RAVENNA

V 1.0

White Paper

This White Paper describes how RAVENNA and AES67 relate to each other.

ALC NetworX

ALC NetworX GmbH

Am Loferfeld 58
81249 Munich
GERMANY

Fon: +49 (89) 44236777-0
Fax: +49 (89) 44236777-1

@: info@alcnetworx.de

www.alcnetworx.de

Version information:

V1.0 - 2014-01-14

Any information contained in this document may be changed or amended at any point in time without further notice. The document may also be replaced as a whole.

RAVENNA & AES67

V 1.0

1 MANAGEMENT SUMMARY

RAVENNA is an open technology for real-time distribution of audio and other media content in IP-based network environments. Utilizing standardized network protocols and technologies, RAVENNA can integrate and operate on existing network infrastructures. Performance and capacity are scaling with the capabilities of the underlying network architecture. Emphasize is put on data transparency, tight synchronization, low latency and reliability. It aims at applications in professional environments, where networks are planned and managed, and where performance has to surpass the requirements of consumer applications.

As an open technology, the functional principles are publically available and RAVENNA technology can be freely implemented and used without any proprietary licensing policy[1]. Numerous industry partners are already supporting the RAVENNA technology[2].

In September 2013, the AES has published the AES67 "Standard on High-performance Streaming Audio-over-IP Interoperability", which defines guidelines and mechanisms to achieve interoperability between different IP-based real-time streaming solutions. Since RAVENNA's fundamental operational principles, protocols and formats are in-line with what has been defined in AES67, RAVENNA is already compatible with AES67.

1 Implementation of RAVENNA may require obtaining licenses from third party IP holders when implementing certain standards employed by RAVENNA, i.e. implementers most likely will need to obtain a license from Agilent Technologies when implementing PTP according to IEEE1588 (whereupon such license is granted under RAND licensing policy).

2 See http://ravenna.alcnetworx.com/partners for current RAVENNA partner list.

2 RAVENNA

2.1 Overview

RAVENNA is a solution for real-time distribution of audio and other media content in IP-based network environments. Utilizing standardized network protocols and technologies, RAVENNA can operate in existing network infrastructures. Performance and capacity are scaling with the capabilities of the underlying network architecture. Primarily targeting the Professional Broadcast Market, RAVENNA matches Broadcaster's requirements for low latency, full signal transparency and high reliability. Fields of application include (but are not limited to): in-house signal distribution for Broadcasting Houses and other fixed installations, flexible setups at venues and live events, OB van support, inter-studio connectivity across WAN links and production facilities.

2.2 Why Audio-over-IP?

The tremendous scale of manufacturing in the data network world ensures large cost-savings on equipment. Using network-based solutions for media transport enables broadcasters to leverage their existing infrastructure and achieve greater flexibility in content sharing and networking configuration.

As an IP-based solution, RAVENNA is based on protocol levels on or above layer 3 of the OSI reference model. IP can be transported on virtually any LAN and is also used as the base layer for communication across WAN connections (including the internet). Although Ethernet will be deployed in most cases as underlying data link layer, IP is infrastructure-agnostic in general and can be used on virtually any network technology and topology. In contrast to Ethernet AVB - an extension to the IEEE802.1 Ethernet standard suite of protocols – which requires full native support by the network infrastructure through deployment of all-new switches, RAVENNA can be operated on most existing network infrastructures of any size – from small studio setups to large corporate environment.

Since IP technology was originally designed to forward a large number of data packets without constraints towards real-time, it bears some handicaps for media transport applications. These handicaps are addressed by utilizing standardized QoS mechanisms usually available in a manageable network infrastructure. Despite this handicap, the benefits of using IP for media transport are increasingly too persuasive for broadcasters and service providers to be ignored.

2.3 RAVENNA Key Features

- Suitable for audio, video and other media data

  RAVENNA’s stream transport format is based on RTP, which makes it payload-agnostic. Any kind of payload can be transported within RTP packets, hundreds of standardized formats already exist.

- Precise media clock distribution

  RAVENNA does not distribute media reference clocks, but relies on distribution of absolute time. Thus, any desired media clock can exactly be derived at any participating device – even anywhere in the world when the time master device is referenced to GPS. Where required, the media clock phase alignment can even match the tight requirements of AES11 for master clocks.

- Sample-accurate play-out alignment across all nodes on the network

  Since RAVENNA's synchronization is based on absolute time, it is possible to achieve a perfect sample-accurate play-out or processing alignment between all streams anywhere in the network.

- Concurrent support of multiple media clocks and data formats

  While most networked solutions are limited to a single payload format or sample rate selected for the whole network, RAVENNA offers the freedom to distribute streams with different sample rates and / or data formats concurrently across the network – no sample rate conversion is required.

- Full bit transparency

  Since no sample rate conversion is applied, all signals are transported with full bit transparency. I.e., AES3 signals can be transported with all meta data bits in place, and Dolby E® signals stay intact.

- Low latency

  While standard latencies on larger corporate networks may stay in the upper single-digits milliseconds range, careful design of network topology and traffic patterns, paired with appropriate stream and QoS configuration also allows sub-milliseconds latencies, if required (i.e. in live monitoring situations). If required, even MADI performance can be achieved with reserved use of selected network segments.

