AUDIO FOR TELEVISION – HOW AES67 AND UNCOMPRESSED 2022/2110/TR03 VIDEO FIT TOGETHER

Leigh Whitcomb

Imagine Communications, Canada

ABSTRACT

Independent audio routing, or SDI audio breakaway, is a standard aspect of today's TV production workflow and is functionality that will need to be implemented in IP as the industry transitions away from SDI. Fortunately, the existing AES67 standard for audio over IP meets this objective and eliminates the need for the industry to reinvent the wheel. Not only is there already AES67 equipment deployed in the audio industry, but using this standard also enables significant new workflow opportunities.

This paper provides an overview of AES67 and explores how it can be used with SMPTE 2022-6 and VSF TR03 uncompressed video to form a complete solution. AES67 uses the IEEE-1588 PTP timing standard, and combining AES67, IEEE-1588, SMPTE ST 2059 and SMPTE ST 2022-6 using the new VSF Technical Recommendation 04 (TR04) provides a solution for maintaining A/V alignment throughout the production workflow.

INTRODUCTION

In today's TV production, audio can take two forms: embedded audio and independent audio. Production environments often mix the two – embedded audio on video sources, alongside separate audio for audio-only devices. As the industry transitions to IP infrastructure, each of these models still has applications, and the industry will need to support both.

With embedded audio, the audio is carried with the video. In SDI, the audio is literally carried in the ANC space of the video. From a routing, management, timing and audio/video alignment point of view, this is a simple model. The IP extension of this model is SMPTE ST 2022-6, wherein the audio is embedded in the video and the video with audio is transported via IP.

With independent audio or SDI audio breakaway, the audio is not routed with the video. This model allows the audio to be routed and processed independently from the video, for example, routing the audio to and from an audio production console. While this model increases flexibility, it also increases complexity for routing, management, timing and A/V alignment. The AES67 standard offers one method of implementing audio breakaway in the IP domain. AES67 transports audio channels over IP in separate audio streams, which can be separately routed within the IP routing fabric.

AES67

Historically, there have been many different standards to choose from when it comes to implementing audio over IP, with some of the best publicised being Ravenna, Dante, LiveWire, Q-LAN, WheatNet-IP and AVB. While many of these standards are similar, each is also a little different. To improve interoperability within the industry, the Audio Engineering Society (AES) developed AES67 "AES standard for audio applications of networks – High-performance streaming audio-over-IP interoperability" (1). This was first published in 2013 and updated in 2015.

At a high level, AES67 creates RTP IP packets with the PCM audio samples. There are no additional headers or overhead.

Key components of AES67 are:

- Synchronization
- Transport
- Encapsulation and Streaming
- Session description

AES67 Synchronization

Synchronization between all the AES67 transmitters and receivers is accomplished using the IEEE 1588-2008 Precision Time Protocol (PTP) (2). PTP distributes a precision time using IP to all the devices in the network. Using this time, transmitters and receivers generate locked media clocks, which are used for sampling the input audio and generating RTP timestamps.

The flexible PTP standard allows the use of profiles for different industries and applications. Profiles restrict parameters such as the rate at which time messages are sent. AES67 allows for the IEEE default profile and an AES67 profile.

Since all the AES67 devices on the network are locked together, audio sample rate conversion (SRC) is not required between the AES67 devices. SRC may be required for unlocked input signals

AES67 Transport

Transport describes how the encoded media data is transported across the network. AES67 is transported using Real-time Transport Protocol (RTP) over User Datagram Protocol (UDP) over IPv4. AES67 allows for both unicast and multicast; however, multicast is the normal mode for the professional broadcast market. For multicast audio essence and PTP messages, AES67 devices are required to support IGMPv2 and may support IGMPv3.

AES67 Encapsulation and Streaming

The AES packets are created by concatenating the PCM audio samples together. AES67 allows for:

- Sampling: 44.1KHz, 48KHz and 96KHz
- Bit width: 16-bit (L16) and 24-bit (L24)

AES67 packets have constant time spacing between packets. This is called the "packet time." AES67 defines a range of packet time intervals, and the recommended values are 125us, 250us, 333us, 1ms and 4ms. Other values are permitted.

