocol. Thus it will also be efficient at carrying protocols such as TCP, and will interoperate with IP networks. This provides a migration path from current systems.

Some of the drafts are available from <http://www.iec62379.org/FN-standardisation.html>, including 29181 part 3 which specifies requirements for the switching and routing technology and 62379 Part 5-2 which specifies a possible control protocol.