- Operation in shared traffic environment

  Since RAVENNA's QoS scheme is based on the widely available *Differentiated Services* (DiffServ) mechanism, the network resource can be shared with other traffic. RAVENNA's real-time packets flow side-by-side with conventional data traffic, receiving the required prioritization through adequate QoS configuration of participating switches.

- Fully redundant operation

  While larger networks may offer self-healing mechanisms through standardized re-routing protocols, these procedures usually may take up to a few seconds and won't conclude without noticeable packet loss. While tolerable in certain application, RAVENNA offers the option of fully redundant stream transport on different physical network paths whereby packet loss or delay on one network path won't result in any audible interruption at all.

2.4 Fields of Application

While primarily targeting the professional broadcast market, RAVENNA is also suitable for deployment in other pro audio market segments like live sound, install market and recording.

Possible fields of application include (but are not limited to) in-house signal distribution in broadcasting houses, theaters, concert halls and other fixed installations, flexible setups at venues and live events, OB van support, inter-facility links across WAN connections and production & recording applications.

Although in a first approach RAVENNA is focusing on audio distribution, the same technology will be utilized to support video transport at a later stage.

2.4.1 Typical Fields of Application in the Broadcast Market

Typical fields of potential deployment are all areas where audio routing mixing or mixing is involved. The major areas comprise of:

- Broadcasting centers: In broadcasting centers usually a large number of signals is concentrated and managed. Typically several hundred up to thousands of sources and destinations are maintained. The main task is to distribute and route incoming signals to specific destinations. Since in traditional approaches resources are usually limited, a very important task is to manage these resources over time. Routes have to be ordered, checked against conflicts, have to be switched in time, and of course, have to be maintained and checked for desired quality.

  These tasks are usually being served by centralized, cost-intensive audio and / or video routing systems, which imaginably could be replaced by an inexpensive and far more flexible approach based on RAVENNA technology, specifically where hundreds of PC-based workplaces require access to selected media sources.

- Regional studios: Although similar to broadcasting centers, regional studios are much more focused on the management of their local sources and their appropriate in-house distribution. Feeds coming from or going to the related broadcasting center are usually distributed across (permanently) leased lines, SDH networks or ATM WAN connections. Some broadcasters also maintain "corporate networks" (utilizing one of the above technologies), which basically will give them the freedom of transferring any type of signal or data without interference from any "public" traffic. Part of this expensive infrastructure could also be replaced by RAVENNA network technology.

- OB vans: OB vans are "mobile studios". They are packed with any type of equipment which is required to feature a remote production (recording, live show etc.). Beside a "central" switching and mixing infrastructure, an OB van of course needs a large number of access points for incoming sources, outgoing feedback and communication streams as well as a stable uplink connection to a related studio (regional studio or broadcasting center).

  The main disadvantage of conventional infrastructure originates from the necessity that most connections to the venue have to be set up physically and administered logically a certain time before they are to be used. Since each venue or event will have a different setup, network-based RAVENNA technology can save a lot of valuable installation and commissioning time.

But also the studio uplinks, which at current most often are leased satellite lines, may benefit cost-wise from IP-based RAVENNA technology.

- Venues: A venue may be any permanent, semi-permanent or temporary installation which requires a professional audio infrastructure (e.g. a sports stadium, summer open-air stage, event installation etc.). These venues usually have a variety of different signal sources and sinks (microphones for ambience, on-stage, off-stage, wireless for multi-purpose use around the venue, reporter cabins, line feeds, monitoring signals etc.) which in best case are routed through a local "mini" switching center; in older installations you often can also discover independent "systems" (consisting of discrete cabling of incompatible type and signal format).

  Most venues also offer "outside" connections, if - for example - broadcasters (eventually with mobile OB vans) desire to access the infrastructure for specific events. This also includes monitoring and communications lines which have to be fed back into the venue infrastructure.

  Since the specific requirements clearly may change with each event, the replacement of "old" discrete infrastructure with a RAVENNA-based solution will offer many advantages towards flexibility, set-up time and costs - especially if the OB vans are equipped with a matching I/O technology.

2.4.2 Recording Market

Although in a recording setup many requirements are similar to those in broadcasting, a few requirements are essential:

- tight synchronization
- support for variable sample-rates
- low latency

It is essential for recording that all devices are tightly synchronized, so that sample-accuracy is guaranteed throughout the complete system setup. Thanks to the sophisticated synchronization scheme of RAVENNA, it is not only possible to guarantee system-wide sample-accuracy, but to provide phase synchronization precision as required for a word clock master according to AES-11.

Although within one recording session sample rates usually stay constant, they may change on a per-session basis. RAVENNA is sample-rate agnostic, it can even transport streams of different sample rates concurrently.

The low-latency requirement may be essential in a recording session when signal feedback (i.e. in-ear monitoring etc.) is required; RAVENNA's capability to configure latency numbers in the sub-milliseconds range is a perfect match.

Through the use of a RAVENNA Virtual Sound Card (RVSC) – a software-only emulation of “real” sound card –, Windows-based PC applications can participate directly in a RAVENNA network without the need for an extra sound card. While a RAVENNA Virtual Sound Card on a standard PC requires more relaxed latency settings, it is still possible to preserve the system-wide sample-accurate synchronization.