AES67 supports a range of audio channels per stream. The maximum is limited by the network maximum transmission unit (MTU), sampling rate, bit width and packet time interval. Some example limits are:

- 80 channels for 125us packet interval, L24, 48KHz sampling, MTU = 1500
- 10 channels for 1ms packet interval, L24, 48KHz sampling, MTU = 1500

Because the range of AES67 parameters is large, the combinations of parameter sets are extremely large. Many of the combinations do not make sense for most applications and industries. To promote interoperability, the standard defines a set of parameters that all transmitters and receivers must support. Most applications and use cases are covered with this required set

- Sampling: 48KHz
- Bit width: 16-bit (L16) and 24-bit (L24)
- Packet time interval: 1ms
- Audio channels: 1-8

Audio channels	Packet Timing in microseconds							
	62.5	125	250	333.3	500	1000	2000	4000
1	9	18	36	48	72	144	288	576
2	18	36	72	96	144	288	576	1152
3	27	54	108	144	216	432	864	
4	36	72	144	192	288	576	1152	
5	45	90	180	240	360	720	1440	
6	54	108	216	288	432	864		
7	63	126	252	336	504	1008		
8	72	144	288	384	576	1152		
12	108	216	432	576	864			
16	144	288	576	768	1152			
20	180	360	720	960	1440			
24	216	432	864	1152				
32	288	576	1152					
40	360	720	1440					
48	432	864						
56	504	1008		AES67 Receivers Shall Support				
64	576	1152		AES67 Receivers Should Support				
72	648	1296		AES67 Receivers May Support				
80	720	1440						
88	792							
96	864							
128	1152							

Table 1 – AES67 Receiver Shall/Should/May audio channel support and packet size for 48KHz 24-bit audio

AES67 Session Description

For a receiver to decode a stream, it needs to know some information about the stream. This includes number of audio channels per stream, bit width of the audio, and the synchronization source. Because there are no AES67 headers or in-band information, this information is sent to the receiver using Session Description Protocol (SDP), as specified in RFC 4566 (3).

AES67 defines fields in the SDP; however, the standard does not mandate how the SDP information is distributed throughout the system. This allows for flexibility in the future, but also causes some interoperability issues.

SDP is a human and machine-readable format; however, it is very cryptic for humans. An example of part of an SDP is

v=0

- Protocol version. E.g. 0

c=IN IP4 239.1.1.2/32

- Connection Data. This is IP address and Time To Live (TTL). E.g. IP address is 239.1.1.2 and TTL is 32.

t=0 0

- Time the session is active. Start and End time. A zero end time is unbounded.

m=audio 5004 RTP/AVP 96

- Media Descriptions. E.g. RTP protocol on port 5004

i=Channels 1-8

- Session Information. Textual information about the session

a=rtpmap:97 L24/48000/6

- RTP dynamic payload type. E.g. 97. Audio sample bit width, sampling rate and audio channels. E.g. 24-bit, 48KHz and 6 audio channels.

a=ptime:0.250

- Packet time in ms. This is the time between packets. E.g. 0.250ms

a=ts-refclk:ptp=IEEE1588-2008:39-A7-94-FF-FE-07-A7-E0: 0

- Network clock used by the stream. Attributes are specified in RFC 7273 (4). E.g. The network clock is IEEE1588-2008, the PTP Grandmaster is 39-A7-94-FF-FE-07-A7-E0 and the PTP domain is 0.

a=mediaclk:direct=113224434

- Offset between the RTP timestamp and the PTP time. In this example, the RTP timestamp was 113224434 at the media clock epoch.

AES67 for Breakaway Audio in Television Production

Today's common baseband breakaway audio use case allows users to route audio in mono channels or stereo pairs of channels depending on the application. Audio channels can also be logically grouped and switched in larger groups, all while maintaining sample-level synchronization between channels. To implement this same functionality in AES67, AES67 packets would need to be limited to 1 or 2 audio channels, because the packet stream is the minimum routable unit on an IP network. While AES67 allows for more audio channels within each AES67 packet, increasing the number will limit the minimum routable unit. The only downsides of having 1 or 2 audio channels per packet are:

- Small increase in audio packet overhead
- Additional IP multicast address to setup

Note that AES67 by design only supports PCM audio. It cannot, therefore, be used for compressed audio or applications that use the C and U bit in an AES3 signal. There are industry groups looking at complementary approaches for this non-PCM use case.

