

IP 방송제작 기술 23

SMPTE ST 2110-30/40

글. 조인준 KBS 미디어기술연구소 차장

C군의 네버엔딩 스토리

이번 연재에서 다룰 내용은 SMPTE ST 2110 계열 내의 오디오 전송 표준인 SMPTE ST 2110-30 'Professional Media Over Managed IP Networks : PCM Digital Audio'와 부가 데이터 전송 표준인 SMPTE ST 2110-40 'Professional Media Over Managed IP Networks : SMPTE ST 291-1 Ancillary Data'이며 이로써 'IP 방송제작 기술' 시리즈의 막을 내릴 예정입니다. 그간 방송제작 기술의 전환점이 될 IP 전송 및 동기화 표준에 관한 연재를 지속하며 부족한 지식으로 인해 독자 여러분께 필요한 내용을 충실히 전달 드리지 못한 것 같아 더욱 열심히 공부해야겠다는 생각이 들었습니다. 언젠가 더 넓어진 지식으로 더 깊은 내용을 담아 다시 한번 'IP 방송제작 기술' 심화 시리즈를 연재하는 꿈을 꾼습니다. 그럼 'IP 방송제작 기술' 시리즈의 마지막 연재를 시작하겠습니다.

SMPTE ST 2110-30

IP 오디오 방식은 RAVENNA, Livewire, Q-LAN, Dante 등의 기술들이 각각의 영역을 구축한 채 호환성을 배제한 형태로 존재하고 있었습니다. 하지만 다양한 제조사의 IP 오디오 기기 간 호환의 필요성을 느낀 AES(Audio Engineering Society)에 의해 2013년 IP 오디오 전송 표준인 AES67이 제안되었습니다. AES67 표준의 성립이 가능했던 것에는 최소 공통 사항을 정의했다는 것과 각 제조사의 IP 오디오 기술이 대동소이한 형태를 하고 있어서 적은 수정만으로도 AES67을 지원하는 데 문제가 없었기 때문이라고 합니다. 이렇게 탄생한 AES67의 특징을 요약하면 [표 1]과 같습니다.

인코딩	PCM 16비트, 24비트
전송	RTP
샘플링 레이트	44.1kHz, 48kHz, 96kHz
미디어 클럭	48kHz
동기화	PTP

표 1. AES67 특징 요약

그런데 SMPTE ST 2110-30에 관한 설명에 왜 AES67이 등장하는 걸까요? 이유는 간단합니다. SMPTE ST 2110-30이 AES67을 기반으로 만들어졌기 때문입니다. 더 자세히 이야기하면 SMPTE ST 2110-30은 AES67 표준의 서브셋에 해당하는 내용을 담고 있습니다.

SMPTE ST 2110-30이 AES67의 서브셋으로서 가지는 한정 사항을 대략 정리하면 [표 2]와 같습니다. 샘플링 레이트는 48kHz, 96kHz만을 지원하고 패킷 타이밍은 125 μ s ~ 1ms 사이의 값을 가지며 48kHz 샘플링 레이트 기준으로 1~8, 또는 1~64개의 채널을 지원하는 것이 골자입니다.

기반 표준	AES67
샘플링 레이트	48kHz, 96kHz
인코딩	PCM 16비트, 24비트
패킷 타임	125 μ s ~ 1ms
채널 수 (패킷 타이밍)	1~8채널(1ms), 1~64채널(125 μ s) 48kHz 샘플링 레이트 기준

표 2. SMPTE ST 2110-30 특징 요약

우선 [표 2]에 나오는 패킷 타임이라는 다소 생소한 개념의 설명이 필요할 것 같습니다. 패킷 타임은 IP 패킷으로 데이터를 전송할 때 패킷 간 시간 차를 이야기합니다. 좀 더 쉽게 설명하면 48kHz의 주파수로 오디오를 샘플링할 경우 1초에 48,000개의 샘플이 나오고, 이 48,000개의 샘플을 1초에 걸쳐 온전히 전달해야 수신 측에서 제대로 오디오를 재생할 수 있습니다. 이때 패킷 타임이 1ms라는 것은 앞의 48,000개의 샘플들을 나누어 패킷에 실어 보낼 때 각 패킷을 1ms 간격으로 보내겠다는 의미이고, 1ms의 간격으로 패킷을 보내면 1초에 1,000개의 패킷을 보낼 수 있으므로 각 패킷은 48개의 오디오 샘플을 실어서 보내면 됩니다.