2.5 Fundamental Technical Principles

2.5.1 IP Transport

The RAVENNA suite of protocols is fully based on layer 3 of the OSI network abstraction model[3], the so-called "IP layer". All employed protocols are widely known and commonly supported protocols standardized by the IETF[4]. Since transport is based on IPv4, virtually any existing manageable switch can be used with RAVENNA.

While some employed protocols operate in unicast mode, multicast operation is a fundamental building block of RAVENNA, since synchronization, stream transport, device discovery and service advertisement all rely on multicast transport.

2.5.2 QoS

As different services can co-exist with RAVENNA on the same network, it needs to be ensured that RAVENNA traffic will be forwarded with expedited priority through the network.

For IP-based traffic, *Differentiated Services* (DiffServ)[5] is widely supported by most modern managed switches and has become the primary Layer 3 QoS mechanisms to provide different levels of service. DiffServ operates on the principle of traffic classification, where each data packet is assigned to a certain traffic class, which receives a configurable forwarding characteristic in the network switches. Through careful network administration it can be ensured, that RAVENNA traffic receives the required expedited forwarding treatment.

2.5.3 Synchronization

While simple streaming across a network can be achieved without any synchronization at all, in pro audio applications a tight synchronization between all devices and streams is absolutely mandatory. While playback synchronization in most applications requires sample accuracy, it has been the goal for RAVENNA to optionally provide superior performance by providing phase-accurate synchronization of media clocks according to AES-11; this would render the separate distribution of an external reference word clock throughout the facility or venue obsolete.

In RAVENNA, synchronization across all nodes is achieved through IEEE1588-2008[6] (Precision Time Protocol version 2 - often referred to as PTPv2), another standardized protocol which can be operated on IP. PTPv2 provides means for synchronizing local clocks to a precision in the lower nanoseconds range with reference to a related master clock - provided that all participating switches natively support PTPv2. But even without native PTP support, the achievable precision - while varying

3 https://en.wikipedia.org/wiki/OSI_model
4 Internet Engineering Task Force - http://www.ietf.org/
5 https://en.wikipedia.org/wiki/Differentiated_services
6 http://www.nist.gov/el/isd/ieee/ieee1588.cfm

depending on size and bandwidth utilization of the network - will be more than sufficient to reach sample accurate synchronization across all nodes.

Sample-accurate synchronization can even be reached across WAN connections, when local master clocks are synchronized to GPS as a common time reference.

2.5.4 Streaming

As IP has been chosen as a basis, it's only natural to use RTP (Real-time Transport Protocol) for streaming of media content. RTP, as defined by the IETF, is widely used and supported by numerous applications and comes with a large number of standardized payload formats. For RAVENNA, specifically RTP/AVP over UDP together with RTCP (the real-time transport control protocol) according to RFC 3550 is used.

Basic payload formats for audio are 16 and 24-bit @ 48 kHz with any desired number of channels. This would allow any standard media player to potentially link to a RAVENNA stream and monitor its content - even without knowledge or support of any of the other RAVENNA-specific features or methods. Of course the payload format is not restricted to those basic formats as with RTP a huge variety of different payload formats for audio and video is already defined; it is even possible to add vendor- or solution-specific formats (e.g. AES3 or 32-bit float), still preserving full RTP compatibility.

Streaming is supported both in unicast and multicast mode[7] on a per-stream basis providing the highest flexibility to match the distinct requirements of different applications. Unicast is preferred in situations where a certain stream needs to be transported to a single destination only or where the network infrastructure or application prohibits the use of multicast (e.g. across most WAN connections). On the other hand, multicast allows resource-efficient usage of network links and faster switching between available streams in situations, where a certain stream will be accessed at different locations.

2.5.5 Device Configuration and Service Advertisement & Discovery

In order to participate in an IP-based network communication, a device must obtain a unique IP address. Then, in order to be recognized by other RAVENNA nodes, a device must announce its existence and advertise information about available services (e.g. IP address and host name, supported protocols, access information, information about available streams etc.).

In order to support a wide range of application environments, RAVENNA supports three different methods for device configuration and service advertisement & discovery: In managed networks usually DHCP[8] and DNS[9] services are operated under management and control of a network administrator. In small networks, where usually no DHCP / DNS server is present, the Zeroconf[10] mechanism (also known as "Bonjour") - a fully automatic, self-configuring method - can be used for auto-IP assignment and network configuration. Fully manual configuration is also supported as a third option.

7 https://en.wikipedia.org/wiki/Multicast
8 https://en.wikipedia.org/wiki/Dhcp
9 https://en.wikipedia.org/wiki/Dns
10 https://en.wikipedia.org/wiki/Zeroconf, http://zeroconf.org/

Once configured, a RAVENNA device advertises its existence and available services on the network. Service advertisement & discovery in RAVENNA is based on the DNS-SD protocol. The mDNS mechanism (part of Zeroconf) is used by default; in larger or more complex environments, dedicated DNS servers can be used alternatively or in addition. Other devices can then discover the presence of a device and may retrieve information on the offered stream services using the RTSP/SDP protocol.