Maturity of AES67

AES67 was first published three years ago (2013). This has given equipment vendors enough time to develop AES67 equipment. Since AES67 draws heavily on precursor technologies found in Ravenna and Dante, it gained a reasonable level of acceptance and deployment fairly quickly. To improve interoperability between AES67 equipment, AES has held two successful interoperability Plug Fest events dedicated to AES67. The first was in Munich in October 2014 and the second was in Washington in November 2015. The Video Services Forum (VSF) included AES67 at their January 2016 interop event in Houston, and more industry events are planned in 2016 and 2017. The number of equipment vendors participating at these events is constantly growing.

AES67 WITH SMPTE ST 2022-6

SMPTE ST 2022-6 “Transport of High Bit Rate Media Signals over IP Networks (HBRMT)” (5) defines the transport of uncompressed SD and HD over IP. The standard takes all the SDI samples and organizes them into 1376 octet groups and puts them into an IP packet.

If there is embedded audio in the SDI stream, SMPTE ST 2022-6 distributes this embedded audio over IP, along with the containing video. For applications that require breakaway audio, AES67 can be used for the audio, alongside SMPTE ST 2022-6 for the video. The AES67 audio can be de-embedded from the SDI or discretely taken. This method duplicates some bandwidth since the audio is sent twice (embedded and separately) in some cases.

Synchronization and Timing between ST2022-6 and AES67

ST 2022-6 does not require the video stream to be synchronized to the plant reference genlock system. AES67 requires the AES67 devices to be synchronized to the PTP master; therefore, the de-embedded audio generally must be made synchronous to the reference by processing through an audio sample rate converter (SRC) to move it from the video clock domain to the PTP clock domain. The same is true if/when the AES67 audio is embedded back into asynchronous SDI.

Audio/Video Alignment

Since the audio and video are separate streams with separate delay, methods must be developed to ensure the two can be aligned to prevent lip sync errors. The separate delays can occur from the different network/processing paths the audio and video signals take. The VSF Technical Recommendation TR-04 “Utilization of ST 2022-6 Media Flows within a VSF TR03 Environment” (6) defines a method for aligning audio and video.

In ST 2022-6, the RTP timestamp for a general ST 2022-6 stream is not locked to anything. VSF TR04 constrains the ST 2022-6 RTP timestamp by requiring it to be locked

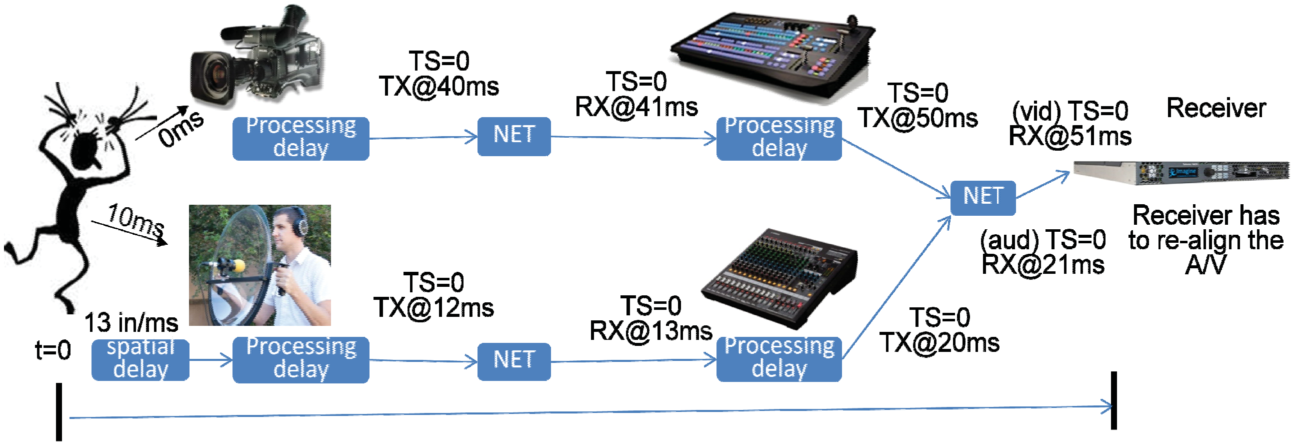

Figure 1 – Audio/Video Alignment in VSF TR04

to the PTP reference. It is then generated in a method similar to the AES67 RTP timestamp. Using the AES67 RTP timestamp, a receiver can generate an audio stream that has a fixed delay from when the audio stream was created. Using the VSF TR04 constrained ST 2022-6 RTP timestamps, a receiver can generate a video stream that has a fixed delay from when the video stream was created. By using the same delay for audio and video and assuming the original audio and video were aligned, the receiver can generate audio and video that are aligned. See Figure 1.

