|  |  |
| --- | --- |
| 3GPP TR 26.805 V1.2.1 (2022-05) | |
| Technical Report | |
| 3rd Generation Partnership Project;  Technical Specification Group Services and System Aspects;  Study on Media Production over 5G NPN Systems  (Release 17) | |
|  | |
|  |  |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners’ Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles – Sophia Antipolis  Valbonne – FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  <http://www.3gpp.org> |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2021, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 5

1 Scope 7

2 References 7

3 Definitions of terms, symbols and abbreviations 11

3.1 Terms 11

3.2 Symbols 11

3.3 Abbreviations 11

4 Review of existing workflows and media protocols 12

4.1 General 12

4.2 Transport Protocols 13

4.2.1 General 13

4.2.2 SMPTE ST 2110 14

4.2.2.1 Introduction 14

4.2.2.2 ST 2110 for audio (ST 2110-30 and ST 2110-31) 14

4.2.2.3 ST 2110 for video (ST 2110-20 and ST-2110-22) 15

4.2.2.4 ST 2110 for metadata (ST 2110-40 and ST-2110-41) 15

4.2.3 Secure Reliable Transport (SRT) 15

4.2.4 Reliable Internet Stream Transport (RIST) 16

4.2.5 Network Device Interface NDI 17

4.2.6 IP Media eXperience (IPMX) 17

4.2.7 Comparison Table 18

4.2.8 Other Protocols 18

4.2.9 Audio Networking Solutions 18

4.3 Codec choice 19

4.5 Review of existing orchestration and control solutions 21

4.5.1 General 21

4.5.2 AMWA Network Media Open Specification (NMOS) 21

4.5.3 Camera control and configuration protocols 23

4.5.3.1 General 23

4.5.3.2 Camera control protocols 23

4.5.4 EMBER+ 24

4.5.5 Other Protocols 24

4.6 Evolving Internet media transport protocols 24

4.6.1 General 24

4.6.2 Tunnelling RTP media sessions over QUIC 25

5 Relevant media production use cases 26

5.1 General 26

5.2 Scenario 1: Wireless cameras within a production workflow 26

5.3 Scenario 2: Outside broadcast contribution 28

5.4 Considerations on remote and cloud-based production 28

5.5 Collaboration models and deployment architectures 29

5.5.1 General 29

5.5.2 Deployment #1: On-site wireless production with Standalone NPNs 30

5.5.3 Deployment #2: On-site wireless production with PNI-NPN or Outside Broadcast Contribution 30

5.5.4 Remote wireless production with Standalone NPNs 31

5.6 Low-Latency Production with 5G mmWAVE 32

6 Potential issues and Candidate Solutions 33

6.1 General 33

6.2 Key Issue #1: Utilizing Available Capacity in Multi-Camera Scenarios 33

6.3 Key Issue #2: Media Protocols on 5G: Traffic segregation and prioritization 35

6.3.1 General 35

6.3.2 Application flow prioritisation 35

6.3.3 Applying Quality of Service to application flows 36

6.3.4 Solutions leveraging 3GPP QoS 37

6.3.4.1 General 37

6.3.4.2 Solution Example A: Coarse-grained separation with separated media 37

6.3.4.3 Solution Example B: Fine-grained separation with separated media 38

6.3.5 Solutions leveraging Network Slices 38

6.3.5.1 General 38

6.3.5.2 Solution Example C: Separation using Multiple Network Slices 39

6.3.5.3 Solution Example D: Separation using Network Slices and QoS 39

6.3.6 Summary 39

6.4 Key Issue #3: Device onboarding, Remote device configuration and remote device control 40

6.4.1 General 40

6.4.1.1 Introduction 40

6.4.1.2 Media production event timeline 40

6.4.1.3 Media production event activities 41

6.4.1.3.1 Introduction 41

6.4.1.3.2 Onboarding and configuration activities before a media production event 41

6.4.1.3.3 Activities during an ongoing media production event: 41

6.4.1.3.4 Activities after a media production event 41

6.4.2 Solutions 42

6.4.3 Discussion 42

6.5 Key Issue #4: Different bit rates for Standby vs Program Cameras 42

6.5.1 General 42

6.5.2 Solutions leveraging QoS 43

6.5.2.1 General 43

6.5.2.2 Solution A: QoS rule modification 45

6.5.2.3 Solution B: Multiple pre-installed QoS rules 46

6.5.3 Discussion 47

6.5.3.1 Discussion of Solution A 47

6.5.3.2 Discussion of Solution B 47

6.6 Key Issue #5: Dynamic bit rate adaptation 47

6.6.1 General 47

6.6.2 Dynamic bit rate adaptation solutions leveraging RTCP-based congestion control signalling 48

6.7 Key Issue #6: Interfacing Audio Channels 48

6.7.1 Circuit-switched audio transmission 48

6.7.2 Packet-based audio channels 50

6.7.2.1 Introduction 50

6.7.2.2 Audio-over-Ethernet 50

6.7.2.3 Audio-over-IP 50

6.7.3 Solutions 51

6.7.4 Discussion 51

6.8 Key Issue #7: Usage of NPN (SNPN or PNI-NPN) 51

6.9 Key Issue #8: Usage of mmWAVE for low-latency television production 52

7 Summary and Conclusions 53

Annex <X> (informative): Change history 54

# Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

Y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**Should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**Can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**Will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# 1 Scope

The present document identifies standardization needs and potential standards gaps when using 5G Systems for media production. More specifically the following aspects are addressed in this document:

- To identify the relevant media production use cases (professional, semi-professional, production, contribution), based on existing use-cases from TR 22.827 as well as requirements from TS 22.263, that may benefit from 5G System functionalities. This includes collaboration use cases between media producers and 5G System operators.

- To develop one or several reference media production architectures and to map the variety of different media and control flows (such as uplink video, return video, tally, etc) involved in media production onto 5G System delivery components.

- To identify relevant QoS requirements for media production workflows, including required bit rates, loss rates, formats, latencies and jitter, and to identify their impact on the relevant KPIs for media production workflows (reliability, mean-time-between failure, service-level agreements, etc.).

- To identify relevant 5G System features like NPNs, Network Slicing, QoS classes, network event reporting and assistance, etc. that are useful for media production, and to clarify their usage for media production.

- To identify the suitability of existing media production content delivery protocols, codecs and service layers for 5G System usage, evaluate benefits and gaps, and recommend profiles or extensions in collaboration with organizations that develop and deploy existing protocols and codecs.

- To study media device and network orchestration solutions (such as AMWA NMOS), and their integration/interactions with the 5G exposure framework.

- To collaborate with relevant other 3GPP groups and external organizations (VSF, 5G-MAG, EBU, etc.) on media-related aspects of Media Production use cases.

- To identify potential normative work on media level for media production use cases in 5G Systems.

The document primarily focuses on the usage of 5G Systems including NPNs (both Standalone NPN and Public Network Integrated NPN).

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 22.261: "Service requirements for the 5G system".

[3] 3GPP TS 22.263: "Service requirements for Video, Imaging and Audio for Professional Applications (VIAPA)".

[4] 3GPP TR 22.827: "Study on Audio-Visual Service Production".

[5] M.P. Sharabayko, M.A. Sharabayko, J. Dube, JS. Kim, JW. Kim: "The SRT Protocol", draft-sharabayko-mops-srt-01

[6] Video Sercices Forum (VSF): "Reliable Internet Stream Transport (RIST) Activity Group",

[7] VSF TR 06-1:2020 "Reliable Internet Stream Transport (RIST) Protocol Specification – Simple Profile", https://vsf.tv/download/technical\_recommendations/VSF\_TR-06-1\_2020\_06\_25.pdf

[8] VSF TR 06-2, "Reliable Internet Stream Transport (RIST) Protocol Specification – Main Profile", [https://www.vsf.tv/download/technical\_recommendations/VSF\_TR-06-2\_2020\_03\_24.pdf](https://protect2.fireeye.com/v1/url?k=cc406e56-93db577d-cc402ecd-866038973a15-a3187c63f11b10f6&q=1&e=1f3c54ba-abd4-4509-b7b2-0816901e7741&u=https%3A%2F%2Fwww.vsf.tv%2Fdownload%2Ftechnical_recommendations%2FVSF_TR-06-2_2020_03_24.pdf)

[9] NewTek: "NDI Encoding/Decoding", <https://support.newtek.com/hc/en-us/articles/218109667-NDI-Encoding-Decoding>

[10] NewTek: "NDI Network Bandwidth, <https://support.newtek.com/hc/en-us/articles/217662708-NDI-Network-Bandwidth>

[11] David Aleksandersen: "What is NDI® (Network Device Interface)?", <https://newsandviews.dataton.com/what-is-ndi-network-device-interface>

[12] Kieran Kunhya and Ciro Noronha: "RIST and SRT: What’s the difference?", <https://www.tvbeurope.com/ip-migration/rist-and-srt-whats-the-difference>

[13] Tofik Sonono: "Interoperable Retransmission Protocols with Low Latency and Constrained Delay: A Performance Evaluation of RIST and SRT", Masters Thesis, KTH Stockholm, 2019.

[14] EBU: "Minimum User Requirements to Build and Manage an IP-Based Media Facility", 15 July 2020, <https://tech.ebu.ch/files/live/sites/tech/files/shared/tech/tech3371.pdf>.

[15] AMWA: "NMOS Overview", <https://www.amwa.tv/nmos-overview>.

[16] EBU: "The Technology Pyramid For Media Nodes", .

[17] EBU: "Technology Pyramid Media Node Maturity Checklist", September 2021, <https://tech.ebu.ch/publications/technology-pyramid-media-node-maturity-checklist?rec=1>.

[18] AMWA: "NMOS Technical Overview", <https://specs.amwa.tv/nmos/branches/main/docs/2.0._Technical_Overview.html>.

[19] AMWA: "Networked Media Systems – the Big Picture",  
<https://static.amwa.tv/networked-media-systems-big-picture-2021-03-05.pdf>.

[20] AMWA: "NMOS specification repository", <https://specs.amwa.tv/nmos>.

[21] SMPTE ST 2110: "Professional Media over Managed IP".

[22] IEEE 1588-2008: "Precision Time Protocol".

[23] SMPTE ST 2022-1:2007: "Forward Error Correction for Real-Time Video/Audio Transport Over IP Networks".

[24] SMPTE ST 2022-6:2012: "Transport of High Bit Rate Media Signals over IP Networks (HBRMT)",

[25] SMPTE ST 2022-7:2019: "Seamless Protection Switching of RTP Datagrams".

[26] SMPTE ST 2059-2:2015: "SMPTE Profile for Use of IEEE-1588 Precision Time Protocol in Professional Broadcast Applications".

[27] SMPTE ST 2110-10:2017: "Professional Media Over Managed IP Networks: System Timing and Definitions".

[28] SMPTE ST 2110-20:2017: "Professional Media Over Managed IP Networks: Uncompressed Active Video".

[29] SMPTE ST 2110-22:2019: "Professional Media Over Managed IP Networks: Constant Bit-Rate Compressed Video".

[30] SMPTE ST 2110-30:2017: "Professional Media Over Managed IP Networks: PCM Digital Audio".

[31] SMPTE ST 2110-31:2018: "Professional Media Over Managed IP Networks: AES3 Transparent Transport".

[32] IETF RFC 4585: "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)".

[33] IETF RFC 8086: "GRE-in-UDP Encapsulation".

[34] Ember+ control protocol.

[35] SMPTE ST 259:2008: "For Television — SDTV Digital Signal/Data — Serial Digital Interface".

[36] SMPTE ST 292-1:2012: "1.5 Gb/s Signal/Data Serial Interface".

[37] 3GPP TR 26.925: "Typical traffic characteristics of media services on 3GPP networks".

[38] Wikipedia: "MADI", last modified 19th April 2021, <https://en.wikipedia.org/wiki/MADI>

[39] Wikipedia: "Time-Sensitive Networking", last modified 23rd June 2021, <https://en.wikipedia.org/wiki/Time-Sensitive_Networking>

[40] AES67 / SMPTE ST 2110: "COMMONALITIES AND CONSTRAINTS", https://aimsalliance.org/wp-content/uploads/2019/04/AES67-SMPTE-ST-2110-Commonalities-and-Constraints-Updated-April-2019.pdf

[41] IETF RFC 5104: "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)".

[42] IETF RFC 4585: "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)".

[43] ISO/IEC 13818‑1: "Information technology — Generic coding of moving pictures and associated audio information — Part 1: Systems".

[44] IETF RFC 3550|STD 64: "RTP: A Transport Protocol for Real-Time Applications".

[45] IETF RFC 8086: "GRE-in-UDP Encapsulation".

[46] IETF RFC 2250: "RTP Payload Format for MPEG1/MPEG2 Video".

[47] IETF RFC 7798: "RTP Payload Format for High Efficiency Video Coding (HEVC)".

[48] OASIS: "MQTT Version 5.0", 7th March 2019, <https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html>

[49] IETF RFC 6416: "RTP Payload Format for MPEG-4 Audio/Visual Streams".

[50] IETF RFC 9134: "RTP Payload Format for ISO/IEC 21122 (JPEG XS)".

[51] SMPTE ST 2110-40:2018 "Professional Media Over Managed IP Networks: SMPTE ST 291-1 Ancillary Data".

[52] SMPTE ST 291-1:2011 "Ancillary Data Packet and Space Formatting".

[53] ITU-R BS.2076-2: "Audio Definition Model".

[54] SCTE-104 2019a: "Automation System to Compression System Communications Applications Program Interface (API)".

[55] SMPTE ST 2020-1:2014 "Format of Audio Metadata and Description of the Asynchronous Serial Bitstream Transport".

[56] ITU-R BS.2125-0: "A serial representation of the Audio Definition Model".

[57] AES67-2018: "AES standard for audio applications of networks – High-performance streaming audio-over-IP interoperability".

[58] GSA: "mmWave Bands 24.25 GHz", May 2021, <https://gsacom.com/paper/mmwave-bands-24-25-ghz-may-2021-executive-summary/>

[59] Li Nian, "The Next Journey for 5G: The Standardization and Application of mmWAVE", <https://www.gsma.com/greater-china/wp-content/uploads/2020/09/%E6%AF%AB%E7%B1%B3%E6%B3%A2%E6%A0%87%E5%87%86%E5%8C%96%E5%92%8C%E8%AF%95%E9%AA%8C%E8%BF%9B%E5%B1%95_%E4%B8%AD%E5%9B%BD%E7%A7%BB%E5%8A%A8_%E6%9D%8E%E7%94%B7-1.pdf>.

[60] GSMA Intelligence: "The economics of mmWave 5G", January 2021, <https://data.gsmaintelligence.com/research/research/research-2021/the-economics-of-mmwave-5g>

[61] 3GPP TR 38.900: "Study on channel model for frequency spectrum above 6 GHz".

[62] Yiqing Cao: "Live 8K production using 5G mmWave", <https://www.5g-mag.com/post/follow-up-workshop-media-production-over-5g-npn-deep-dive-into-protocols>.