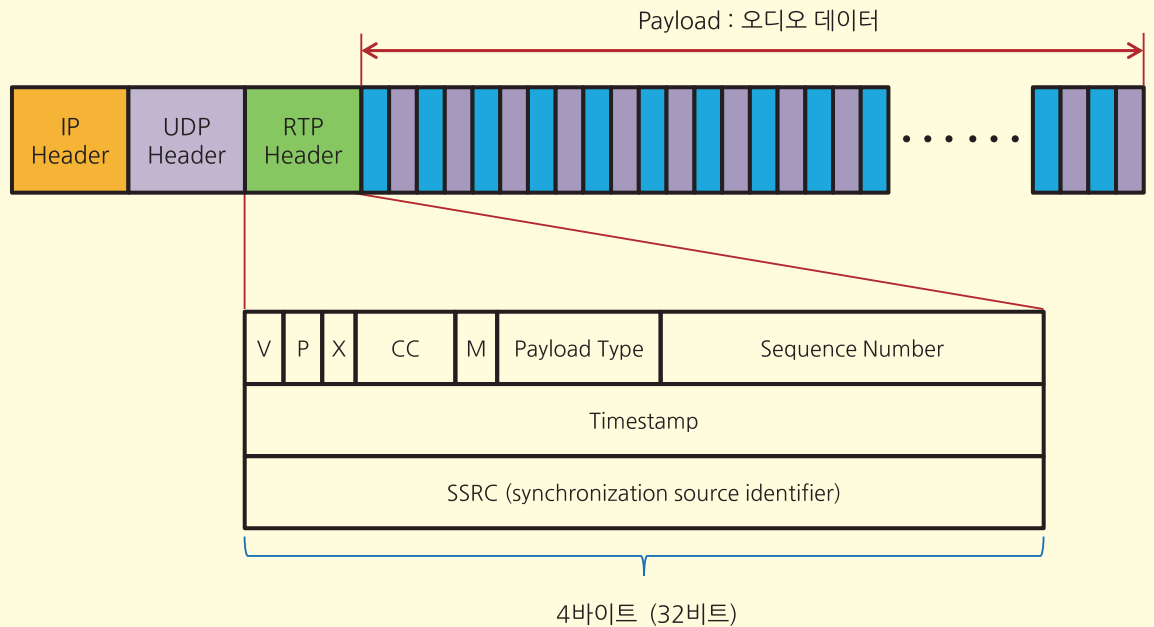


그림 1. SMPTE ST 2110-30 IP 패킷 및 RTP 헤더 구조

[그림 1]에 SMPTE ST 2110-30 IP 패킷 및 RTP 헤더 구조를 나타내었습니다. SMPTE ST 2110-20 비디오 전송 표준에서는 RTP 헤더 뒤에 RTP Payload Header가 붙고 그다음에 실제 비디오 데이터가 Payload로 붙었지만, SMPTE ST 2110-30에서는 RTP Header 뒤에 곧바로 PCM으로 인코딩된 오디오 샘플 데이터가 Payload로 붙습니다. 그런데 앞서 말씀드린 바와 같이 샘플링 주파수 48kHz의 오디오 데이터를 1ms의 패킷 타임으로 보내면 패킷당 48개의 오디오 샘플을 실어 보내게 되고, 48개의 오디오 샘플이 3바이트(24비트)로 PCM 인코딩되었다면 $48 \times 3\text{바이트} = 144\text{바이트}$ 가 됩니다. 이는 네트워크상에서 단편화를 피할 수 있는 UDP 패킷의 최대 크기인 1,460바이트에서 UDP Header 8바이트와 RTP Header 12바이트를 제외한 실제 전송 가능한 데이터 1,440바이트에 비하면 1/10에 지나지 않습니다. 그래서 오디오의 경우 한 패킷에 여러 채널을 동시에 실어 보내는 방식을 택하고 있으며, 샘플링 주파수 48kHz의 오디오 데이터를 패킷 타임 1ms로 전송할 경우 $1,440/144 = 10$ 이므로 이론적으로 최대 10개 채널의 오디오를 하나의 패킷으로 묶어 전송할 수도 있습니다. 이와 관련해 SMPTE ST 2110-30 표준에서는 오디오 데이터 수신장치의 레벨을 처리 가능한 채널 수에 따라 A, B, C로 나누어 [표 3]과 같이 정의하였습니다.