2.5.6 Stream Connection Management

A receiver can connect to any existing stream through RTSP / SDP[11] protocol. Again, this scheme is supported by most common media players (i.e. Windows Media Player, VLC media player et al). While RTSP is used for control communication between receiver and sender, the SDP record contains any relevant information about one or more streams - like stream name, payload formats, number of channels, access information etc. Although a typical RAVENNA SDP contains some specific extensions (i.e. clock domain and sync information), any non-RAVENNA-aware media player can still receive and play-out a RAVENNA stream by just ignoring the specific extensions.

2.5.7 Additional vendor-specific Control

Device configuration is suggested to be provided via local web service accessible through HTTP protocol. This has the advantage, that any device-specific configuration can be executed through any common web browser without the need for a vendor-specific utility. However, as IP provides a platform for operation of any type of protocol, vendor-specific commands and functions can be executed on the same network interface concurrently. Some RAVENNA devices are already supporting the open Ember+ control protocol[12].

2.6 Open Technology Approach

In the past, the pro audio market has seen numerous technological innovations created or invented by some of the most ingenious minds of our industry. Unfortunately, most of this valuable intellectual property ended up in being used as proprietary or patented technology. It was apparent that a new audio distribution technology will not gain significant market acceptance if it would not be supported by an ample number of different companies. Thus, ALC NetworX decided to make the underlying technology and mechanisms used with RAVENNA publicly available[13]. In order to emphasize this open approach, ALC NetworX has teamed-up with renowned companies from the Pro-Audio market, and a variety of RAVENNA-enabled devices are already commercially available.

2.7 Profiles as a Means for Interoperability

The full breadth of variability and flexibility offered by RAVENNA might be intimidating, particularly for the newcomer. This is true for both users and manufacturers. In such a situation, incompatibility may result if all individual parties exercise their freedom of choice independently, and arrive at sets of choices which do not overlap.

---

11 https://en.wikipedia.org/wiki/Session_Description_Protocol
12 http://code.google.com/p/ember-plus/
13 RAVENNA Operating Principles - Draft 1.0 - http://ravenna.alcnetworx.com/infos-downloads/white-papers.html

Profiles are a way around this. They apply to a certain range of applications, and collect a minimum set of features these applications are likely to require from devices and from the network. Thereby they establish a baseline of compatibility the user can rely upon without having to check the details.

RAVENNA defines a set of profiles for application areas which are regarded as important. Manufactures have the freedom to implement these profiles in their devices as appropriate for their designated use cases. Of course devices can support multiple profiles concurrently, thus allowing a wider field of application and increased interoperability.

A *Generic Profile* has been defined to describe a base line of features considered to be indispensable for most devices; it contains a small set of requirements almost every device should be able to meet, and can therefore be expected by the user to be available in most RAVENNA devices. This is enough for some common use cases where particularly stringent requirements are not present, and can also serve as the default setting in more capable devices. Examples of commonly used stream formats being defined in the Generic Profile include:

- Low-latency stereo stream with 16 or 24 bit at 48 kHz sampling rate and a packet time of 1 ms
- surround stream with 16 or 24 bits, 48 kHz, 1 ms packet time
- standard stereo stream with 16 or 24 bits, 48 kHz, 4 ms packet time

Other profiles cover high-performance operation with MADI-like channel assembly and sub-milliseconds latencies or backbone interconnectivity with ultra-low latencies in the microseconds range and channel counts beyond 256. More profiles can be defined as application requirements demand. And manufacturers have the freedom of adding their own formats and profiles for their individual needs (i.e. DSD/DXD audio transport with 384 kHz as supported by Merging Technologies' Horus & Pyramix devices).

3 AES67

3.1 Motivation / Background

A number of IP-based audio networking systems have been deployed into the market prior of the development of RAVENNA. Despite being based on IP and using alike synchronization mechanisms and payload formats, none of these systems were interoperable, and no recommendations for operating these systems in an interoperable manner were in existence.

3.2 Standardization Work

In autumn 2010 the AES inaugurated the SC-02-12-H Task Group on *High-performance Streaming Audio-over-IP Interoperability*. This project has been designated AES-X192[14] and was chaired by Kevin Gross of AVA Networks, Boulder, Co., who is a highly-respected industry veteran on media networking (he is the inventor of *CobraNet*[15]). The approach was to identify common mechanisms and protocols and to suggest and standardize means for interoperability, based on existing protocols and mechanisms.

The technical scope was on higher performing networks which allow high-quality (16 bit / 48 kHz and higher), high-capacity (up to several hundred channels) and low-latency (less than 10 milliseconds) digital audio transport. The level of network performance required to meet these requirements is typically available on local-area networks and achievable on enterprise-scale networks.

Since its inauguration at the end of 2010, the Task group membership list grew up to more than 100 individuals, representing a prestigious spectrum of companies and organizations from the professional audio community. Through bi-weekly web conferences and several face-to-face meetings, the Task Group concluded on a 40+ pages draft, which had been handed over to the hosting SC-02-12 Work Group for further processing end of April 2013. Following a public commenting period, on September 11th, 2013 the AES finally published the ***AES67-2013 standard for audio applications of networks - High-performance streaming audio-over-IP interoperability[16].***

3.3 Ingredients

It became immediately apparent that the commonalities between all solutions and technologies considered did not cover enough ground for a complete interoperability definition. Thus, the work group had to define a set of mechanisms and protocols to be used for this new AES standard, following the proposition to use standardized and existing principles and protocols wherever possible.