[63] IETF RFC 9000: "QUIC: A UDP-Based Multiplexed and Secure Transport", May 2021.

[64] IETF Internet-Draft draft-ietf-quic-datagram-08: "An Unreliable Datagram Extension to QUIC", 14th January 2022, <https://www.ietf.org/archive/id/draft-ietf-quic-datagram-08.html>.

[65] IETF Internet-Draft draft-gruessing-moq-requirements-00: "QUIC Encapsulation for Media over RTP – Requirements and Use Cases", October 2021, https://www.ietf.org/id/draft-gruessing-moq-requirements-00.html.

[66] IETF Internet-Draft draft-hurst-quic-rtp-tunnelling-01: "QRT: QUIC RTP Tunnelling", 28th January 2021, <https://datatracker.ietf.org/doc/html/draft-hurst-quic-rtp-tunnelling-01>.

[67] IETF Internet-Draft draft-engelbart-rtp-over-quic-01: "RTP over QUIC", 25th October 2021, https://www.ietf.org/archive/id/draft-engelbart-rtp-over-quic-01.html.

[68] IETF Internet-Draft draft-ietf-quic-http-34: "Hypertext Transfer Protocol Version 3 (HTTP/3)", February 2021, https://www.ietf.org/archive/id/draft-ietf-quic-http-34.html.

[69] IETF RFC 2543: "SIP: Session Initiation Protocol", March 1999.

[70] IETF Internet-Draft draft-dawkins-avtcore-sdp-rtp-quic-00: "SDP Offer/Answer for RTP using QUIC as Transport", 28th January 2022, <https://www.ietf.org/archive/id/draft-dawkins-avtcore-sdp-rtp-quic-00.html>.

[71] IETF RFC 4588: "RTP Retransmission Payload Format", July 2006.

[72] IETF RFC 8888: "RTP Control Protocol (RTCP) Feedback for Congestion Control", January 2021.

[73] IETF RFC 8698: "Network-Assisted Dynamic Adaptation (NADA): A Unified Congestion Control Scheme for Real-Time Media", February 2020.

[74] IETF RFC 8298: "Self-Clocked Rate Adaptation for Multimedia", December 2017.

[75] IETF Internet-Draft draft-ietf-rmcat-gcc-02: "A Google Congestion Control Algorithm for Real-Time Communication", 8th July 2016, https://datatracker.ietf.org/doc/html/draft-ietf-rmcat-gcc-02.

[76] IETF RFC 8382: "Shared Bottleneck Detection for Coupled Congestion Control for RTP Media", June 2018.

[77] IETF Internet-Draft draft-ietf-quic-multipath-00: "Multipath Extension for QUIC", February 2022, <https://www.ietf.org/archive/id/draft-ietf-quic-multipath-00.html>.

[78] ITU-R BT.2069-7, "Tuning ranges and operational characteristics of terrestrial electronic news gathering (ENG), television outside broadcast (TVOB) and electronic field production (EFP) systems", October 2017, https://www.itu.int/dms\_pub/itu-r/opb/rep/R-REP-BT.2069-7-2017-PDF-E.pdf.

[79] BBC White Paper WHP 277, "BBC halfRF MIMO Radio-camera Programme Trial", February 2014, http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP277.pdf.

[80] AES-R16-2021: "AES Standards Report – PTP parameters for AES67 and SMPTE ST 2059-2 interoperability", October 2021, <https://www.aes.org/publications/standards/search.cfm?docID=105>

[81] ETSI TS 103 589: "Higher Order Ambisonics (HOA) Transport Format".

[82] Offical Audinate DANTE FAQ: <https://www.audinate.com/learning/faqs/clocking>

[83] ITU-R BS.1548-7 "User requirements for audio coding systems for digital broadcasting".

[84] 3GPP TS 23.501: "System architecture for the 5G System (5GS)".

[85] IETF RFC 7950: "The YANG 1.1 Data Modeling Language".

[86] IETF RFC 6241: "Network Configuration Protocol (NETCONF)".

# 3 Definitions of terms, symbols and abbreviations

## 3.1 Terms

For the purposes of the present document, the terms given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

**Non-Public Network**: See definition in TS 22.261 [2].

NOTE 1: Not all media production scenarios need a Non-Public Network.

**Tier 1, 2, 3**: Different categories of media production with differences in importance and usage characteristics.

**Production link**: A connection, usually bidirectional with strict QoS and latency requirements, between one or more devices used in a production environment to carry audio, video or other data.

**Contribution link**: A connection between a production location and a broadcast centre that is usually a single path for tier 3 production but may be a dual path for Tier 1 events.

NOTE 2: Link technologies that support contribution include fibre, satellite, microwave and bonded cellular.

NOTE 3: Not all production scenarios use both types of link. A recorded event may use production links with no contribution element and a single-camera tier 3 event may just use a contribution link.

## 3.2 Symbols

For the purposes of the present document, the following symbols apply:

Symbol format (EW)

<symbol> <Explanation>

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

AES Audio Engineering Society

AIMS Alliance for IP Media Solutions

AMWA Advanced Media Workflow Association

ARQ Automatic Repeat Query

CCU Camera Control Unit

DNS Domain Name System

FEC Forward Erasure Correction, Forward Error Correction

HDCP High-bandwidth Digital Content Protection

HDR High Dynamic Range

HFR Higher Frame Rates

HOA Higher Order Ambisonics

HTF HOA Transport Format

IPMX IP Media eXperience

mDNS Multicast DNS

MADI Multichannel Audio Digital Interface

MQTT Message Queuing Telemetry Transport

NDI Network Digital Interface

NMOS Networked Media Open Specifications

NPN Non-Public Network

PA Public Address

PLMN Public Land Mobile Network

PTP Precision Time Protocol

PTZ Pan, Tilt, Zoom

QRT QUIC RTP Tunnelling

RIST Reliable Internet Stream Transport

SDI Serial Digital Interface

SMPTE Society of Motion Picture and Television Engineers

SRT Secure Reliable Transport

VSF Video Service Forum

WAN Wide Area Network

# 4 Review of existing workflows and media protocols

## 4.1 General

There is a variety of different scenarios for media production operations supporting different workflows across multiple genres, editorial ambitions and budgets. Over time, solutions have evolved from analogue- to digital-based workflows and the production community is currently migrating to IP based architectures. IP-based production presents, among others, benefits in terms of:

- Handling existing and new video and audio technologies as media is transported as data IP packets.

- T he carriage of media, control and data on the same network.

- The harmonization of wide area and local media networks around same technologies, therefore increasing flexibility and scalability of equipment and infrastructure.

The largest single challenge is in the transport of high-quality video and audio content from multiple cameras and microphones to tools that combine these into an output such as a television programme or video stream that can then be used for onward distribution. This requires high bandwidth in the uplink path, often low latency in the network, time synchronization, as well as meeting challenging network Quality of Service requirements.

Alongside the uplink of video and audio there is often network traffic in the downlink direction which consists of a number of different functions such as control, reverse audio and video and other forms of data, all having different UEs at both the source and receiving ends.

Activities in media operations can be broadly broken down into three categories:

1. *Production:* All of the activity that happens locally on location. This activity often involves multiple sources of audio and video content used alongside different technologies to produce content.

2. *Contribution:* The act of moving content from a production location to a broadcast centre to be distributed. The content is often a single source of Audio/Visual (AV) content that is moved over large distances.

3. *Installed and Live Sound:* Operations and workflows related to the provision of live sound (usually an audio mix of the activity that happens during production) to performers on stage, through in-ear monitoring devices and/or to the on-site audience through Public Address (PA) systems. This provisioning involves an audio transmission “closed loop” scenario, thus requiring extremely low latency transmission of the audio content.

There are also different tiers of production activities that can be broadly broken down as follows:

- Tier One production:

- Usually heavily planned in advance with high budgets.

- Examples may include sports, cultural or historical events and studio production.

- Audio is usually separated and may have extra requirements such as live audio feedback to performers, Public Address (PA) distribution on site, or television/radio feeds.

- These events usually demand the highest-level requirements in terms of bandwidth and latency.

- Tier Two production:

- Usually planned in advance, but with lower budgets than Tier One productions.

- Examples include smaller scale sport and cultural events.

- Audio production is usually separated and may have extra requirements such as live audio feedback to performers or PA distribution on site.

- Audio for contribution may be taken from a local source such as a PA or venue system.

- Large potential for cloud-based and distributed production.

- Tier Three production:

- Usually less planned and with constrained budgets.

- Examples include live news and current affairs.

- Simple solutions and often mixed production and contribution workflows.

- Sometimes nomadic and growing in scale over time.

- Best efforts transmission, and often highly compressed.

- Audio is usually contributed locally to the camera.

In order to meet the requirements of different production scenarios a number of different solutions have been developed. For instance, certain protocols and codecs are better suited to different tiers of production. Alongside these, different solutions have emerged: some based on proprietary solutions that meet certain aspects of the workflows and some more open and interoperable. In addition to the media transport layer there are also requirements around control and orchestration which are generally based on specific technologies built on open infrastructures while some others are proprietary to support specific vendor implementations.

## 4.2 Transport Protocols

### 4.2.1 General

Transport protocols describe the way data is carried over networks. For media operations there are a number of potential options. Different Transport protocols support a variety of different wrappers (or payload formats), which allow carriage of different media codecs and other data.

Transport protocols typically support reliability (e.g. using ARQ or FEC), support for packet pacing and/or traffic shaping, features to allow network address translation (NAT) and security features such as firewalls in the network path. Some transport protocols also support some form or congestion control to handle different network load conditions.

Transport protocols include the carriage of a timestamp in their protocol header fields, which allow media from different sources to be related to a common production wall-clock time reference. Depending on the protocol and the usage, a sender may need to be time-synchronized with the system, so that the system can align streams from different media source devices.

Editor’s Note: The TR should clarify, how the different protocols can work with 5G QoS

### 4.2.2 SMPTE ST 2110

#### 4.2.2.1 Introduction

SMPTE ST 2110 [21] specifies an RTP-based media transport protocol intended for the carriage of uncompressed media streams in managed production networks. Its primary goal is to provide a viable replacement for the Serial Digital Interface (SDI) [35] [36] in professional media production environments using commodity networking infrastructure and interconnects. It is designed to be format-agnostic, handling various video formats such as 720/1080/4k raster lines, progressive/interlaced raster scan, High Dynamic Range (HDR) sampling, Higher Frame Rates (HFR), audio formats and ancillary formats. There are standards for both compressed and uncompressed audio and video workflows, even though the first round of work in SMPTE has focused on uncompressed workflows. ST 2110 is currently optimised for use in studios and production facilities.

SMPTE ST 2110 keeps apart audio, video and ancillary data in separate elementary streams. This is done to provide flexibility, allowing different elementary streams to be routed and worked on independently.

ST 2110 also takes into consideration that the underlying infrastructure is no longer synchronous (in contrast with the precursor Serial Digital Interface). The enabler for separating audio, video and data streams on an asynchronous infrastructure is timing, making sure that each elementary stream is time stamped and that timing information is carried in the RTP header as part of the stream. In the case of ST 2110 this is achieved using the Precision Time Protocol (PTP)  [22] [26].

In addition to timing, another challenge of moving to asynchronous infrastructure is burstiness. With a synchronous infrastructure the concept of burstiness does not exist, as traffic is delivered in one continuous flow. With IP, that is no longer the case. Being packet-based, each device along the traffic path contains buffers that are not synchronized. That means each device and buffer acts independently, resulting in traffic being delivered in bursts rather than as a continuous flow. For this reason, ST 2110 defines several sender and receiver profiles describing the different packet pacing patterns and the burst sizes accepted in different environments.

#### 4.2.2.2 ST 2110 for audio (ST 2110-30 and ST 2110-31)

In SMPTE ST 2110, audio transport is based on AES67 [57], specifying how to carry uncompressed 48 kHz, or 96 kHz Pulse Code Modulated (PCM) audio. Up to 64 channels can be bundled in one stream and both 16- and 24-bit depth is supported. In addition to this the ST 2110-31 [31] standard specifies how to bit-accurately transport PCM and non-PCM AES3 (AES/EBU) audio payloads over IP.

ST 2110 relies on ST 2110-30 [30] that is based on AES67 for the audio transport. However, ST 2110-30 and ST 2110‑10 [27] introduce additional constraints compared to AES67. Mainly, ST 2110 constraints refer to the area of timing and synchronization.

Regarding the use of PTP, while AES67 mandates the use of gPTP and a specific media profile, ST 2110-30 devices require the use of the SMPTE 2059-2 PTP [26]. Fortunately, AES67 PTP Media profile and SMTPE 2059-2 PTP profile share many commonalities so that it is possible to configure devices to interwork. These commonalities are described in the AES-R16-2016 report [80] that defines preferred PTP profile variables range that can be used in mixed ST 2110/AES67 environments. Further, most AES67 devices support the SMPTE 2059-2 PTP profile. Another more important constraint impacts the offset of the media clock and the network clock. ST 2110-30 requires that the offset of the media clock with respect to the network clock is 0.

ST 2110- 31 builds on RAVENNA’s AM824 (IEC 61883-6) payload definition, which retains AES67 definitions for synchronization and RTP usage while it extends the AES67 payload definition in one byte. All non-linear audio data formats that fit into this pattern can be transported over ST2110-31.

With elementary streams, a key challenge for audio transport over a Wide Area Network (WAN) is how to protect against loss. This is typically done using Forward Error Correction (FEC) and/or “1+1 protection”, but FEC on low-bandwidth services such as audio introduces too much delay. The solution is WAN architecture that can group together multiple streams into a high bandwidth bundle, on which FEC can be applied.

Editor’s Note: The TR should give some idea if packet losses for uncompressed audio can happen and which QoS requirements are available / known for ST 2110 for audio

#### 4.2.2.3 ST 2110 for video (ST 2110-20 and ST-2110-22)

Besides the RTP wrapper, another new thing about how uncompressed video is carried is that only the active part of the image, i.e. the pixels actually used, is sent. In contrast to SDI, ancillary data in the vertical blanking interval is not transported.

Defined to support resolutions up to 32×32k pixels, ST 2110 is future-proof with regards to supporting coming high-resolution formats and specifications. Support for colour modes and colour depths are flexible and include HDR.

Editor’s Note: The TR should give some idea if packet losses for uncompressed video can happen and which QoS requirements are available / known for ST 2110

#### 4.2.2.4 ST 2110 for metadata (ST 2110-40 and ST-2110-41)

SMPTE ST 2110-40 [51] provides a mechanism for carriage over IP of SMPTE ST 291 [52] metadata that has traditionally been carried in the VANC (Vertical ancillary data) space of an SDI multiplex. Because ST 2110-40 [51] supports all ST 291 [52] payloads it is convenient in systems where metadata is being converted between IP and SDI domains. However, it is also limited by the constraints of ST 291 [52] ANC packets. Typical applications for ancillary metadata are:

- Carriage of timing markers for down-stream ad insertion.

- Enforcement of Digital Rights Management (DRM) restrictions on specific programs and regions.

- Carriage of audio metadata.