Using these timestamps, additional functionally is possible. For example, the microphone in Figure 1 could adjust the audio timestamp to compensate for the delay in the audio signal from the object to the microphone. Each microphone in the stadium could have a different adjustment so all the audio and video channels are aligned.

AES67 AS PART OF VSF TR03 (AND SMPTE ST 2110)

VSF Technical Recommendation TR-03 "Transport of Uncompressed Elementary Stream Media over IP" (7) defines a method to transport video, audio and ANC data essence over IP. SMPTE is creating a standard (ST 2110) for transporting video, audio and ANC data essence over IP that is largely based on TR03. VSF TR03 uses RFC 4175 (8) for the transport of video and AES67 for the transport of PCM audio.

VSF TR03 (containing AES67 audio) differs from VSF TR04 (SMPTE 2022-6 plus AES67 audio) in the following ways:

- TR03 does not transmit the audio twice
- TR03 does not have an embedded audio option – a receiver must subscribe to the separate audio streams in order to receive audio.

VSF TR03 Synchronization and Timing

VSF TR03 streams are required to be locked to a PTP master. It is possible to have multiple independent PTP masters within the network, e.g., asynchronous feeds coming into a TV plant. By examining the PTP master field in the SDP, a receiver can determine which audio and video streams are locked to the same master. For streams locked to the same master, no audio SRC is required. This is the typical use case. For streams locked to different masters, audio SRC and video frame syncs are required.