To achieve a sound interoperability definition, AES67 had to address these functional areas:

- Synchronization – defines the mechanism for a common clock system
- Media clocks – defines which media clocks have to be supported and how they are related to the common clock system

14 http://www.x192.org/
15 http://www.cobranet.info/
16 http://www.aes.org/publications/standards/search.cfm?docID=96

- Transport – describes how media data is transported across the network
- Encoding and streaming – describes the means in which audio is digitized and formatted into the sequence of packets that constitutes a stream
- Stream description – required for connection management, specifies relevant stream information such as network addressing, encoding format and origination information
- Connection management – the procedure and protocols used to establish a media stream connection between a sender and one or more receivers
- Advertisement and discovery – the mechanism to make devices and streams discoverable on the network

3.3.1 Synchronization

The ability for network participants to share an accurate common clock distinguishes high-performance media streaming from its lower-performance relatives such as Internet radio and IP telephony. A common clock allows for an identical packet rate among all senders and receivers and thus for a fixed and determinable latency between sender and receivers. A common clock assures that all streams are sampled and presented at exactly the same rate. Streams running at the same rate may be readily combined in receivers. This property is critical for efficient implementation of networked audio devices such as digital mixing consoles.

Synchronization in AES67 is achieved by using IEEE 1588-2008 Precision Time Protocol (PTPv2) to establish a network-wide common wall clock. IEEE 1588-2008 defines a protocol for distributing absolute time between a Grandmaster and any number of slave clocks (Ordinary Clocks). The achievable precision is depending on various factors like network segment size, number of clients, traffic conditions and other network-specific parameters, but may well reach into the dual-digit nanoseconds range.

The protocol includes a *Best Master Clock algorithm* (BMCA), which results in the most accurate clock being elected to serve as the Grandmaster. Although not required, a Grandmaster clock is typically synchronized to an external traceable reference (i.e. GPS).

3.3.2 Media clocks

Media clocks are used by senders to sample and by receivers to play digital media streams. AES67 supports two sampling frequencies: 48 kHz and 96 kHz. The media clocks are derived from the synchronized, local wall clock time and thus have a fixed relationship to each other, i.e. the media clock for an audio stream sampled at 48 kHz advances exactly 48,000 samples for each elapsed second on the wall clock – at exactly the same pace in each participating node.

Figure 1 illustrates the synchronization and media clock generation scheme:

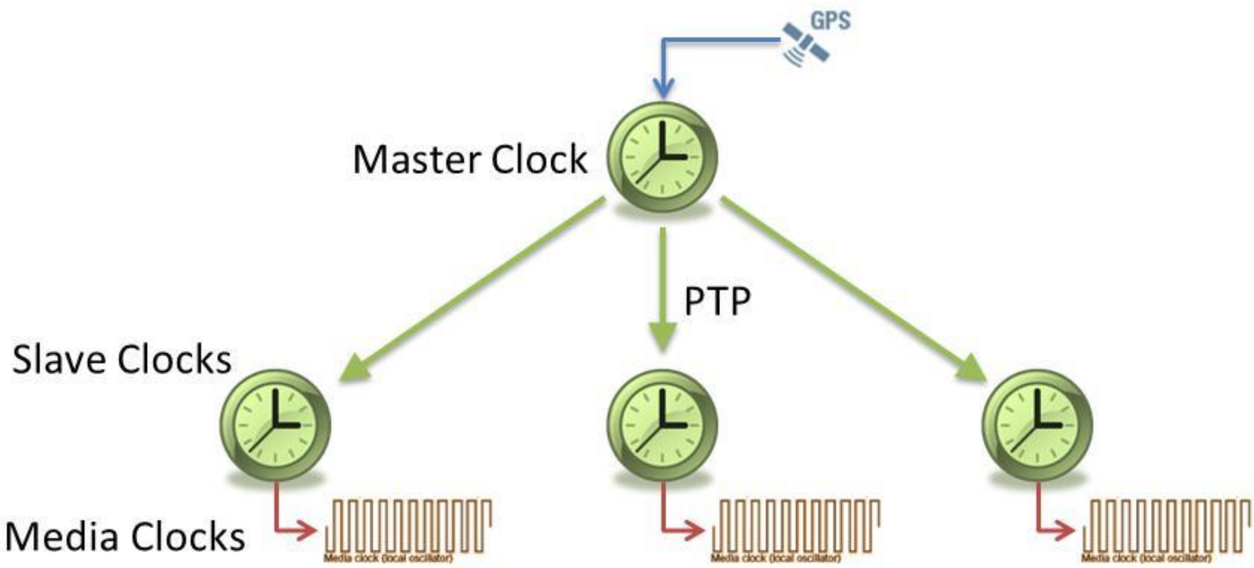

Figure 1: Synchronization & media clock model

3.3.3 Transport

Transport aspects describe how media data, once encoded and packetized, is transported across the network. Media packets are transported using IP version 4 as defined in RFC 791. Although care has been taken in design so as to facilitate future support for IPv6, it is currently outside the scope of AES67.