Ad insertion timing and DRM is typically achieved via requests defined SCTE-104 [53]. Audio metadata carriage over ST 291 [52] is defined in SMPTE ST 2020 [55].

The SMPTE ST 2110-41 updated standard, to be published in 2022, is not based on ST 291 [52] but rather provides a new metadata transport that is relatively unconstrained in terms of functionality or bandwidth limitations. While existing applications like synchronized subtitles (closed captions) are in scope, audio metadata is expected to be one of the first applications to use ST 2110-41 [51].

Immersive and personalized audio requires the conveyance of a document describing the audio elements in the stream. This is a departure from current practice where the audio stream configuration (monaural, stereophonic, multi-channel) is inferred from the number of audio channels. The Audio Definition Model [53] provides these metadata and is used in file-based immersive audio production today. Serial ADM [54] builds on ADM by providing a frame-based serialization method. A proposal was submitted to SMPTE to define a packing method for Serial ADM [54] in ST 2110-41 thus enabling the carriage of advanced audio bundles over IP infrastructure. While the repetition rates for Serial ADM [54] is not limited, it is expected that the video frame rate will be typically used. Using 50 fps, a typical multi-language use-case requires a bit rate of 8 Mbps. As the metadata is required to render the audio at the endpoint, the QoS requirements defined in AES67 [57] apply.

### 4.2.3 Secure Reliable Transport (SRT)

Secure Reliable Transport (SRT) [5] is an open-source media transport protocol that uses the UDP transport protocol. It has been presented to IETF as a potential candidate for standardisation. SRT provides connection and control, reliable transmission similar to TCP at the application layer. It supports packet recovery while maintaining low latency. SRT also supports encryption using AES.

The protocol was derived from the UDT project, designed for fast file transmission. UDT provides its reliability mechanism by using similar methods for connection, sequence numbers, acknowledgements and retransmission of lost packets. UDT uses selective and immediate (NACK-based) retransmission.

SRT has all these features, but also adds several more to support live streaming mode:

1. Controlled latency, with source time transmission (timestamp-based packet delivery).

2. Sender bandwidth control.

3. Conditional “too late” packet dropping (prevents head-of-line blocking caused by a lost packet that wasn’t recovered on time).

4. Eager packet re-transmission (periodic NACK report).

SRT can be used to convey any suitable application payload, including MPEG‑2 Transport Stream [43] and RTP [44].

### 4.2.4 Reliable Internet Stream Transport (RIST)

Reliable Internet Stream Transport [6] is an open source, open specification transport protocol designed for reliable transmission of media over lossy networks (including the internet) with low latency and high quality. It is currently being developed and maintained by the Video Services Forum (VSF).

Technically, RIST seeks to provide reliable, high performance media transport by using RTP/UDP at the transport layer to avoid the limitations of TCP. Reliability is achieved by using NACK-based retransmissions to realise an Automatic Repeat Query (ARQ) capability. SMPTE-2022 Forward Error Correction can be combined with RIST but is known to be significantly less effective than ARQ.

RIST Simple Profile [7] was initially published by the VSF in October 2018 and revised in June 2020. It includes the following features:

- The base stream uses RTP for compatibility with existing equipment.

- Retransmission requests use RTCP. Two types of retransmission requests are defined:

- A Bitmask-based NACK, defined as a Transport Layer Feedback message in section 6.2.1 of RFC 4585 [42].

- A Range-based NACK, defined by [7] as an application-specific (APP) RTCP packet (see also section 6.7 of RFC 3550 [44].

- Bonding of multiple links for load sharing.

- Seamless switching using SMTPE-2022-7 [25].

- Out-of-band transmission of protection data (retransmissions may use a separate link).

- RTT Echo Request / Response procedure to estimate the round-trip time.

RIST Main Profile [8] was published in March 2020 and adds the following features to Simple Profile:

- GRE-in-UDP encapsulation based on RFC 8086 [45], with bidirectional send/receive in the same tunnel.

- Multiplexing of multiple streams into the same tunnel.

- In-band data support in the tunnel, useful for remote management.

- Client/Server architecture.

- Firewall traversal.

- DTLS encryption or Pre-Shared Key encryption, with multicast support, access control, and authentication.

- Advanced authentication options using either public key certificates or TLS-SRP.

- Bandwidth optimization based on null packet deletion.

- Support for high bit-rate streams by extending the size of the RTP sequence number space.

NOTE: RIST Simple Profile does not require or recommend any RTP payload format. As result, deployments may embed for example HEVC frames into an MPEG2-Transport Stream container according to RFC 2250 [46] or directly into RTP according to RFC 7798 [47].

### 4.2.5 Network Device Interface NDI

Network Device Interface (NDI®) [11] is a software solution developed by NewTek™ to enable video-compatible products to communicate, deliver, and receive high-definition video over a network in a high-quality, low-latency manner that is frame-accurate and suitable for switching in a live production environment. In contrast to SRT and RIST, NDI is intended to transfer media streams within a facility, not for contribution over the public networks.

NDI is designed to run over gigabit Ethernet. The table below lists the approximate bandwidth required by NDI codec [x6] for different video streams.

Table 4.2.5-1:

|  |  |
| --- | --- |
| Video stream | Approximate bit rate required by NDI codec |
| 2160p60 | 250 Mbps |
| 2160p30 | 200 Mbps |
| 1080p60 | 125 Mbps |
| 1080i60 | 100 Mbps |
| 720p60 | 90 Mbps |
| SD | 20 Mbps |

By default, NDI uses the multicast DNS (mDNS) discovery mechanism to advertise sources on a Local Area Network (LAN), although two other discovery modes (NDI Access, NDI Discovery Server) allow for operations across different subnets. When a source is requested, a TCP connection is established on the appropriate port with the NDI receiver connecting to the NDI sender. NDI 3.x has options to use UDP multicast or unicast with Forward Error Correction (FEC) instead of TCP, and can load balance streams across multiple Network Interface Controllers (NICs) without using link aggregation. NDI 4.0 introduces multi-TCP connections.

NDI carries video, multichannel uncompressed audio and metadata in XML form. Metadata messages can be sent in both directions allowing the sender and receiver to message one another over the connection with arbitrary metadata. This directional metadata system allows for functionality such as active tally information (on-air program/preview). NDI Receivers can opt to connect to various combinations of streams, to support things like audio-only or metadata-only connections where video is not required.

### 4.2.6 IP Media eXperience (IPMX)

IPMX (IP Media eXperience) is a recent initiative of the Alliance for IP Media Solutions (AIMS) to provide a standards-based approach for “Pro-AV” IP applications, such as in conference rooms, for digital signage etc., which might otherwise use HDMI or an Ethernet- (rather than IP-) based protocol such as SDVoE or HDBaseT.

IPMX adapts the SMPTE ST 2110 [21] specifications to provide a lower-cost approach to synchronisation – it still uses PTP but does not require boundary switches – and a timing model that is possibly better suited to software implementation. It uses mezzanine compression (JPEG-XS [50]) and NMOS discovery and connection (see below). It supports HDCP content protection.

At this time IPMX is still in development with few products available and it is too soon to comment on its interoperability.

### 4.2.7 Comparison Table

Table 4.2.7-1: Comparison of media transport protocols

| Parameter | ST 2110 | SRT | RIST | NDI | IPMX |
| --- | --- | --- | --- | --- | --- |
| Intended use | High quality facility and OB operations | Contribution over unreliable links (e.g., public internet) | Contribution over unreliable links (e.g., public internet) | Transfer of media streams within a facility | “Pro-AV” applications such as conference rooms, digital signage, etc |
| Proprietary/Opensource | Open standard | Opensource | Opensource | Proprietary | Standards |
| Based on protocol | RTP | UDT | RTP, e.g. TS-over-IP | TCP/UDP | RTP |
| Interoperability | wider vendor support and community of practice | Can be limited between different vendors | Good | Partially limited due to proprietary nature | Too soon to comment |
| Latency | uncompressed very low  compressed under 2 lines | Configurable, 4 × RTT of the link is recommended | Configurable, 4 × RTT of the link is recommended | Practically one field latency, might be as low as 8 scan lines | “Sub frame” |
| Error correction |  | FEC/ARQ | FEC/ARQ | TCP or FEC |  |
| Security | Designed for closed networks | Transport encryption | Transport encryption | Designed for closed networks | Support for HDCP |
| Authentication | NMOS | Supported, PSK based | Supported, PSK and DTLS based | Not supported natively |  |
| Multicast | Supported | Not supported | Supported | Supported | Supported |
| Multiple links | Supported | Not supported | Supported | Supported | Supported |
| Codec | Uncompressed, JPEG XS, ST 2042-1 (VC-2), potentially more in future | Codec agnostic | Codec agnostic | Built in | JPEG XS or other |

Editor’s Note: it would be excellent of we can add an idea on reliability requirements.

### 4.2.8 Other Protocols

A number of other protocols exist for the carriage of audio and video data such as ST 2022-6 (encapsulated SDI) as well as various proprietary solutions. There are also solutions such as HDBaseT, AVLC, SDVoE Dante AV which support other workflows such as conference and event production.

### 4.2.9 Audio Networking Solutions

DANTE, RAVENNA, QLAN, LiveWire+, WheatNet-IP can be considered as complete audio networking solutions, i.e. offering a complete networked audio systems. While each audio networking solution offers in-system connectivity, previous to the appearance of AES67 there was no standard to provide inter-system connectivity, thus leading to incompatibility between devices implementing different audio networking solutions.

AES67 [57] is not a complete audio networking solution but it does specify a mode of operation that allows interoperability between audio devices implementing different audio networking technologies (or audio “complete” networking solutions). Thus, AES67 is a complement to the existing audio networking technologies but not in direct competition with them.

AES67 defines a set of common protocols and standards to achieve that compatibility/interoperability. Like ST 2110 it uses RTP streams, and (with care) AES67 and ST 2110-320 audio systems can interoperate.

## 4.3 Codec choice

In order to transport audio and video data over bandwidth-constrained networks there is a need to encode and decode video and audio.

To achieve the optimum balance of needed bandwidth, quality and latency there are a number of different codecs solutions that are found in a production workflow.

Different categories of production tend to use different codecs. For instance, a Tier 1 event would prioritise a high-quality, low-latency mezzanine codec over a highly compressed codec that would be better suited to a news environment. This choice is influenced by both the subject matter being captured and the time taken to encode and decode the video and audio. The table below describes some common use of various codecs.

There are many options for audio and video codecs and they have different applications. Some are more suited to distribution of content, some for file-based processes such as post-production and some for live production and contribution. Table 4.3-1 below highlights some common usage scenarios for live production and contribution, but specific applications may substitute similar types of codecs or codec structures which may depend on proprietary infrastructure, licensing issues or interoperability with downstream process.

Table 4.3‑1: Codec comparison by production type

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Production Type | Codec | Bandwidth for Full HD | Common Use | Reasons | Strength | Weakness |
| Tier 1 | JPEG XS/‌VC2 | >100 Mbit/s | Compressed high quality low complexity | Very low latency encoder can handle complex scenes | High quality and low latency. ST 2110 compatibility | Requires high bandwidth |
| H.264/AVC | <20 Mbit/s | Reverse video, monitoring | Lower quality video with low bandwidth so suitable for not critical applications | Lower latency encode requiring less compute than H.265 | Not as efficient as H.265 |
| H.265/HEVC | <20 Mbit/s | higher quality video but still compressed | Efficient coding for load bandwidth applications | Requires more compute power to encode than H.264 |
| Tier 2 | H.264/‌H.265 | ~50 Mbit/s | Production/‌contribution | Highest quality video with reasonable compression | Large user base, common decoders | Highly compressed so noticeable artifacts on complex scenes |
| NDI | ~110–120 Mbit/s | Multi-camera IP production remote working | Large knowledge base and easy for smaller scale workflows | Wide user community | No timing and does not scale to large facility/OB operations |
| Tier 3 | H.264/‌H.265 | <20Mbit/s | Contribution links | Reasonable picture at low bandwidth | Good for ‘talking heads’ and non complex scenes | Not good for fast movements |
| NDI – HX | ~ 8-20 Mbit/s | Mobile journalism contribution | Low bandwidth | Easy to deploy on mobile devices and runs on poor quality networks | Very low bandwidth |
| NOTE 1: H.266/VVC is currently too complex for low latency applications but as it develops we may see its usage increase to replace H.264 and/or H.265.  NOTE 2: Codecs are defined for full HD (1920×1080) but all will support higher resolutions but with an increase in bandwidth and latency. | | | | | | |

## 4.5 Review of existing orchestration and control solutions

### 4.5.1 General

The professional broadcast industry uses a range of legacy and proprietary approaches that have been developed over many years to provide operational control. The lack of a consistent approach to interoperability has caused complexity in the architecture and integration of broadcast facilities.

A broadcast facility typically uses equipment from multiple vendors accessed through a “broadcast control system” which integrates with the different vendor-specific control protocols. Examples of broadcast control systems (alphabetically by manufacturer) include:

- Atos BNCS

- BFE Silknet

- EVS Cerebrum

- GrassValley Orbit

- Lawo VSM

- Nevion VideoIPath,

- Pebble Control

- TSL TallyMan

The Networked Media Open Specifications [15] have been developed as a response to this problem as the industry transitions to an all-IP approach. The set of specifications is primarily used for media orchestration and control purposes. Media orchestration refers to the procedures of instantiating needed media processing functions in virtualized environments and providing the control functionality for workflow management. The control functionality can be broken down into three main areas:

1. Discovery and registration: Procedures to register and identify all available functions in the media production network and their capabilities.

2.- Media Routing configuration: Define sources and sinks for media related traffic flows.

3. Operational control: Changes during operations, such as changing capture setting.

### 4.5.2 AMWA Network Media Open Specification (NMOS)

The Networked Media Open Specifications (NMOS) [15] is a family of specifications produced by the Advanced Media Workflow Association (AMWA) related to networked media for professional applications. NMOS was created to help enable automation in live IP-based architectures through control plane APIs that are built on typical patterns used for web services (REST, publish-subscribe). NMOS specifications are increasingly being adopted for applications using SMPTE ST 2110, and are part of the EBU’s Technology Pyramid for Media Nodes [14][16] reproduced in Figure 4.5.2-1 below.