샘플링 주파수/패킷 타임	A	AX	B	BX	C	CX
48kHz, 1ms	1~8	1~8	1~8	1~8	1~8	1~8
48kHz, 125μs			1~8	1~8	1~64	1~64
96kHz, 1ms		1~4		1~4		1~4
96kHz, 125μs				1~8		1~32

표 3. 오디오 수신장치 레벨과 처리 가능 채널 수

[표 3]에서 A, B, C 레벨의 장치는 샘플링 레이트 48kHz 오디오만 처리 가능한 장치이며 뒤에 X가 붙은 AX, BX, CX 레벨의 장치는 샘플링 레이트 96kHz 오디오까지 처리 가능한 장치들입니다. 또한 레벨 A는 패킷 타임 1ms만 처리 가능하며, 레벨 B, C는 패킷 타임 1ms와 125μs 모두 처리 가능합니다. 그리고 마지막으로 레벨 B와 C의 구분 기준은 패킷 타임에 따른 처리 가능한 채널 수입니다. [표 3]에서 48kHz, 1ms 오디오 데이터의 경우 모든 장치가 8개의 채널을 지원하는 것을 볼 수 있습니다. 그런데 125μs에서 레벨 C 장치는 64개의 오디오 채널을 지원합니다. 패킷 타임이 125μs라는 것은 초당 8,000개의 오디오 데이터 패킷이 만들어진다는 것이고, 48kHz의 샘플링 레이트를 가정하면 패킷 타임마다 채널 당 48,000/8,000 = 6개의 오디오 샘플을 가지게 됩니다. 이를 3바이트(24비트)로 PCM 인코딩하면 6×3바이트 = 18바이트가 되어 125μs마다 1,440바이트의 UDP 전송용량을 최대로 사용하여 오디오 데이터를 송수신할 수 있는 장치에서는 이론적으로 최대 1,440/18 = 80개의 오디오 채널 데이터를 보내거나 받을 수 있습니다. 그리고, 받은 오디오 데이터를 모두 다 처리할 수 있다면 80개의 오디오 채널을 지원할 수 있습니다. 하지만 80개는 이론적으로 가능한 수치이므로 표준에서는 현실적인 부분을 감안하여 64개의 채널을 지원하는 것으로 레벨을 정한 것 같습니다. [표 3]의 다른 채널 숫자들도 앞의 내용과 같은 배경으로 정해진 숫자들이므로 관심이 있으신 독자분들은 하나씩 헤아려 보시기 바랍니다.

그렇다면 전송되는 오디오 패킷 안에 어떤 오디오 채널들이 들어있는지는 어떻게 알 수 있을까요? 이는 NMOS(Networked Media Open Specifications)라고 부르는 IP 방송장비 관리 및 제어를 위한 프로토콜에서 사용되는 SDP(Session Description Protocol) 내의 정보를 이용하여 알 수 있습니다(NMOS에 대해서는 추후 따로 연재를 기획해 보겠습니다). IP 방송제작 네트워크 내에 SMPTE 2110-30 표준을 지원하는 오디오 장치가 있을 경우 해당 장치가 NMOS 프로토콜을 통해 제공하는 SDP에는 다음과 같이 ‘channel-order’ 태그를 통해 오디오 채널 정보가 담겨 있습니다.

channel-order=SMPTE2110.(51, ST)

채널 그룹 심볼	그룹 내 채널 수	그룹 형식
M	1	모노
DM	2	듀얼 모노
ST	2	표준 스테레오
LtRt	2	매트릭스 스테레오
51	6	5.1 채널
71	8	7.1 채널
222	24	22.2 채널
SGRP	4	SDI 오디오 그룹
U01~U64	U 뒤의 숫자와 같음	기타

표 4. 오디오 수신장치 레벨과 처리 가능 채널 수

예의 ‘channel-order’ 뒤에 붙은 SMPTE2110.(51, ST)에서 SMPTE2110은 SMPTE 2110-30 표준을 사용하는 오디오 패킷이라는 정보이고 (51, ST)는 오디오 패킷이 포함하는 채널 정보이며, 이 채널 정보에 사용되는 심볼의 뜻은 [표 4]와 같습니다. [표 4]를 이용해 (51, ST)의 의미를 풀어보면 오디오 패킷 안에 6개의 채널로 된 5.1 채널 오디오 데이터와 2채널로 된 표준 스테레오 데이터를 합해 총 8채널의 오디오 데이터가 들어있다는 뜻이 됩니다.