The Real-time Transport Protocol (RTP) in accordance with the AVT profile as defined in RFC 3550 / 3551 is used for media data transport. For maximum payload size, the standard 1500 byte maximum Ethernet MTU[17] is assumed. To assure future compatibility with IPv6, the maximum allowed RTP payload size has been limited to 1440 bytes. Streams can be transported unicast (one-to-one) or multicast (one-to-many), depending on the use case.

3.3.4 Encoding and Streaming Composition

Encoding describes the means in which audio is digitized and formatted into the sequence of packets that constitutes a stream. The following payload formats are supported by AES67:

- L16 – 16-bit linear format
- L24 – 24-bit linear format

All devices must support 48 kHz sampling rate and are recommended to support 96 kHz sampling rate.

Packet time defines the number of samples per RTP packet. Shorter packet times allow for lower latency but introduce overhead and high packet rates that may overtax some devices or networks. Longer packet times imply higher latency and require additional buffering which may not be available on memory-constrained devices. Interoperability is addressed by the requirement that all devices

17 https://en.wikipedia.org/wiki/Maximum_transmission_unit

must support a 1-millisecond packet time; further interoperability is encouraged through additional packet time recommendations (see table below).

Packet time	Samples @ 48 kHz	Samples @ 96 kHz	Notes
Required:			
1 ms	48	96	Required common packet time for all devices
Recommended:			
125 µs	6	12	Compatible with class A AVB transport
250 µs	12	24	High-performance, low-latency operation. Interoperable with class A and compatible with class B AVB transport.
333.33 µs	16	32	Efficient low-latency operation
4 ms	192	384	For applications desiring interoperability with EBU Tech 3326, transport over wider areas or on networks with limited QoS capability

Table 1: Required and recommended packet times

Streams may be composed of 1 to 8 audio channels, but more than 8 channels per stream may optionally be supported as well; however, the maximum number of channels per stream is limited by the packet time, encoding format and network MTU. The table below indicates the maximum channels count per stream at different encoding settings:

Format, sampling rate	Packet time	Maximum channels per stream
L24, 48 kHz	125 µs	80
L16, 48 kHz	250 µs	60
L24, 48 kHz	250 µs	40
L24, 48 kHz	333.33 µs	30
L24, 96 kHz	250 µs	20
L24, 48 kHz	1 ms	10
L24, 48 kHz	4 ms	2

Table 2: Maximum channel capacities per stream

3.3.5 Stream Description

Stream description is used by connection management to specify relevant information about each stream including network addressing, encoding and origination information.

SDP as specified in RFC 4566 is used to represent the relevant stream information for connection management. However, interoperability under this standard imposes additional SDP requirements and recommendations:

- Packet Time – the *ptime* attribute indicates the preferred packet time. If more than one packet time is supported, *maxptime* indicates the maximum permitted packet time.
- RTP clock, media clock and clock source – the clock source and the relationship of media clock to RTP clock are specified for each stream with additional attributes as specified in *draft-ietf-avtcore-clksrc*
- Payload types – as none of the encoding formats used within AES67 are defined as static payload types within RFC 3551, an *rtpmap* attribute is required for each stream to describe its current encoding formats

3.3.6 Quality of Service

On a network shared with unregulated non-real-time traffic, time-critical media traffic generally requires prioritized handling known as QoS (Quality of Service). For AES67, the widely deployed *Differentiated Services* method (DiffServ) as described in RFC 2474 is being used. DiffServ uses a tag in each IP packet header (the DSCP field) to mark packets according to their traffic class so that the network can easily recognize packets that need to be treated preferentially.

For AES67, at least three traffic classes have to be supported by the network. Devices should tag outgoing traffic with an appropriate DSCP value according this table:

Class name	**Traffic type**	**DiffServ class (DSCP value)**
Clock	IEEE 1588-2008 *Announce*, *Sync* and all *Delay request* & *Delay response* messages	EF (46)
Media	RTP and RTCP media stream data	AF41 (34)
Best effort	IEEE 1588-2008 signaling and management messages; connection management messages; other (non-AES67) network traffic	DF (0)

Table 3: QoS classes and DiffServ associations

Other traffic classes / DSCP markings may be used in certain network situations; however, these situations require specific knowledge about network transport policies and requirements of other traffic services in order to establish a suitable system configuration through administrative means.

3.3.7 Connection Management

Connection management is the procedure and protocols used to establish a media stream between a sender and one or more receivers.

Within AES67, connection management for unicast connections is accomplished using the Session Initiation Protocol (SIP)[18] as defined in RFC 3261. SIP is widely being used in IP telephony and by codec devices utilizing the ACIP protocol as specified in EBU Tech 3326[19].

SIP is conventionally used with the assistance and participation of SIP servers. Different types of servers perform different tasks for a SIP network. Servers may be located anywhere on the network where they are reachable by end stations. The use of servers creates a flexible and scalable connection management system.

For modest installations (i.e. with limited scale), where servers produce minimal benefit and the overhead of installing and configuring SIP servers introduces unnecessary complication, the server-less mode is appropriate. Server-less mode is used to perform connection management between user agents in direct peer-to-peer fashion.

In order to perform peer-to-peer connection management, the caller must have some means of determining network contact information (that is, host name or IP address) of the callee. In peer-to-peer connection management, all SIP messages are directed to the target device instead of the server. Under AES67, devices are required to support server-less operation. However, this does not release devices from requirements to operate in a normal SIP environment featuring servers. Specifically, devices will still need to attempt to discover, and register with, SIP registration servers and respond to messages originating from servers.