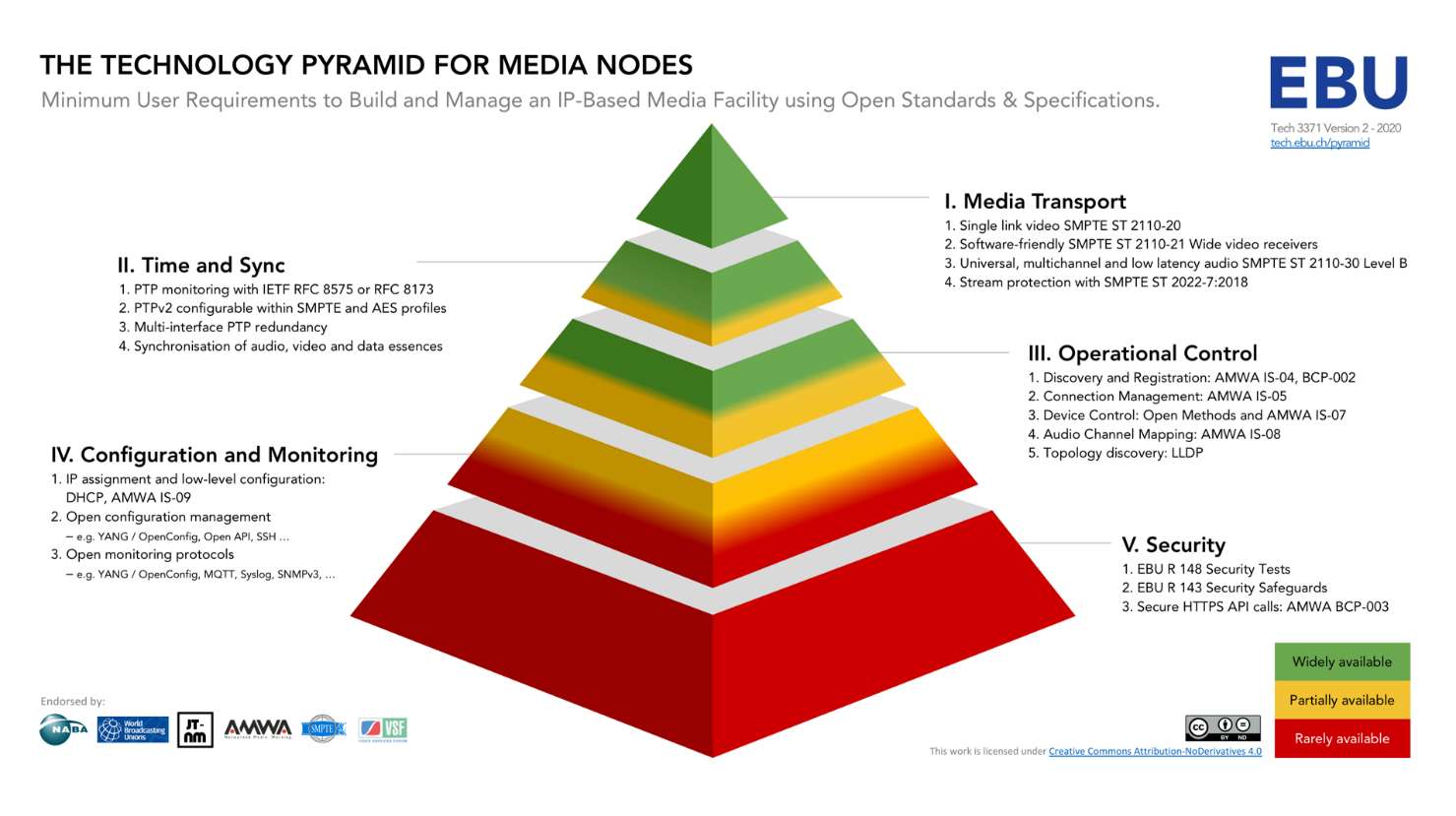
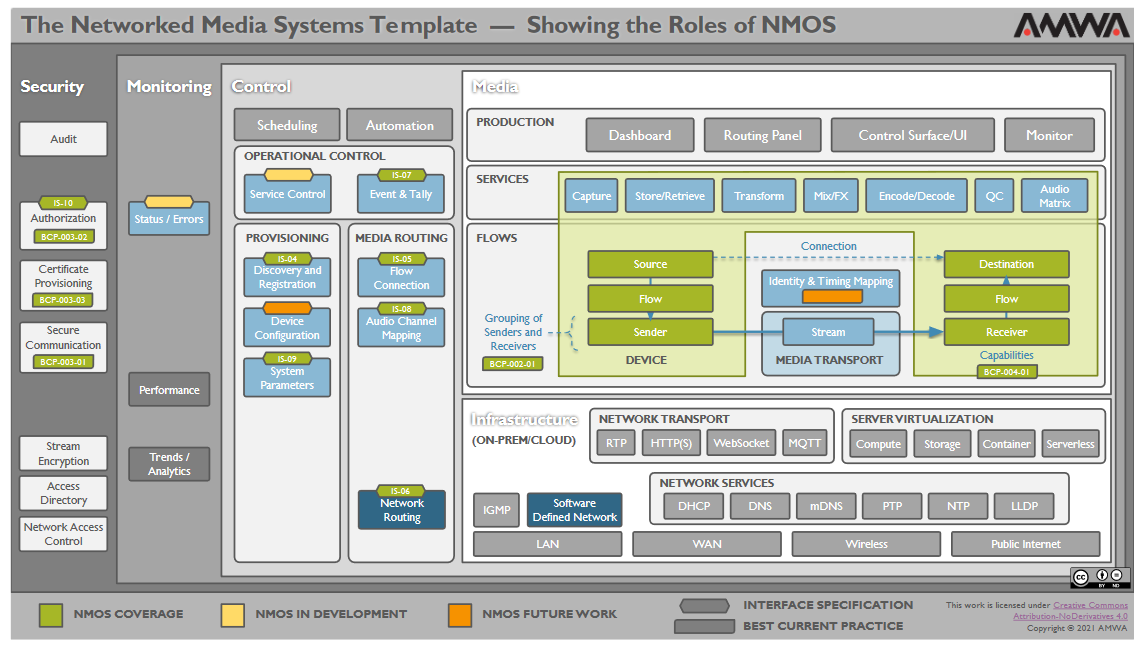


Figure 4.5.2-1: EBU’s Technology Pyramid for Media Nodes (with the permission of EBU)

AMWA has defined a system template containing several building blocks in [18]. The system template contains four distinct layers, namely Media & Infrastructure, Control, Monitoring and Security. Figure 4.5.2-2depicts Figure 3 from [18] for convenience.

Figure 4.5.2-2: Networked Media Systems Template – Showing the Roles of NMOS (Figure 3 from [18]) (with the permission of AMWA)

The Control layer contains:

- *Provisioning functions:* Discovery and Registration, Device Configuration and System Parameters.

- *Media Routing functions:* Flow Connection, Audio Channel Mapping and Network Routing.

- *Operational Control functions:* Service Control and Event & Tally.

The Media layer is subdivided into Production, Service and Flows. For the present study, the content within the Flows box is mostly of interest. Flows in this context are sequences of video, audio or time-related data, and are configured and controlled using the Flow Connection tool from the Control layer.

Further details of NMOS can be found at [15], and the specifications are documented at [20].

The most relevant NMOS specifications are depicted in also Figure 4.5.2-2:

- AMWA IS-04 allows media nodes (i.e. networked media devices) to register themselves, along with what they are (or are capable of) sending or receiving, and allows control applications to query this information.

- AMWA IS-05 allows control applications to set up and remove connections between media nodes.

- AMWA IS-07 provides a publish-and-subscribe channel for sending time-based events such as tally information.

- AMWA IS-08 specifies how to handle audio channels in NMOS APIs.

- AMWA BCP-002-01 provides grouping of related resources, e.g. video, audio and data senders.

- The AMWA BCP-003 suite of specifications (including IS-10) covers secure communication and authorisation of NMOS APIs.

- AMWA BCP-004-01 lets a receiver describe any constraints on the types or parameters of streams it can receive.

To date NMOS has mostly been used with ST 2110 [21] uncompressed multicast video and audio streams within wired facilities. However, NMOS can be used with other types of streams, including unicast. There is growing interest in other areas, such as professional audio-visual applications using compressed video (e.g. IPMX – see clause 5.2.6 above – uses NMOS), and where media is streamed between facilities over WAN connections ([VSF WAN group](https://protect2.fireeye.com/v1/url?k=f78a2ded-a81114d1-f78a6d76-86b1886cfa64-7f2369187e6a709e&q=1&e=e5962c62-ed05-4343-9d31-d4289015984d&u=https%3A%2F%2Fvsf.tv%2FSMPTE_ST_2110_over_WAN.shtml)).

## 4.5.3 Camera control and configuration protocols

#### 4.5.3.1 General

Control of UE equipment such as cameras, microphone and monitors can be broadly divided into two functions.

1. *Configuration*: The act of setting up a specific set of equipment to support specific production workflows. This includes the choice of codec, frame and sample rates as well as vendor-specific functions.

2. *Control:* Used to denote functions that will change during the production process such as focus, exposure or zoom.

In general, configurations are vendor-specific as they access root layer functions that are not common to all manufacturers. Control tends to be more open and indeed may need to support devices from more than one manufacturer e.g. a camera from Sony mounted with a Canon lens.

#### 4.5.3.2 Camera control protocols

For basic camera control such as pan, tilt, zoom, focus, iris, start, stop, etc. there are a number of relevant technologies, some of which include:

- LANC is an old serial remote control protocol for camcorders that is still widely supported.

- VISCA is a serial protocol, now mapped to IP, for control of PTZ surveillance and similar cameras

- ONVIF is an industry group that produces (SOAP/WSDL) web services for control of PTZ surveillance and similar cameras.

- Vendor-specific protocols and APIs (e.g. Blackmagic Camera, NDI PTZ API).

For more advanced control (as required for some broadcast applications) interoperability is more of a problem, because cameras typically use proprietary and vendor-specific control protocols via a camera control unit (CCU).

### 4.5.4 EMBER+

EMBER+ is a lightweight control and monitoring protocol designed by L-S-B Lawo Group that is supported by devices from broadcast manufacturers. It has an open source SDK [34], with the last significant features added in February 2019.

### 4.5.5 Other Protocols

NDI (see clause 4.2.5) provides discovery on a local network using multicast DNS-SD or between networks using NDI Acces or NDI Discovery Server. NDI also provides an API for camera pan/tilt/zoom (PTZ) control.

A number of control/management standards and specifications are used with audio devices, including:

- AES70 aka OCA (Open Control Alliance), a full-featured control architecture developed by Bosch.

- IEEE 1722.1 provides Discovery, Enumeration, Connection management and Control for AVB applications.

- MIDI and OSC, in particular for music applications. MIDI 2.0 provides significant enhancements over 1.0.

- SNMP is used in some applications.

However, none of these are universally adopted, and in practice many networked audio environments rely on the control layer provided with DANTE [82].

Recently, there has been interest in use of YANG [85] and NetConf [86] for device control.

## 4.6 Evolving Internet media transport protocols

### 4.6.1 General

Clause 4.2 describes existing media transport protocols used in media production operations. Because of the constantly evolving nature of the Internet, there exist new media transport protocols that are either incomplete or not yet adopted in the media production space. These may include support for new features or payload formats that are not included in existing solutions.

Table 4.6.1-1 below expands the feature comparison in table 4.2.7‑1 with the media transport protocols discussed in the following clauses.

Table 4.6.1-1: Comparison of future media transport protocols

| Parameter | QRT |
| --- | --- |
| Intended use | Contribution over unreliable links (e.g., public Internet) |
| Proprietary/‌Opensource | Open standards |
| Based on protocol | RTP, QUIC |
| Interoperability | Experimental |
| Latency | Configurable |
| Error correction | FEC/ARQ |
| Security | Transport encryption |
| Authentication | TLS client certificate |
| Multicast | Not yet specified |
| Multiple links | Supported, using multipath extension |
| Codec | Codec-agnostic |

### 4.6.2 Tunnelling RTP media sessions over QUIC

RTP media sessions [44] can be carried over the QUIC transport protocol specified in RFC 9000 [51] using the unreliable datagram extension specified in [64]. A survey of some recent proposals to standardise this usage are found in [65], along with relevant use cases and requirements. Of particular interest in the present document, the QUIC RTP Tunnelling (QRT) [66] and RTP over QUIC [67] proposals both specify a means to multiplex RTP media over a QUIC transport connection, allowing for secure transmission of media flows over lossy IP networks (including the Internet) with tuneable latency and quality parameters.

QRT [66] and RTP over QUIC [67] specify a mutually interoperable lightweight multiplexing layer on top of the QUIC unreliable datagram extension [64], allowing multiple RTP sessions to be multiplexed together into a single encrypted packet flow.

In addition, reliable streams may be multiplexed into the same QUIC connection as the RTP media flows to exchange data using application protocols requiring reliability, such as HTTP/3 [68] and/or the Session Initiation Protocol (SIP) specified in RFC 2543 [69] used in combination with appropriate session announcements [70]. The former may be used, for example, to convey NMOS [15] configuration and control messages as introduced in clause 4.5.2.

NOTE: The use of SIP or NMOS may be convenient for in-band call control in certain scenarios, such as remote production contribution links.

Using RTP as the basis for media transport, QRT and RTP over QUIC can leverage the substantial existing feature set of already-deployed RTP solutions, including:

- Support for any codec and packaging format that has an associated RTP payload format.

- Support for *Forward Erasure Correction (FEC)*, including SMPTE 2022-1 [23].

**-** *Automatic Repeat Query (ARQ)* requests by means of in-band RTCP packets using the bitmap-based RTP Retransmission payload format specified in RFC 4588 [70].

- RIST’s range-based NACK retransmission mechanism [7] (as described in clause 4.2.4) may additionally or alternatively be used in this context.

In this respect, QRT and RTP over QUIC offer a similar feature set to RIST Main Profile, as described in clause 4.2.4.

The following optional RTP-related features are additionally available in comparison with current RIST specifications:

**-** *Congestion Control (CC)* through the use of RTCP-signalled feedback mechanisms, such as that described in RFC 8888 [72], alongside congestion control algorithms such as Network-Assisted Dynamic Adaptation (NADA) [73], Self-Clocked Rate Adaptation for Multimedia (SCReAM) [62], Google Congestion Control (Google-CC) [75] or Shared Bottleneck Detection [64].

By using QUIC, which has a predominant use in underpinning HTTP/3, QRT also inherits the following features:

- It is well understood by many application firewalls and proxies.

- Because connections are identified by a pair of abstract connection identifiers (rather than by a traditional 5‑tuple) active QUIC connections can be migrated between network endpoints without performing the security connection handshake again and without interrupting application-level flows.

- By probing additional network links before performing a connection migration, minimal delay and/or interruption is incurred.

A draft multipath extension [77] recently adopted by the IETF QUIC Working Group allows a pair of QUIC endpoints to use multiple network paths simultaneously (i.e. "link bonding"), either to increase the aggregate capacity of a connection, or to improve the robustness/resilience of the connection, or a combination of these.

# 5 Relevant media production use cases

## 5.1 General

Audio/Visual (AV) production includes television and radio studios, outside and remotely controlled broadcasts, live news gathering, sports events and music festivals, among others. All these applications require a high degree of reliability, since they are related to the capturing and transmission of data at the beginning of a production chain. This differs drastically when compared to other multimedia services because the communication errors will be propagated to the entire audience consuming that content either live or via recording. Furthermore, the transmitted data is often post-processed with nonlinear filters which could actually amplify defects that would be otherwise not noticed by humans. Therefore, these applications call for high quality data, and very low probability of errors. These devices will also be used alongside existing technologies which have a high level of performance and so any new technologies will need to match or improve upon the existing workflows to drive adoption of the technology.