SMPTE ST 2110-40

SMPTE ST 2110-40 표준의 제목은 'Professional Media Over Managed IP Networks : SMPTE ST 291-1 Ancillary Data'이며 말 그대로 SDI 신호의 부가데이터를 IP 스트림으로 전송하기 위한 표준입니다. 부가데이터는 비디오와 오디오에 비해 중요도 및 관심이 높지 않으므로 지면 관계상 아주 간략한 설명으로 마무리 하겠습니다(사실대로 고백하면 C군이 오디오도 잘 모르지만 부가데이터는 더 모릅니다). [그림 2]에 SDI 신호의 부가데이터 포맷을 나타내었습니다. SDI 신호에서 부가데이터는 Horizontal Blanking이나 Vertical Blanking 구간에 삽입되며 16진수 코드 0003FF3FF로 시작하여 부가데이터에 관한 정보를 담은 Data Identifier와 Secondary Data Identifier, 부가데이터의 길이를 나타내는 Data Count, 그리고 실제 부가데이터와 마지막의 체크섬으로 구성됩니다. SMPTE 2110-40에서는 [그림 3]과 같이 SDI 신호 내부의 0003FF3FF 플래그 이후의 데이터와 동일한 데이터에 Header와 패딩(패킷의 길이를 32비트의 배수로 만들기 위함)을 붙여 '부가데이터 패킷'으로 만든 후 이를 여러 개 묶어서 RTP Payload Header와 RTP Header를 붙여 전송합니다. [그림 3]의 RTP Header는 SMPTE ST 2110-20이나 2110-30 표준의 RTP Header와 같은 구조이며, RTP Payload Header에는 뒤에 몇 개의 부가데이터 패킷이 붙어 있는지, 그 데이터 길이는 얼마인지 등에 관한 정보가 들어있습니다.

SDI 신호 내 부가 데이터

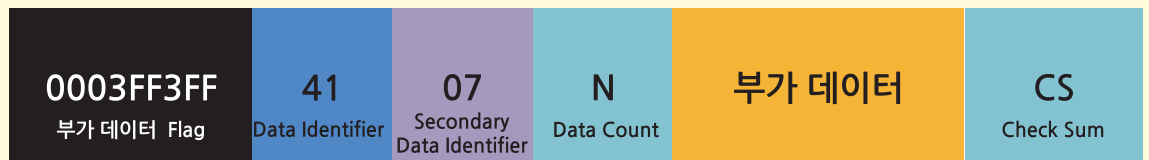


그림 2. SDI 신호 내 부가 데이터 구조

SMPTE ST 2110-40

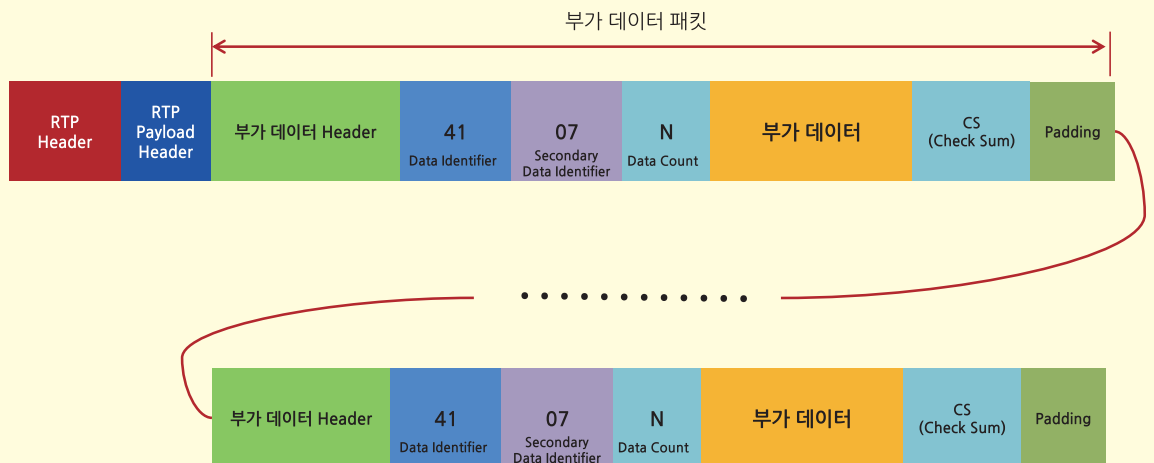


그림 3. SMPTE ST 2110-40 패킷 구조

이상으로 부족하나마 SMPTE ST 2110 표준에 관한 긴 연재를 마치겠습니다. C군의 부족한 지식으로 인해 독자 여러분께 쉽고 세밀하게 전달 드리지 못한 점 양해 부탁드립니다. 다음 연재부터는 좀 더 유익하면서 쉽고 재미있는 주제로 찾아뵙겠습니다. 📺