Multicast connection management may be accomplished without use of a dedicated connection management protocol (i.e. SIP). In this connection management scenario, a receiver can simply connect to an existing multicast stream by using the IGMP protocol[20], given that the relevant stream parameters (i.e. multicast address, stream format and clock parameters) are known by the receiver (i.e. through a discovery mechanism or administrative means).

3.3.8 Advertisement & Discovery

Advertisement and discovery is the network service which allows participants to build a list of the other participants or streams available on the network. Such a list can be presented to users to assist with connection management.

Device and stream discovery has intentionally been excluded from the AES67 standard works, as a number of different discovery mechanisms are available, with certain preference for one or the other in various network environments and application scenarios. However, interoperability can be achieved even without a common discovery mechanism by distributing the relevant parameters for connection management by other (i.e. administrative) means. Some recommendations for suitable discovery protocols (i.e. Zeroconf aka Bonjour, SAP or other existing mechanisms) have been given in the annex of AES67. Future extensions or supplemental standards may further address this topic.

18 https://en.wikipedia.org/wiki/Session_Initiation_Protocol
19 http://tech.ebu.ch/docs/tech/tech3326.pdf
20 https://en.wikipedia.org/wiki/Internet_Group_Management_Protocol

3.4 Application

The standard is expected to be useful for interoperability scenarios within broadcast, music production and post-production facilities as well as for commercial audio applications including fixed and touring live sound reinforcement.

AES67 is not intended to be a solution on its own, but rather providing means for exchanging audio streams between areas with different networking solutions or technologies already in place. It can be expected that the various IP-based solutions enhance their capabilities to adopt an AES67-compliant stream mode to facilitate inter-system interoperability. It is expected that the various solutions will stay in the market as their individual features exceed the commonalities defined by AES67. In this respect, AES67 may be seen as the "O-negative of audio networking" - in analogy to the blood group "O" which is the universal donor[21].

The following picture sketches a typical application scenario:

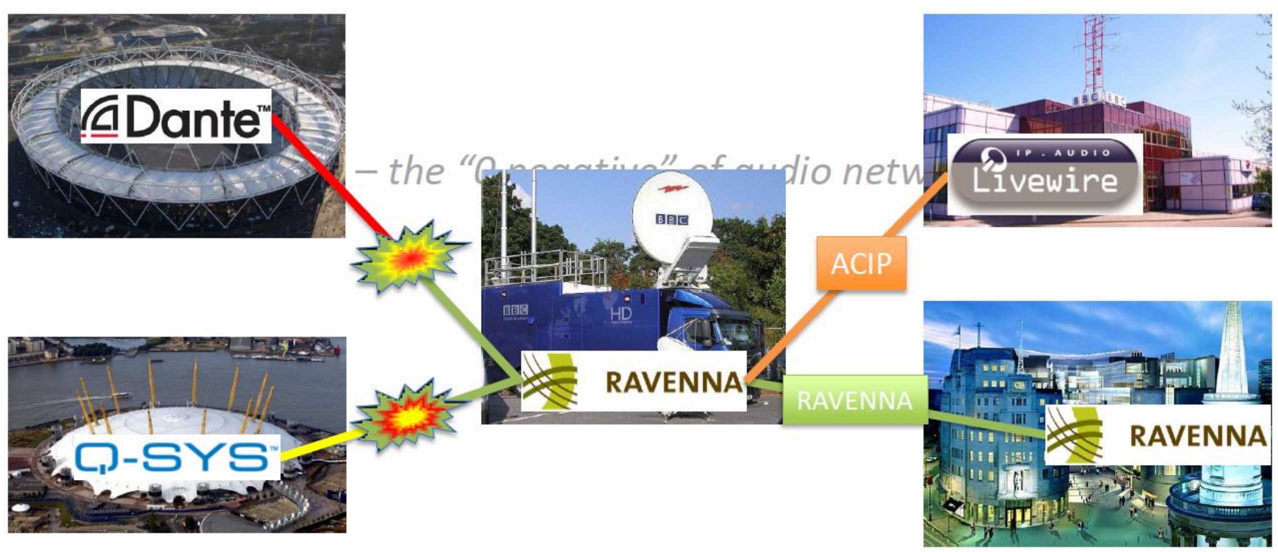

Figure 2: Typical interoperability impairments

This situation may be resolved without the need to replace the already installed systems – provided that the participating systems support AES67:

[21] https://en.wikipedia.org/wiki/Blood_type

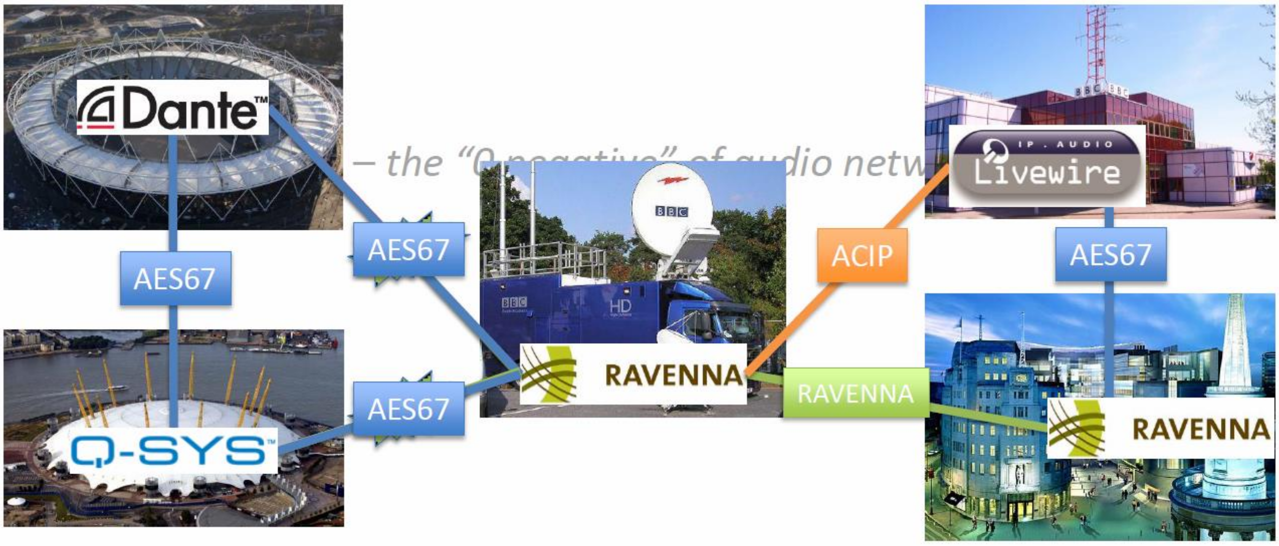

Figure 3: Interoperability established through AES67

4 RAVENNA & AES67

Since all relevant standard ingredients of AES67 are either identical or very similar to RAVENNA's operating principles, RAVENNA can naturally fully support interoperability as defined within AES67. However, since the RAVENNA technology framework offers performance and functionality superior to the AES67 interoperability guidelines, AES67 can be seen as one of many possible operational profiles for RAVENNA. In fact, the AES67 specification is very close to the RAVENNA Generic Profile, so that it can be expected that all RAVENNA-enabled devices supporting the RAVENNA Generic Profile will also support AES67. Other RAVENNA profiles offer faster performance and lower latency capabilities, higher channel counts, better applicability to routed network environments or means of transporting different types of media.

4.1 The RAVENNA AES67 Profile

While the fundamental mechanisms and protocols for synchronization, transport and payload formats of AES67 are identical to RAVENNA's operating principles, AES67 calls for SIP connection management for unicast stream operation. SIP is not specified for RAVENNA as RAVENNA uses RTSP/SDP for connection management. RTSP/SDP is a standardized, widely supported mechanism for internet stream connection management and is used and understood by most media servers and players (i.e. Windows Media Player, VLC media player et al). SIP is widely being used with VoIP telephony, but also within the ACIP protocol specification for codec-based audio contribution in broadcast applications. Since the RAVENNA framework allows for extensions and protocol variations beyond its core definition, SIP can easily be added as an ingredient for the RAVENNA AES67 operating profile. In other words, RAVENNA devices continue to use RTSP/SDP connection management, while SIP will be used when communicating to other devices under the AES67 operating profile.

While service advertisement and discovery are not defined within AES67, RAVENNA devices continue to use their advertisement and discovery mechanism based on Zeroconf, offering automatic discovery even when operating under the AES67 profile. However, since all RAVENNA devices also support manual configuration for stream setup, connections to other AES67-compatible devices not supporting Zeroconf-based discovery can of course also be established.

5 ASSESSMENT & OUTLOOK

RAVENNA is a technology framework based on well-known and widely supported mechanisms. All employed protocols are based on IP and conform to established industry standards. The operating principles are fully published and implementation is license-free. Consequently, RAVENNA is well-accepted and can be enhanced and extended as application requirements demand.

Other IP-based technologies were in existence prior to RAVENNA, but – although featuring some commonalities – were not able to interoperate with each other. The motivation for AES67 was to define interoperability guidelines to which existing solutions can be adapted with reasonable effort in order to facilitate synchronized inter-system stream exchange.

AES67 and RAVENNA share the same fundamental principles for synchronization and transport, while AES67 packet setup and payload formats are functional subsets of RAVENNA. Minor differences exist with stream connection management, which can easily be adopted by the RAVENNA technology framework. Consequently, AES67 will be supported by RAVENNA devices as an operational profile.

Other audio-over-IP solutions will most probably continue to exist, as most of them already have a significantly large installed base. Most of them also offer application-specific functionality beyond the scope of AES67. Following X192 Task Group participation and contribution, it can be expected that several other existing solutions will be modified or enhanced to support AES67 in the future. This will open up the landscape of product diversity and widen the field of application for AES67-capable devices. Limited interoperability can also be expected with devices compliant to the next revision of the EBU Tech 3326 protocol suite, which is currently work in progress in the EBU ACIP2 Work Group.

This all will help broadening the basis of acceptance for high-performance network-based media distribution systems like RAVENNA.

Document information:

Author: Andreas Hildebrand

Editor: Andreas Hildebrand

Version: 1.0

Last revision: 2014-01-15 by Andreas Hildebrand

File: RAVENNA & AES67_V1.0.docx