The performance aspects that are covered by/in TS 22.263 [3] (Service requirements for Video, Imaging and Audio for professional applications) also target the latency that these services experience.

In recent years, production facilities have moved from bespoke unidirectional highly specialised networks to IP-based systems and software-based workflows. This migration is expected to continue, and wireless IP connectivity is key to a number of these workflows.

Typical setups consist of multiple devices such as cameras, microphones and control surfaces that require extremely close synchronisation to maintain consistency of video and audio. Often devices need to communicate directly to each other, for instance a camera to a monitor or a microphone to a Public Address (PA) system.

Video and audio applications also require extremely high quality of service metrics as the loss of a single packet can cause picture or sound breakup in the downstream processing or distribution. Often this is a legal, regulatory or contractual agreement to maintain a high-quality, stable and clear video or audio signal.

Today’s digital AV network transport is typically handled separately for wireless and wired transfers. Wireless AV transmissions are implemented with application-specific solutions that allow deterministic data transport of a single isolated audio or video link. Wired AV transmissions are typically either Ethernet- or IP-based. Network Quality of Service in AV IP networks is mainly achieved with IP DiffServ/DSCP-based prioritization of packets in network switches. This method is sufficient for most AV use cases since jitter resulting from packet collisions is small, for example in the order of 10 µs per concurrent data stream in gigabit Ethernet.

Live video production is a complex subset of production activity that typically is served by evolving specialized technologies, networks and radio solutions. The high bandwidth and low latency required to produce real-time high-definition video requires dedicated point-to-point connections that have evolved from analogue production, via digital, to IP-based solutions. Current IP solutions for the studio are based on managed wired networks and the mobility required by cable-free cameras, microphones and monitoring have been adapted to interface with these networks via gateway devices but still supporting legacy integrations.

The COVID-19 pandemic has also led to an increase in distributed production where control surfaces are not necessarily co-located with the equipment they control. Cloud-based solutions are emerging to support these workflows and this use case should support distributed compute functionality.

Other technologies used include optical fibre for fixed links, satellites and the physical transport of media storage devices with previously recorded content. In this sense, wireless connectivity plays a major part in production where there is a need to have mobility, flexibility and reliability.

## 5.2 Scenario 1: Wireless cameras within a production workflow

Different types of network may be deployed depending on how the camera is used and the number of them. For a single point-to-point link, a dedicated peer-to-peer solution can be achieved with a simple transmitter and receiver set up. These may use either omnidirectional or directional antennas. For more complex setups, such as a studio or sporting event, a network providing wider area coverage (e.g. a mesh network with multiple receivers) may be set up. This allows the cameras to move freely within the coverage area while maintaining a consistent network Quality of Service. Finally, for large area events, aerial relays may be deployed to cover a moving camera on the ground.

While these solutions are extremely robust, they do require specialist skills and knowledge to set up.

When deployed in real world scenarios these types of camera are usually incorporated into production workflows consisting of other cameras that are connected directly to the production network by fibre or coaxial cables. In this scenario it is important that the latency of any radio-connected device is minimised so that any cuts between a wired and wireless camera are synchronised. This is currently done by sending a special signal to an on-board clock generator that time-synchronises the various functions of the camera to match other cameras in the network.

There are also requirements for near-real-time responses to instructions or control of a camera. If, for instance, the focus of the camera is controlled remotely then the operator will need to see the image in under 100 ms in order to be able to respond and control the lens on the camera.

Cameras used for this type of production are usually highly specialised and have a modular design with various elements such as a lens, viewfinder and microphones added as required. Different cameras rely on different protocols to control various elements, but there are also some standard protocols that are used where specialist control is not required. Some signals, such as lens control, will pass through the camera unit itself, while others will connect directly to the end user device.

Within Media Production scenarios, the wireless camera act as a UE. Multiple, partially optional application flows are between the wireless camera and one or more network side media production function.

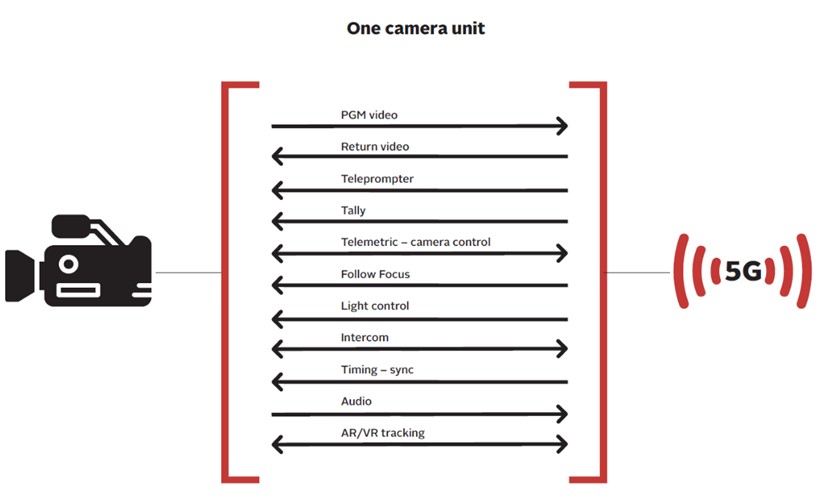


Figure 5.2-1: Flows by one camera unit

Figure 5.2-1 illustrates a set of important data flows, namely:

- *PGM Video (Program Video):* The uplink video stream.

- *Return video:* In some production events the camera receives a return video and renders it in the viewfinder. The return video may be a CGI- enhanced version of the captured video, or a video stream from a different camera. The camera operator considers the return video when composing the camera shot.

- *Teleprompter:* In some production events a speaker in front of the camera reads from a rolling script projected directly in front of the camera lens through a half-silvered mirror.

- *Tally:* the small red light indicating which camera is “on-air”.

- *Telematics – Camera Control:* Different functions of the camera, such as the shutter speed and iris can be locally or remote controlled. The telematics signal may also contain information about the camera status, such as battery level.

- *Follow Focus:* A focus control mechanism to help the operator be more precise while adjusting the focus and maintaining it while the camera is moving relative to the subject/object.

- *Intercom:* In some production events, the camera operators can talk to each other and the programe director using a separate speech channel.

NOTE: Intercom is traditionally integrated into a camera. However, an intercom might become more and more independent in media production, since intercom typically is set up first and torn down last.

- *Timing – Sync:* The camera needs to time synchronized, (A) for timestamping the media packets and (B) for synchronizing the frame capture pulse (GenLock).

- *Audio:* In some production events (specifically news gathering), the camera is equipped with a microphone to capture audio. In other production events (like sports), the microphone positions are different from camera positions to capture “atmosphere”.

- *AR/VR tracking:* Accurate camera positioning is of paramount importance to incorporate virtual and augmented reality studio sets in live productions.

## 5.3 Scenario 2: Outside Broadcast contribution

Over the past few years, broadcasters have been using mobile networks for some workflows, specifically using 4G networks to send a live video stream to a production centre. This type of communication has helped revolutionise the way news and events are produced, as reporters and teams can work from anywhere, at any time if an acceptable coverage is available. To do this, a backpack or camera-mounted device is used to encode and broadcast video without the need for mobile units (vans) and/or many cables and devices.

However, the current use of these technologies requires additional functionalities to overcome the lack of QoS on unmanaged mobile networks, which brings several disadvantages. For example, due to the bandwidth required, mobile solutions require multiple connections and therefore multiple SIM cards to provide adequate service; this method of connection aggregation is known as “link bonding”. Additionally, when these devices are outside the mobile network provider coverage area, other SIM cards are required to use an alternative network. The video must be highly compressed due to network bandwidth restrictions, which degrades content quality in later stages of the production and distribution chains. These technologies usually provide a single video link. When more than one camera is required it either needs multiple units (that are often timed differently) or staff to support multiple camera operation with on-site infrastructure. There is also no differentiation between the networks to which these devices connect and public networks, so in large events 4G connections become unreliable as they compete for connectivity and bandwidth with other users.

It can be expected that 5G solutions will evolve to meet these workflows with little or no interventions but there is also a demand for a technology that allows multiple audio and video sources to be connected and synchronized as well as better interoperability with existing workflows.

The scenarios for contribution may be focused on newsgathering and lower budget production. In these scenarios content may be more static with less temporal change or fixed backgrounds, so more intense compression may be applied.

## 5.4 Considerations on remote and cloud-based production

Productions typically require long preparation times with large audio and video equipment that is physically moved to external event sites, as well as configured and adjusted for a specific production activity. Remote Production enables remote control of audio–visual capture equipment (such as microphones and cameras) deployed at an outside broadcast site from a more convenient production location, typically a broadcast centre. Remote Production thereby reduces the requirement to move all production equipment to the outside broadcast site. This may lead to cost reductions or allow more coverage of complex events. For example, multimedia sources such as cameras or microphones would be deployed at the outside broadcast site, but much of the equipment may be in production centres and be connected over the network to the remote site. Examples include audio and video mixers, switching matrices, storage devices and multi-viewers.

Some functions are coordinated in master control rooms (MCRs). These MCRs pull together multiple internal and outside sources and organise them for presentation to operational galleries. Large broadcast centres have signal routing matrices that allow multiple audio and video signals to be organised and packaged for both incoming and outgoing feeds.

TR 22.827 [4] includes the following definition:

***Remote Production****: Content being acquired is remote to the broadcast centre but configured and controlled from the broadcast centre. This may include video or audio content but also command and control functions to operate the technical facilities located at the outside broadcast site.*

Cloud-based production is a special case of Remote Production in which workflows are executed in a cloud-based infrastructure. This cloud-based infrastructure can be public or private and may even be deployed within the 5G operator’s infrastructure itself (e.g. leveraging Edge Computing capabilities close to the production location).

A 5G NPN could allow audio–visual capture equipment (such as cameras and microphones) deployed at an outside broadcast site to connect to a production facility, whether the latter is local or remote, and whether it is operated within a central broadcast centre with the support of fixed equipment or deployed in a cloud infrastructure. The various application flows, latency and bit rate requirements depend on the scenario envisaged and should be studied.

## 5.5 Collaboration models and deployment architectures

### 5.5.1 General

This clause describes various collaboration models with different NPN deployments, targeting the different media production scenarios, which are introduced in previous clauses.

The following (simplified) media functions are deployed in different models:

- *Technical Manager:* This function represents a role within the media producer that decides on various options, e.g. how many media production devices (cameras, monitors, etc) are used in the deployment, their connections, etc.

- *Dynamic Configuration:* This function translates the decisions of the technical manager role into (dynamic) configurations. For each device, it determines the network connectivity and media configuration, such as codec configuration (separately for uplink and downlink return path), selection of Media Gateway (IP address and port), media protocols, etc. When the traffic crosses trust domains, the Dynamic Configuration function also configures security functions, e.g. to secure the media plane traffic. When the configuration setup is finished, the Dynamic Configuration function provisions the needed QoS flows in the PCF/NEF. For each QoS flow, the Dynamic Configuration function provides traffic detection information (e.g. a Packet Filter Set or a PFD) and information about the needed QoS class.

- *Configuration Application:* A UE component, which interacts with the network-based Dynamic Configuration function. Typically, the Configuration Application listens to dyanamic configuration instructions from the Dynamic Configuration Function.

- *Media Client* (sender and receiver): The media level function. In the case of a wireless camera, this function is captures, encodes and sends the media data (typically video, optionally with audio). When return video is configured, this function is also capable of receiving, decoding and rendering the media. When the device is a display, then this Media Client only receives, decodes and renders media data.

- *Media Gateway:* A network function for sending or receiving encoded media. The Media Gateway may act as proxy.

### 5.5.2 Deployment #1: On-site wireless production with Standalone NPNs

A straightforward realization of Scenario 1 (clause 5.2) is the usage of a Standalone NPN. Here, a dedicated 5G System is deployed for exclusive use for media production. The media producer also acts as the Mobile Network Operator; thus, all Application Functions are trusted and may interact with other network functions as needed.

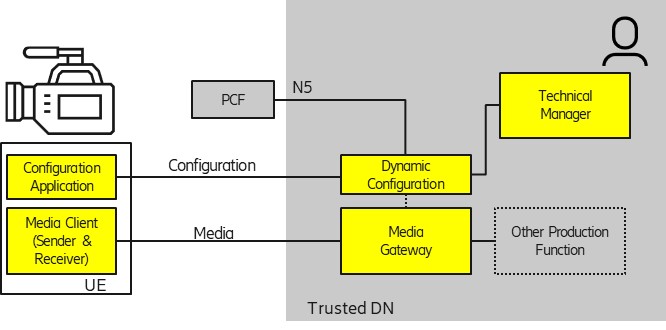


Figure 5.5.2-1: On-site production with a Standalone NPN

### 5.5.3 Deployment #2: On-site wireless production with PNI-NPN or Outside Broadcast Contribution

This deployment model contains three sub-scenarios, based on the distance between media production site and the media production network.

1. *Local PNI-NPN production with support for on-site edge computing:* A media producer may leverage the network of a Mobile Network Operator for an on-site media production event. When a local breakout in a local edge computing environment is provided, the deployment is very similar to an on-site wireless production with an SNPN (clause 5.5.2). For example, the media producer connects the equipment of an OB Van through a local breakout at an event location with the 5G PNI-NPN. Low latency communication is enabled due to close proximity of devices.

2. *Remote production:* A media producer may leverage the network of a Mobile Network Operator for remote media production. Here, media production equipment is kept more centrally in the network in order to reduce equipment and people movement, as described in clause 5.4.

3. *Contribution:* A media producer may leverage the network of a Mobile Network Operator for an Outside Broadcast contribution event, for example Electronic News Gathering (ENG) including mobile journalism. Here, the media production network elements are located more centrally within the studio facility of the media producer.

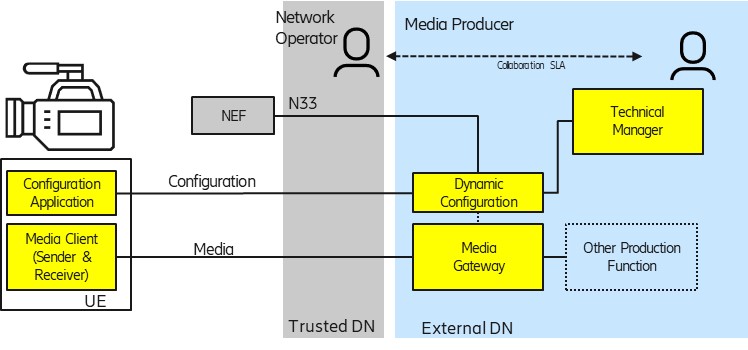


Figure 5.5.3-1: PNI-NPN collaboration model for on-site productions or OB contributions

In this collaboration scenario, the NEF APIs are the key enabler for the collaboration. Some procedures, such as the SLA definition and agreement, may be outside of the scope of the NEF APIs.

### 5.5.4 Remote wireless production with Standalone NPNs

This deployment model addresses remote production scenarios, reducing the need for moving equipment (and people) to a local production site. Remote production is described with cloud production in clause 5.4.

NOTE: The usage of cloud computing does not necessarily refer to remote production. Cloud computing instances may be deployed anywhere, including locally at the event production site.