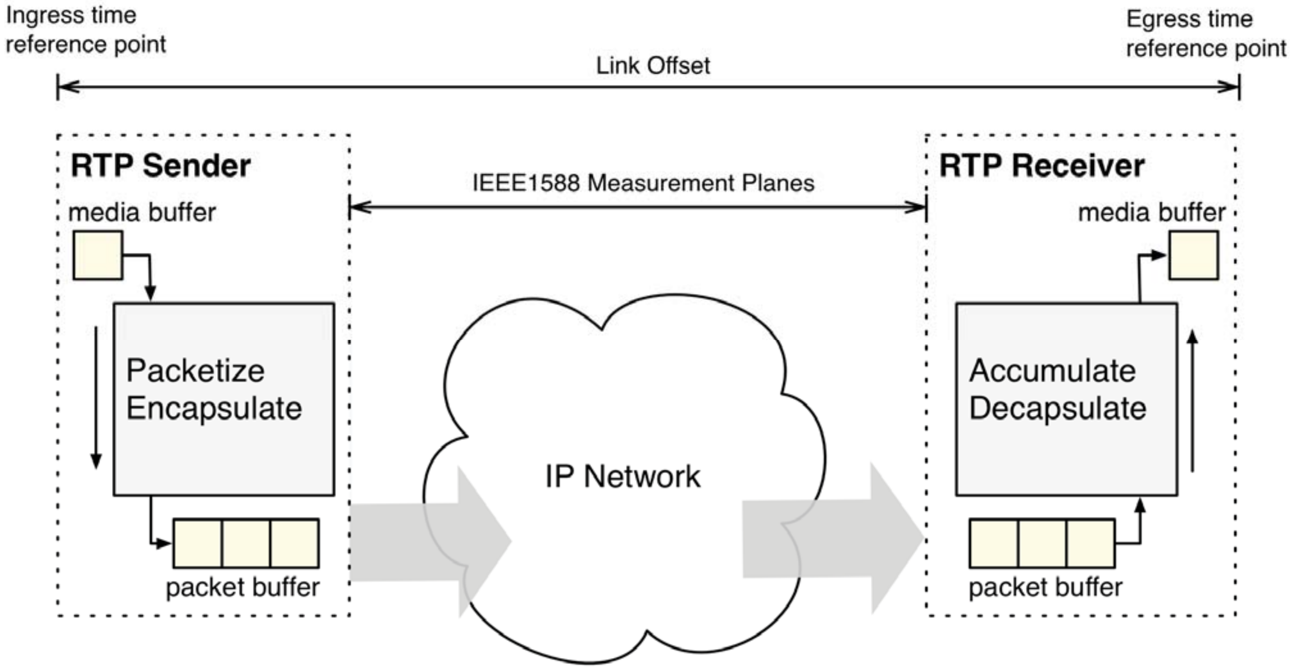

Figure 2 – Audio/Video Alignment in VSF TR03

VSF TR03 Audio/Video Alignment

VSF TR03 requires the video RTP timestamp to be locked to the PTP reference. Therefore, the same techniques used in TR04 can be applied to the flows within VSF TR03. See Figure 2.

AES67 WITH SMPTE ST 2059 GENLOCK

To migrate from analog black burst + DARS + LTC in today's reference system to IP networked timing, SMPTE has created the SMPTE ST 2059 family of standards. These standards build on the IEEE 1588 PTP standard. ST 2059-1 "Generation and Alignment of Interface Signals to the SMPTE Epoch" (9) defines generating locked signals from time. SMPTE ST 2059-2 "SMPTE Profile for Use of IEEE-1588 Precision Time Protocol in Professional Broadcast Applications" (10) defines an IEEE profile.

The ST 2059 PTP profile and AES67 PTP profile were developed at the same time. While liaison activities tried to keep the profiles compatible, there are some differences between the profiles. Some differences are due to parallel development and differing requirements. At the time of writing, there is ongoing harmonization work in both SMPTE and AES to ensure these profiles are compatible. To better facilitate this work, a joint AES-SMPTE taskforce is being created. AES, with input from SMPTE, has published a document that lists PTP parameters for AES67 and SMPTE ST 2059-2 interoperability. In addition, in June 2016, a weeklong interop workshop is being held to work through this issue. By the end of 2016 this issue should be resolved.

EVOLUTION OF THE STANDARDS FOR IP CARRIAGE OF AUDIO

VSF TR03 and VSF TR04 are being transformed from Technical Recommendations into standards through SMPTE ST 2110. SMPTE ST 2110 draws heavily on TR03 and TR04 however it is improving the contents of TR03 and TR04 and drawing on other sources. TR03 and TR04 will be superseded when SMPTE ST 2110 is published.

AES67 is a flexible standard. SMPTE ST 2110 uses AES67 however it puts some constraints on AES67 for the Professional Media industry. AES67 has some limitations and areas that are not fully standardized. For example since AES67 is limited to PCM audio, SMPTE 2110 will be standardizing how to transport non-PCM audio.

CONCLUSIONS

Independent audio, or SDI audio breakaway, is a standard aspect of today's TV production workflow and is functionality that will need to be implemented in IP as the industry transitions away from SDI. The existing AES67 standard for audio over IP meets this requirement for PCM audio.

AES67 can be used with either ST 2022-6 or VSF TR03. Combining these with ST 2059 and TR04 forms a complete solution for maintaining Audio/Video alignment throughout the production workflow.

When published, the SMPTE ST 2110 family of standards will supersede VSF TR03 and TR04. AES67 is key component of ST 2110.

REFERENCES

1. AES67-2015 AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability, 2015

2. IEEE-1588-2008 - IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, 2008

3. IETF RFC 4566 SDP: Session Description Protocol, 2006

4. IETF RFC 7273 RTP Clock Source Signalling, 2014

5. SMPTE ST 2022-6:2012 Transport of High Bit Rate Media Signals over IP Networks (HBRMT), 2012

6. VSF TR-04 (2015) Utilization of ST 2022-6 Media Flows within a VSF TR03 Environment, 2015

7. VSF TR-03 (2015) Transport of Uncompressed Elementary Stream Media over IP, 2015

8. IETF RFC 4175 RTP Payload Format for Uncompressed Video, 2005

9. SMPTE ST 2059-1:2015 Generation and Alignment of Interface Signals to the SMPTE Epoch, 2015

10. SMPTE ST 2059-2:2015 SMPTE Profile for Use of IEEE-1588 Precision Time Protocol in Professional Broadcast Applications, 2015

ACKNOWLEDGEMENTS

The author would like to thank John Mailhot, Steve Sulte and Karen Plumley for their contributions to this work.

Figure 2 is reproduced with permission of the VSF.