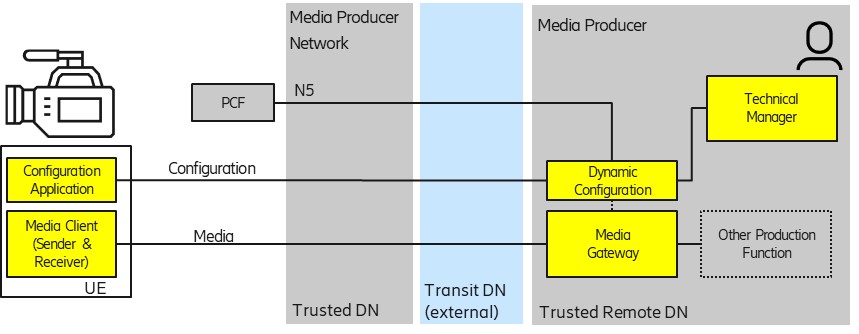


Figure 5.5.4-1: Remote production with SNPNs

This deployment model specifically addresses the use of Standalone NPNs for remote production (clause 5.5.3 describes the usage of a PNI-NPN for remote production).

No extra Service Level Agreement is needed between the Mobile Network Operator and the media producer because the network is operated by the same entity. However, SLAs may be needed for Transit DNs, which connect the Local SNPN to the Remote Data Network.

In some cases, the Local SNPN may leverage a 5G System as a Transit Data Network, so that the Media Production Network becomes portable. For example, an SNPN system for local production could be installed in an Outside Broadcast van together with other media production equipment.

## 5.6 Low-Latency Production with 5G mmWAVE

mmWAVE is also known as Frequency Range 2 defined in RAN specifications, which covers the spectrum above 24 GHz. At WRC-2019 in November 2019, several new frequency ranges for use by IMT-2000 (5G) were identified. These encompass many of the existing 3GPP bands plus the following additional spectrum ranges:

- 24.25–27.5 GHz

- 37–43.5 GHz

- 45.5–47 GHz

- 47.2–48.2 GHz

- 66–71 GHz.

Since Release 15, 5G mmWAVE technology, as part of 5G NR specifications, is well defined in 3GPP. It allows users to access the full potential of 5G by utilizing untapped frequency bands above 24 GHz. This abundant spectrum can deliver the fastest available speeds, extreme capacity and low latency.

3GPP has done several studies on the channel model for frequency spectrum above 6 GHz, e.g. in TR 38.900 [61]. According to [59], extreme higher propagation loss and penetration loss in mmWave spectrum is expected, and the frequency is sensitive to blockage, e.g. by foliage or the human body. However, again according to [59], performance tests for mmWAVE provide extraordinary KPIs:

- 14.7/3 Gbps cell peak throughout (DL/UL) in an 800 MHz spectrum band.

- One-way user plane latency between 1–1.5 ms.

- Farthest access distance: 2.6 km in line-of-sight with a few small trees.

According to [62], commercial 5G modems support mmWAVE as well as dual connectivity, and upload speeds of 2.2 Gbps can be achieved by the aggregated Frequency Ranges FR2 400 MHz (on n261) and FR1 100 MHz (n77). According to [61], mmWave bands can accommodate more capacity and bandwidth than any other band. And since spectrum in these bands is abundant, mmWave spectrum is ideally placed to deliver high speeds, low latency and high capacity, all at the same time. The short wavelength of mmWave allows for very small antennas, which helps with beam forming for enhanced coverage and spectral efficiency. Promoted by the whole industry, 5G mmWAVE commercialization grows rapidly. GSA’s report [58] indicates that, up to May 2021, twenty-eight operators in sixteen countries/territories are known to be already deploying 5G networks using mmWave spectrum at 24 GHz.

Based on this analysis, mmWAVE is an attractive technology for production scenarios that typically rely on fully wired data rates. In particular, interactive video live production scenarios, i.e. video production with almost real time interaction between video director and cameras, require extreme low latency, as they allow camera direction in almost real time. In some UHD and 8K media production cases, lightweight compression codecs (e.g. JPEG-XS [50]) are used. In some field tests, JPEG-XS could achieve less than 3 ms camera-to-screen latency with time synchronisation based on Precision Time Protocol (PTP) [22, 26]. Whereas without PTP decoder vendors usually buffer one frame to compensate for timing fluctuation, the use of PTP allows this buffer to be shrunk, reducing the camera-to-screen latency to less than 23 ms.

To achieve “Master Copy” quality, the preferred compression rate is between 1/8 – 1/12 which requires 1–1.5 Gbps transmission capacity for UHD video. Realistically, according to [62], for HD at least 150 Mbps and for UDD / 8K more than 800Mbps are needed. Due to the large payload, lightweight compression has been used over IP and Ethernet links until now. 5G mmWAVE is capable of providing data transmission at several gigabits per second, and this could be leveraged to replace existing cabled infrastructure for the transmission of lightly compressed video signals. An example setup for a mobile television studio is shown in Figure 5.6-1. In this case, the camera sends data in a local environment though 5G mmWAVE to a 5G UPF (labelled as "5G Router Server" in the figure), that is connected through cable/fiber to the telestudio.

Graphical user interface, application

Description automatically generated

Figure 5.6-1 Mobile telestudio scenario with 5G mmWAVE

# 6 Potential issues and Candidate Solutions

## 6.1 General

This clause describes and discusses a set of candidate solutions for addressing the different production use-cases and deployment options as introduced in clause 4.1 and clause 5.

- **Local Production:** The expectation for a local production is low-latency operation (a few frames of delay) at high bit rates. With today’s COFDM technology [78][79], a fixed capacity (e.g. maximum 40 Mbps in an 8 MHz UHF channel when using a DVB-T2 transmitter) is allocated to a production device (like a camera). This enables a camera to operate at low latency. The rate control of the video encoder is tuned to never exceed the allocated capacity. With 5G NPN, this can only be achieved when allocating a Guaranteed Bit Rate (GBR) QoS flow to the media. Non-GBR QoS flows may lead to situations where the UE needs to discard packets because the capacity is currently not available. Local Production may leverage specifically Deployment #1 (local SNPN deployment) or Deployment #2 (Local PNI-NPN with support for on-site edge computing).

- **Remote Production:** The expectation for a remote production is also low-latency operation at high bit rates. However, there is a possibility to compromise, e.g. slightly higher latencies than in local production but lower latencies as in contribution. Solutions need to account for some packet loss and potentially for automatic bit rate adaptation. Remote Production may leverage specifically Deployment #2 (Remote Production) or Deployment #3.

- **Contribution:** The expectation for contribution is different than for production. Contribution links often operate at higher latencies than production links. However, a very flexible and on-demand (or even spontaneous) capacity allocation is important, e.g. as needed for Electronic News Gathering (ENG). Specifically Deployment #2 (Contribution) is in focus for most contribution cases. In some cases, Deployment #3 (with a nomadic SNPN) may also be applicable.

Editor’s Note: Considerations for *Installed and Live Sound* is ffs.

## 6.2 Key Issue #1: Utilizing Available Capacity in Multi-Camera Scenarios

There is in several scenarios a need to dynamically and proactively control media rates such that not all cameras use the maximum bit rate all the time. Specifically, within a group of cameras that are used for the same live programme, there is need for reducing the rate for lower-prioritized cameras in order to protect the camera that is currently “live” (production camera) and the camera that is next to go “live” (according to the producer’s wishes). This should be done proactively, considering the radio conditions and load in the network, to avoid loss of quality on important feeds.

Usual fiber-based studio setups use 3-24 Gbit/s per camera (uncompressed, see [37]). A 5G cellular setup is obviously limited in uplink capacity compared to that. Considering this, SA1 produced a table in [3] containing also somewhat lower numbers, assuming various degrees of compression:

Table 6.2.2-1: reproduced from [3] table 6.2.1-3

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Profile | # of active UEs | UE Speed | Service Area | E2E latency | Packet error rate (Note 1) | Data rate UL | Data rate DL |
| Uncompressed UHD video | 1 | 0 km/h | 1 km2 | 400 ms | 10-10 UL  10-7 DL | 12 Gbit/s | 20 Mbit/s |
| Uncompressed HD video | 1 | 0 km/h | 1 km2 | 400 ms | 10-9 UL  10-7 DL | 3 .2 Gbit/s | 20 Mbit/s |
| Mezzanine compression UHD video | 5 | 0 km/h | 1000 m2 | 1 s | 10-9 UL  10-7 DL | 3 Gbit/s | 20 Mbit/s |
| Mezzanine compression HD video | 5 | 0 km/h | 1000 m2 | 1 s | 10-9 UL  10-7 DL | 1 Gbit/s | 20 Mbit/s |
| Tier one events UHD | 5 | 0 km/h | 1000 m2 | 1 s | 10-9 UL  10-7 DL | 500 Mbit/s | 20 Mbit/s |
| Tier one events HD | 5 | 0 km/h | 1000 m2 | 1 s | 10-8 UL  10-7 DL | 200 Mbit/s | 20 Mbit/s |
| Tier two events UHD | 5 | 7 km/h | 1000 m2 | 1 s | 10-8 UL  10-7 DL | 100 Mbit/s | 20 Mbit/s |
| Tier two events HD | 5 | 7 km/h | 1000 m2 | 1 s | 10-8 UL  10-7 DL | 80 Mbit/s | 20 Mbit/s |
| Tier three events UHD (Note 2) | 5 | 200 km/h | 1000 m2 | 1 s | 10-7 UL  10-7 DL | 20 Mbit/s | 10 Mbit/s |
| Tier three events HD (Note 2) | 5 | 200 km/h | 1000 m2 | 1 s | 10-7 UL  10-7 DL | 10 Mbit/s | 10 Mbit/s |
| Remote OB | 5 | 7 km/h | 1000 m2 | 6 ms | 10-8 UL  10-7 DL | 200 Mbit/s | 20 Mbit/s |
| NOTE 1: Packets that do not conform with the end-to-end latency are also accounted as error. The packet error rate requirement is calculated considering 1500 B packets, and 1 packet error per hour is 10-5/(3\*x) , where x is the data rate in Mbps.  NOTE 2: Could use either professional equipment or mobile phone equipped with dedicated newsgathering app | | | | | | | |

Further, Table 6.2.2‑1 in the present document shows a range of bit rates for different event types.

**Observation 1**: The data rate requirements per camera in [3] span a range of more than 1000 times, from 10 Mbit/s to 12 Gbit/s, depending on the profile/scenario.

**Observation 2**: The overall uplink capacity of a 5G system with realistic amount of radio spectrum and realistic ratio between downlink and uplink time resources, is in the same order of magnitude as the required/desired data rate for a *single* camera for tier 2 and tier 1 events.

Editor’s note: example values for uplink cell capacity are invited.

**Conclusion 1**: For multi-camera scenarios, there is a need to dynamically control media rates such that not all cameras use the maximum rate all the time, in particular when this may exceed total cell capacity.

**Conclusion 2**: For multi-camera scenarios, there is a desire from the producer’s point of view to see all cameras in pristine quality but in case of increased cell load or worsening radio conditions, there is also a need to quickly reduce media rates of some cameras to avoid data loss on important camera feeds. Specifically, within a group of cameras that are used for the same live programme, there is need for reducing the rate for lower-prioritized cameras in order to protect the camera that is currently “live” (production camera) and the camera that is next to go “live” (according to the producer’s wishes).

See clause 7.1 for candidate solutions to this issue.

## 6.3 Key Issue #2: Media Protocols on 5G: Traffic segregation and prioritization

### 6.3.1 General

This clause focuses on the usage of 5G Systems, assuming that multiple application flows – either from multiple cameras or from a single camera unit (see Figure 5.2-1) – would experience a different priority treatment by the RAN traffic scheduler and likely by the traffic policing function in 5GC. Different protocols may be used to carry media and other data.

An application flow is typically described by a 5-tuple, i.e. source and destination IP addresses (Layer 3), Layer 4 protocol and Layer 4 source and destination ports. Some protocols may multiplex multiple elementary streams (and potentially other data) into one application flow. Other protocols map one elementary stream to one application flow.

The traffic characteristics and the main flow direction (uplink or downlink) depend on the usage. For example, a program video stream, produced by a camera, is typically of higher bit rate than a return video stream.

NOTE: Some application flows may carry non-media content, for example camera control, telematics (e.g. battery status), and position information for AR tracking.

Editor’s Note: Solutions may use IP multicast or IP unicast packet routing to transport media streams. IP multicast is popular in AV Production because the same feed from a camera, microphone or talkback circuit can then be consumed by monitoring devices (screens, headphones, etc.) as well as feeding into vision mixers, sound mixers, etc. However, there are challenges to be overcome in using IP multicast over Wide-Area Networks and therefore in Remote Production scenarios.

### 6.3.2 Application flow prioritisation

Figure 6.3.2‑1 depicts the same media flows of a single camera as shown in figure 5.2-1, but categorized into three priority groups:

- Group 1, with the highest priority, comprises essential media essence flows.

- Group 2, with medium priority, comprises communications flows.

- Group 3, with the lowest priority, comprises control flows.

Depending on the media production scenario, a certain set of media flows are present. For example, a teleprompter application flow is only present when a teleprompter (autocue) device is attached to the camera.

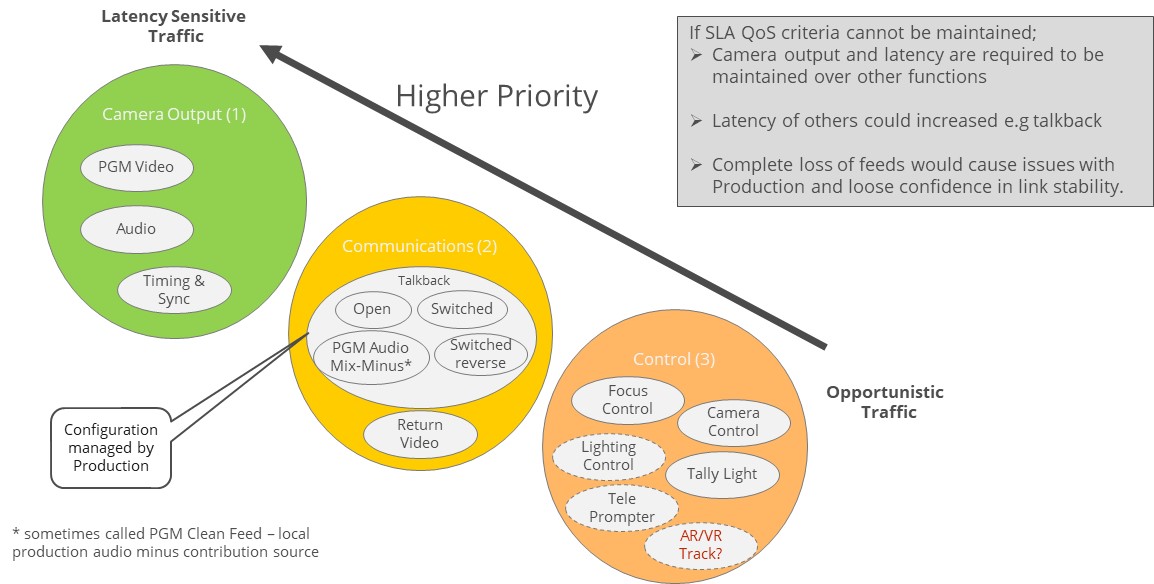


Figure 6.3.2‑1: Flow Priority

Typically, highest priority is program (PGM) video, which is always present when using a 5G-enabled camera. An audio media flow related to the program video flow is not necessarily present in all scenarios, but when present, it often has an even higher priority than the program video. A time synchronization related media flow (e.g. PTP or NTP) is always present and is essential for production.

### 6.3.3 Applying Quality of Service to application flows

Quality of Service (QoS) is a tool which only becomes relevant at times of high network utilisation. In these situations, the 5G System may need to prioritize some packets over others. In a well-planned production scenario, the 5G System is dimensioned to fit the needs of the media production and high network utilisation only occurs rarely. However, proper planning and dimensioning might not be achievable in all media production scenarios. Thus, it might be preferred to degrade the output of a camera, keeping the most essential traffic intact. Depending on the scenario, different media flows are more essential than others to the media production.

An example communication protocol stack is illustrated in Figure 6.3.3‑1 below. The different media flows may use different higher layer protocols. For audio and video streams, the RTP protocol is often used, which typically uses UDP as its Layer 4 protocol. Data streams such as tally light control may be carried using, for example, MQTT [48] (AMWA NMOS recommendation) which uses TCP as its Layer 4 protocol. (NMOS [15] is described in clause 4.5.2.)

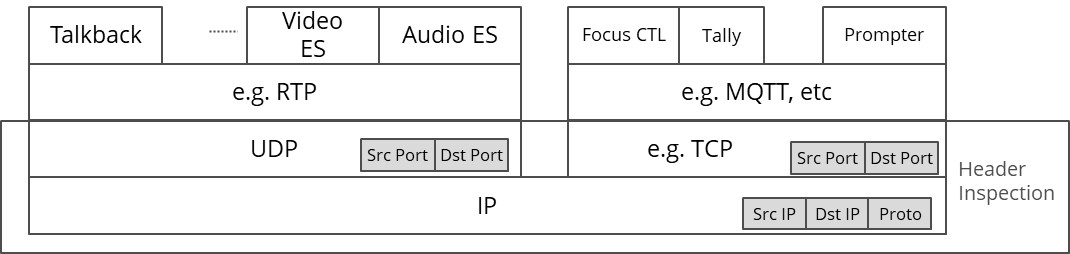


Figure 6.3.3‑1: Example protocol stacks for different media application flows

MQTT [48] is a message-oriented application protocol based on the publish–subscribe paradigm. It was developed as an OASIS open standard and published as ISO/IEC 20922. MQTT uses TCP as its transport protocol. MQTT adds some message headers, which allow (among other things) the byte-stream-oriented TCP protocol to be used for message separation.

The different combinations of media flows (figure 5.2-1) depend on the media production scenario. In the following, the mappings for some example scenarios are presented and discussed.

### 6.3.4 Solutions leveraging 3GPP QoS

#### 6.3.4.1 General

The 3GPP Quality of Service framework contains many tools to define media flow specific treatment with respect to relative priority, target bit rate, packet delay budget and packet error rate. To apply these tools, the 5G System must be able to identify the associated media flow, based on network-level parameters such as a UDP port number or an IP address. The 5G System (UE and UPF) uses packet header inspection techniques for traffic detection. Based on header inspection, each individual IP packet is associated with a QoS flow and is marked accordingly in the 5G System.

#### 6.3.4.2 Solution Example A: Coarse-grained separation with separated media

It is very common in IP-based media production scenarios to keep elementary streams like audio and video separated in independent UDP/IP flows. Thus, audio and video are not multiplexed together.

It is assumed here that all media flows within one group can be treated with the same QoS class. Thus, audio is equally important as video. All the control data flows are also treated with equal priority.

For Group 1, the application traffic can be identified by a (wildcarded) 5-tuple of packet headers comprising:

- Layer 3 parameters:

- *UE IP:* Any (wildcard).

- *Server IP:* IP address of media gateway or vision/sound mixer.

- *Transport Protocol:* Indicating that UDP is used as the Layer 4 protocol.

- Layer 4 Parameters:

- *UE UDP Port:* Any.

- *Server UDP Port:* Separate UDP ports for audio and video on the Media Gateway or Vision Mixer side.

For Group 3, the application traffic can be identified by a different (wildcarded) 5-tuple comprising:

- Layer 3 parameters

- *UE IP:* Any (wildcard).

- *Server IP:* IP address of MQTT Broker or WebSocket server.

- *Transport Protocol:* Indicating that TCP is used as the Layer 4 protocol.

- Layer 4 parameters

- *UE TCP Port:* Any.

- *Server TCP Port:* TCP Port of the MQTT Broker or WebSocket server.

In cases where all video and audio elementary streams are treated with the same priority, the elementary streams can be multiplexed onto the same UDP/IP flow, e.g. using a multi-programme MPEG‑2 Transport Stream.

NOTE: When using MPEG‑2 Transport Stream as a Payload Format, all multiplexed elementary streams are treated with the same QoS by the 5G System.

#### 6.3.4.3 Solution Example B: Fine-grained separation with separated media

In this example, a finer-grained separation of media is used:

- Within Group 1, the audio elementary stream has a higher priority than the video elementary stream.

- Talkback (Group 2) audio has a lower priority than Group 1 traffic.

- In Group 3, tally light control has a higher priority than general camera control.

As result, the individual media flows should be separated into separate application flows, e.g. UDP/IP flows or TCP/IP flows.

In order to enable the 5G System to prioritise the audio elementary stream higher than the video elementary stream in Group 1, the elementary streams need to be carried as individual UDP/IP media flows.

- RIST Simple profile (see clause 4.2.4) allows usage of separated RTP sessions for different elementary streams, when a native RTP payload format (like RFC 7798 [47] for HEVC or RFC 6416 [49] for AAC) is used.

- RIST Main profile (see clause 4.2.4) uses GRE tunnelling to multiplex all media flows in order to simplify NAT/firewall traversal.

- Alternatively, QRT [54] and RTP over QUIC [55] (see clause 4.6.2) can leverage the multiplexing capabilities of QUIC [51] to achieve the same multiplexing effect as GRE does in RIST Main profile, while using standardised QUIC proxies to further simplify NAT/firewall traversal. (These benefits can also be exploited without multiplexing when carrying a single RTP stream.)

NOTE: The usage of a tunnelling protocol prevents the 5G System from differentiating individual media flows multiplexed inside the tunnel, and thus inhibits its ability to apply different network QoS to the flows multiplexed inside a tunnel.

The talkback audio flow needs to be separated from the main output using dedicated TCP/IP or UDP/IP transmission resources.

If tally light control requires a higher priority than other camera control messages, the event messages should be carried using uniquely identifiable network resources. When MQTT is used for carrying control event messages, the camera needs to set up two MQTT/TCP connections, which can then be clearly prioritized by the 5G System. When WebSockets are used for carrying the event message, the camera should set up two WebSocket/TCP connections to enable separate message prioritization.

### 6.3.5 Solutions leveraging Network Slices

#### 6.3.5.1 General

Network Slicing is a feature which allows a Mobile Network Operator to provide customized networks. Network resources are logically separated so that they can be individually controlled. Each Network Slice contains at least one PDU Session. PDU Sessions cannot be shared across multiple Network Slices.

The UE obtains one IP address for each established PDU Session (Type IP). When a UE establishes multiple PDU sessions, either within a single Network Slice or in different Network Slices, the UE obtains a corresponding number of IP addresses.

#### 6.3.5.2 Solution Example C: Separation using Multiple Network Slices

In this example, the traffic separation is realized using multiple Network Slices. Here, similar to Example A, a coarse-grained separation is assumed: all application flows belonging to Group 1 are carried by Network Slice #1, Group 2 uses Network Slice #2 and Group 3 uses Network Slice #3.

In general, a Network Slice may contain one or more PDU Sessions. For this example, however, it is assumed that each Network Slice contains only a single PDU Session.

The UE obtains an IP address for each PDU Session. The camera then sends all Group 1 traffic with the IP address for PDU Session in Network Slice #1. All Group #2 application flows are sent with the IP address associated with the PDU Session in Network Slice #2, and all Group 3 traffic has the IP address of the PDU Session in Network Slice #3.

The 5G System then handles the traffic according to the Network Slice priority.

#### 6.3.5.3 Solution Example D: Separation using Network Slices and QoS

In this example, traffic separation is realized by combining Network Slices with QoS. Here, a more fine-grained separation is assumed, as in Example B:

- Within Group 1, the audio elementary stream has a higher priority than the video elementary stream.

- Group 2 talkback audio has a lower priority than Group 1 traffic.

- In Group 3, tally light control has a higher priority than general camera control.

In this example, all talkback related traffic uses a dedicated Network Slice for talkback. Meanwhile, all Group 1 camera traffic, all Group 3 traffic, and the return video from Group 2 share a second Network Slice.

As in Example C, each Network Slice is configured with a single PDU Session. The camera is therefore assigned a different IP address for the PDU Session in each Network Slices.

The camera uses the IP address associated with the talkback Network Slice for all talkback audio flows. All other application flows use the IP address associated with other Network Slice. Fine-grained prioritization using QoS is then applied for application flows within the second Network Slice.

### 6.3.6 Summary

This section describes different solutions for traffic segregation in order to prioritize different application flows according to the needs of a media production. In principle, all the different production use-cases (see clause 6.1) require traffic prioritization in some shape or form. Depending on the collaboration scenario, the network can be provisioned (e.g. for QoS or Network Slicing) based traffic prioritization.

Generally, the traffic prioritization configuration can be provisioned statically (e.g. using Operation and Maintenance interfaces) or dynamically (using control APIs). For dynamic provisioning, the 5G System offers a set of APIs which can be used based on the collaboration.

In local production scenarios, the target is to operate at low latency, providing a constant media quality. No additional latency is accounted for in IP-level retransmissions or similar. It is recommended to use QoS flows with a Guaranteed Bit Rate (GBR) for media flows (e.g. program video and audio). Other application flows can use QoS without a bit rate guarantee (i.e., non-GBR).

Remote production scenarios still target low latency. However, certain application-level adaptation schemes may be involved.

For contribution scenarios, the target is to provide a constant quality during a production event. The actual quality can vary from one production event to another.

## 6.4 Key Issue #3: Device onboarding, Remote device configuration and remote device control

### 6.4.1 General

#### 6.4.1.1 Introduction

One general issue is the onboarding of devices into an NPN and the initial device configuration. Temporary devices such as rental devices are often used in media production events. Other media production events even happen at Outside Broadcast locations where a temporary production network is set up before the event and torn down afterwards.

The Media Production vertical has very use-case specific device usage. Devices in a studio evnvironment are often statically assigned to a network; devices deployed as part of an Outside Broadcast (OB) or a Live Touring event are often configured specifically for each event. Rented devices are frequently integrated into a media production. These devices should only have access to the NPN during the duration of the event; after the event, their access needs to be revoked.

#### 6.4.1.2 Media production event timeline

A simplified time sequence for a media production event is illustrated in figure 6.4.1.2-1 and shows different activity periods of the devices during a media production event. A *media production event* is defined here as the time period during which the 5G System is used to produce media. A media production event may, for example, be the airing of a breaking news story or the capturing of a sporting event.

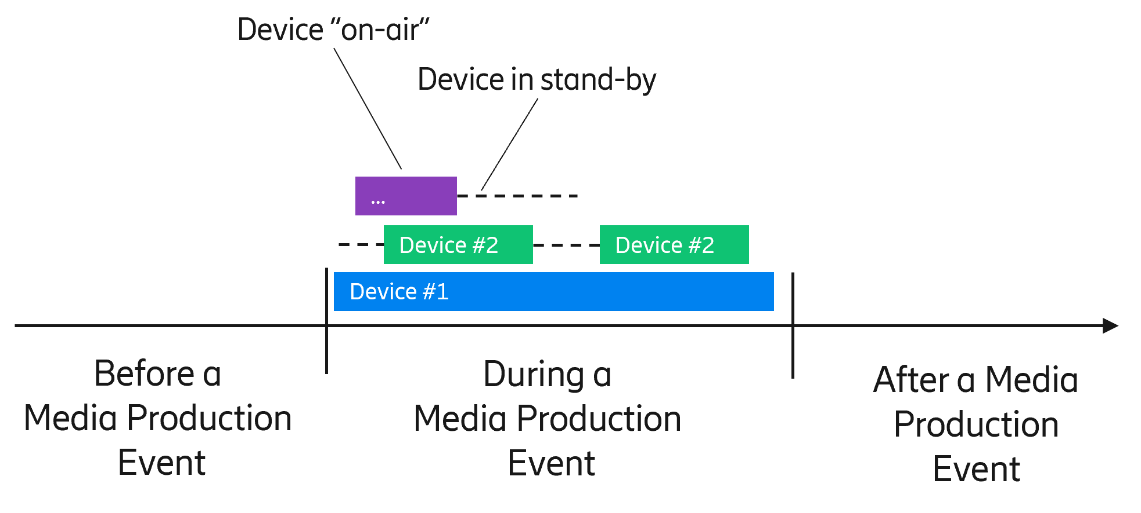


Figure 6.4.1.2-1: Media production event timeline

- Before the media production event, devices such as 5G-enabled professional cameras or displays need to first get access to the 5G System and to receive some initial configuration information. Some devices have conveniently large displays even with touch capability. Other devices like wireless microphones have only small or no displays and only simple buttons.

- During the media production event, devices may not be used continuously.

- At the end of the media production event, devices that were temporarily added to the 5G System need to be "offboarded", i.e. revoking access rights to the NPN.

#### 6.4.1.3 Media production event activities

##### 6.4.1.3.1 Introduction

The following clauses describe the basic activities before (clause 6.4.1.3.2), during (clause 6.4.1.3.3) and after (clause 64.1.3.4) a media production event.

##### 6.4.1.3.2 Onboarding and configuration activities before a media production event

1: The *5G network configuration* activity includes IP address planning and configuration, data network configurations (inlcluding DNN) and provisioning of device authorization information (including access credentials provisioned on a SIM). For permanent installations (e.g. a media production studio), there is a high degree of re-use across multiple media production events; for temporary setups (e.g. an Outside Broadcast), each production event is unique and the configuration is therefore bespoke.

2: The *Device onboarding* activity is the procedure that provides basic 5G network access information to devices, such as the Network Id (e.g. PLMN ID or a NID) and the network access credentials. The device then acts as either an SNPN-enabled UE (a UE configured to use standalone Non-Public Networks with PLMN ID and NID – see clause 5.30.2 of TS 23.501 [84]) or a regular UE with the appropriate PLMN subscription for the PNI-NPN acces through e.g. a Closed Access Group (CAG – see clause 5.30.3 of TS 23.501 [84]) and needs to connect to the correct network to successfully authenticate itself using the network access credentials provisioned above. The network access credentials are typically stored on a SIM card, which is unlocked using a PIN code, although modern devices may support an embedded SIM (eSIM) solution, which removes the dependency on a physical SIM card. For SNPNs, authentication based on EAP-TLS may instead be used.

3: The *Event-specific network configuration* activity realises the requirements placed on the 5G System supporting a specific media production event. Appropriate configurations need to be applied to devices according to the number used and the respective application. This activity may also include the configuration of different QoS flows and other network services.

4: The *Device (remote) configuration* activity is the procedure to provide event-specific configuration information to the device e.g, depending on the Radio Access Network used for the production event (e.g. mid band or high band), and the uplink bit rate consequently available, the device may use a different codec configuration. Further, the IP senders and receivers may change from one event to another and the device configuration may need to be adjusted accordingly. See Key Issue #3 and Key Issue #8.

Editor’s Note: Media Network Configuration and Flow Configuration.

##### 6.4.1.3.3 Activities during an ongoing media production event:

5: The *Device configuration changes* activity involves dynamic configuration changes, e.g. to change the encoding profile of a camera. See Key Issue #4.

6: The *Device control changes* activity covers device-specific control operations, such as camera iris control or pan/tilt/zoom (PTZ) camera control.

7: The *Network monitoring* activity is the monitoring of devices deployed in the NPN and potential actions due to connectivity changes.

##### 6.4.1.3.4 Activities after a media production event

Editor’s Note: For temporary devices like rental devices, the network access credentials should be revoked. When using a physical UICC, this can be done by unplugging the SIM card. In case of eSIM or other authorization mechanisms, the Network Access Credentials should be disabled in the AUSF / AAA-S.

### 6.4.2 Solutions

Editor’s Note: List of solutions:

### 6.4.3 Discussion

## 6.5 Key Issue #4: Different bit rates for Standby vs Program Cameras

### 6.5.1 General

In professional multi-camera production use-cases, several cameras simultaneously capture an event, with each camera focusing on different parts of the scene, for example close-up shots of certain performers, wide angle shot, audience reaction shot. The video of all cameras is displayed in the local studio gallery(which could be in an Outside Broadcast van). The program director selects which of the cameras contribute to the outgoing program feed. On request from the program director, the video mixer changes the source from from one camera to another.

- The camera currently used for the program feed is referred to as the Program Camera in the following. In case of fast cutting between cameras, typically only a single camera is the Program Camera at any given point in time. However, when cross-fading effects are applied during the transition from one camera to the next, two or more cameras are contributing to the program feed at the same time. The tally light is typically illuminated whenever a camera is the active Program Camera.

- All other cameras, which are not directly contributing to the program feed, are just rendered on the display gallery, and are referred to as Standby Cameras.

In principle, the video quality from Standby Cameras does not need to be as high as that from Program Cameras. This key issue explores how the output quality of a video camera can be varied under the control of the video mixer function using a dedicated control flow.



Figure 6.5.1-1: Multi-camera setup showing current Program Cameras and Standby Cameras

The setup is depicted in figure 6.5.1-1. Five Cameras UEs are depicted on the left side. The different Camera UEs are assumed to offer equivalent functionality. However, Camera UE #1 and Camera UE #4 are currently configured to operate in a program quality profile, while Camera UEs #2. #3 and #5 are operating in a stand-by quality profile. The Mixer/Selection function is controlled based on the input from from the program director.

### 6.5.2 Solutions leveraging QoS

#### 6.5.2.1 General

In this set of solutions, the 3GPP QoS Framework is used to separate the different data flows. Different resource types (like GBR) may be used for the QoS flows. Figure 6.5.2-1 illustrates the usage of QoS flows in the multi-camera setup introduced by figure 6.5.1‑1.



Figure 6.5.2-1:

In the figure:

- The video outputs of Camera UE #1 and Camera UE #3 are mapped to program quality QoS flows.

- All other camera outputs are mapped to stand-by quality QoS flows.

- The traffic for switching the camera profile configuration is handled by yet another QoS flow.

Two basic solution alternatives for realising the above QoS flow configuration are described in the following clauses:

A: Modify the QoS rule within a Camera UE whenever the quality profile is switched between program and stand-by.

B: Pre-install a unique QoS rule for each camera quality profile with different packet filters.

#### 6.5.2.2 Solution A: QoS rule modification

Solution alternative A leverages the procedure to modify a QoS rule. When switching a camera quality profile from program to standby, the QoS rule is modified and a new QoS policy (with a different QFI) is installed on the UE and the UPF.



Figure 6.5.2-2: Solution Alternative A

Before the actual media production event:

1: The QoS rules are configured in the 5G System using a 5-tuple as packet filter. The QoS rules are installed in the UE SDAP entity and the UPF.

During the media production event:

2: The program director triggers the mixer function to select the Camera as Program Camera. As result, the Mixer sends an instruction to the Camera to change the Profile to Program.

3: The Camera sends a video stream towards the mixer according to the Program profile.

4: The UE SDAP entity detects a packet filter match and forwards the packets as QoS Flow #1.

5: Video packets traverse the 5G System

6: After some time, the program director selects another camera as Program Camera. As result, the Mixer sends an instruction to the Camera to change the Profile to Standby.

7: At the same time (or before), the Mixer interacts with the PCF (possibly via the NEF) to modify the QoS rule for the traffic flow.

8: The NEF/PCF sends a modified PCC rule to the UE SDAP entity and the UPF. Depending on the parameters for the Standby video, the PCF may select a different 5QI (resulting in a resource type change e.g. from GBR to Non-GBR) and/or change the value for the Guaranteed Bit Rate (GBR).

9: The Camera sends a video stream towards the Mixer according to the Standby profile.

10: The UE SDAP entity detects a packet filter match and forwards the packets as QoS Flow #2.

11: Video packets traverse the 5G System.

#### 6.5.2.3 Solution B: Multiple pre-installed QoS rules

For Solution alternative B, two QoS flows (one for program quality and one for standby quality) are established in advance for each camera. The Packet Filters for the two QoS flows are different, e.g. using different 5-tuples or different ToS/DSCP values.



Figure 6.5.2-3: Solution Alternative B

Before the media production event:

1: Two QoS rules are configured in the 5G System, using different TOS values (TOS#1 and TOS#2) as Packet Filter. The QoS rules are installed in the UE SDAP entity and the UPF.

Depending on the parameters for the Standby video, the PCF may select a different 5QI (resulting in a resource type change e.g. from GBR to Non-GBR) and/or change the value of the Guaranteed Bit Rate (GBR).

During the media production event:

2: The program director triggers the mixer function to select the camera as Program Camera. As result, the Mixer sends an instruction to the Camera to change the Profile to Program.

3: The Camera sends a video stream (marked with TOS#1) towards the Mixer according to the Program profile.

4: The UE SDAP entity detects a packet filter match (TOS field) and forwards the packets as QoS Flow #1.

5: Video packets traverse the 5G System.

6: After some time, the program director selects another camera as Program Camera. As result, the Mixer sends an instruction to the Camera to change the Profile to Standby.

7: The Camera sends a video stream (marked with TOS#2) towards the mixer according to the Standby profile.

8: The UE SDAP entity detects a packet filter match (TOS field) and forwards the packets as QoS Flow #2.

9: Video packets traverse the 5G System.

### 6.5.3 Discussion

#### 6.5.3.1 Discussion of Solution A

Pro:

- Only a single QoS rule installed.

Contra:

- Additional signalling latency from SMF to the UE for modifying the PCC rule for the QoS flow. The signalling from the 5G System to the UE is not frame-synchronized with the camera source.

#### 6.5.3.2 Discussion of Solution B

Pro:

- When the Mixer selects a camera as Program Camera, the Camera changes the QoS flow on application UE side by changing the packet marking. The marking change can be synchronized with the first packets generated by the camera source in the new Profile.

Contra:

- Two QoS rules are installed at the same time. The camera can use the Program Profile QoS flow at any time, just by changing the TOS/DSCP value. Care needs to be taken to ensure that the UE Camera is only selecting TOS values according to the selected profile.

- The network may reserve resources for two QoS flows, while only one QoS flow is ever used at a time.

Editor’s Note: Something more around Cue Pilots, etc. In some production scenarios (like music), everything can be fully timed and scripted in advance, e.g. the ligthing changes. The list of switching cues (the “camera script”) can also be pre-loaded onto the Cameras.

## 6.6 Key Issue #5: Dynamic bit rate adaptation

### 6.6.1 General

Dynamic bit rate adaptation describes the capability to adjust the encoding bit rate of a compressed stream during operation in order to handle short term network variations, by varying the quality of the encoded media stream. Those network variations can be caused e.g. by high load, interference or mobility events. There can be different triggers for rate adaptation, e.g. a control signal from the network or continuous monitoring the network performance (e.g. by estimating the available bandwidth). Such a capability may not be required for Tier 1 AV productions, since Tier 1 AV productions are typically well planned from a capacity and coverage perspective. Dynamic bit rate adaptation could, however, become an important tool for Tier 2 or Tier 3 production scenarios to improve the overall robustness of the system, e.g. to increase the usage flexibilty and simplify SLA negotiations and fulfillment.

This type of adaptive bit rate is not widely available for professional applications so adoption by the media production industry is needed.

- Solutions can describe different realizations (e.g. using the Temporary Maximum Media Bit Rate (TMMBR) RTCP transport layer feedback message defined in RFC 5104 [41] and section 6.2 of RFC 4585 [42], the RTCP-based congestion control feedback message defined in RFC 8888 [72], etc)

- Support can be an optional feature of a media protocol.

- Tradeoff between packet loss, quality, etc (different parameters to fit into the bit rate budget) should be studied

Editor’s Note: More input needed on acceptable performance, potential SLA requirements, bit rate boundaries, such as accaptable minimal bit rate, etc needed from media producer side.

NOTE: Dynamic bit rate adaptation is typically applied to video signals, but can also be applied to audio.

### 6.6.2 Dynamic bit rate adaptation solutions leveraging RTCP-based congestion control signalling

RTP-based media transport protocols such as RIST (see clause 4.2.4) as well as QRT and RTP over QUIC (see clause 4.6.2) can make use of RTCP-based congestion control mechanisms to apply rate control to RTP media flows [44] and to influence the bit rate of a media encoder in real time so that it operates within a capacity envelope dicatated by prevailing network conditions.

A congestion control feedback packet is defined in RFC 8888 [72] for this purpose and it may be combined with a number of different congestion control algorithms, including (but not limited to) Network-Assisted Dynamic Adaptation (NADA), defined in RFC 8698 [73], Self-Clocked Rate Adaptation for Multimedia (SCReAM), defined in RFC 8298 [74], Google Congestion Control [75] and Shared Bottleneck Detection, defined in RFC 8382 [76].

## 6.7 Key Issue #6: Interfacing Audio Channels

### 6.7.1 Circuit-switched audio transmission

The Multiple Audio Digital Interface (MADI) [38] and the Serial Digital Interface (SDI) [35][36] embed audio channels together with video channels onto the same physical medium. Multiple Audio Digital Interface (MADI) [38] supporting [serial digital transmission](https://en.wikipedia.org/wiki/Serial_transmission) over [coaxial cable](https://en.wikipedia.org/wiki/Coaxial_cable) or optical [fibre](https://en.wikipedia.org/wiki/Fibre-optic) lines of 28, 56, 32, or 64 channels; and [sampling rates](https://en.wikipedia.org/wiki/Sampling_rate) to 96 kHz and beyond with an [audio bit depth](https://en.wikipedia.org/wiki/Audio_bit_depth) of up to 24 bits per channel. Where encapsulated audio and video are used then fewer channels are likely to be deployed. As a minimum, this should consist of two audio channels.

5G System resources are shared among devices and radio resources should preferably not be allocated and left idle. This key issue should study, how in particular audio channels are allocated in existing media productions and how 5G based media productions can interwork with existing media productions, when a more dynamic allocation of audio channels is used on 5G Systems.

Audio may be carried as an encapsulated signal multiplexed with video and data, or as a separate set of streams. For tier one or audio-only productions, the audio is treated as separate discrete streams per channel. For tier two and three productions and contribution workflows, it may be desirable to carry audio and video multiplexed with the video.

A channel is usually a mono signal. An audio channel can be considered as:

- *Active* or *inactive:* Not all channels (allocated in MADI or SDI) may be required for all applications so it should be possible to describe a channel as either active or inactive so as to make more efficient use of available bandwidth.

- *Muted* or *unmuted*: An active channel may be temporary muted where it may be required but the UE is not transmitting any data.

- *Silent:* A silent channel is active and unmuted but with a low-level audio signal. This may be used to provide atmospheric or spot effects.

In MADI [38] and SDI [35][36] there is a mode bit to indicate whether a channel is active or not.

The first four bits of a MADI subframe are used for frame synchronization and to indicate the state. If a channel is set to inactive, all payload data must be set to zero. Thus, an individual audio channel cannot be taken out of the multiplex compared to a packet-based transmission approach.

Also, MADI is synchronized in itself: it does not lock to the audio sampling frequency. This indicates an additional need for audio clock synchronicity via a word clock.

In audio production networks a word clock is a dedicated (physical) distributed timing signal. It is the one master clock providing the sampling frequency for all audio processing devices.

Editor’s Note: It should be checked, whether there is a DVB or SMPTE threshold definition for “silence”.

Communication channels are usually speech-only and of a lower quality than main programme audio but do require low-latency solutions. There is also a requirement for one-to-many solutions so that a director can speak to multiple end users at the same time.

SDI (Serial Digital Interface) [35][36] is a family of standards widely used in the media production domain to transport uncompressed video signals. Various SDI interface (SD-SDI, HD-SDI, 3G-SDI, 6G-SDI, 12G-SDI and 24G-SDI) are available to support from standard definition up to UHD video resolutions.

SDI can carry also embedded audio.

3G-SDI, known as the 3Gbit/s interface, defined different mapping levels (A, B-DL, B-DS) for the carriage of 1080-line image formats and associated ancillary data. With respect to the audio, 3G-SDI may contain up to 16 audio channels or 32 if dual-link applications are considered or SMPTE ST 299-2 is used.

NOTE: 3G-SDI and later supports 32 channels but in practice it is limited to 16 channels as it is rare to find products that support more than 16 channels. In fact many products only support 8 channels. For transporting higher than 16 channels of Higher Order Ambisonics (HOA) over SDI, refer to the HOA Transport Format (HTF) as defined in ETSI TS 103 589 [81]. For transporting more than 16 audio channels of channel-based audio content in live production cases over 16 channel SDI, refer to Annex 1 of ITU-R BS.1548-7 [83].

Each audio channel in SDI (and in MADI) is AES3-encoded, removing the need for data conversion between both interfaces. Since the audio is data embedded in the ancillary data, the synchronisation is relative to the video signal. The timing information as well as additional metadata (i.e. sample rate, bit depth, channel status) are transmitted in additional audio control packets.

In Tier one scenarios, in general, the audio signals come from the microphones installed in the studio/location (and not from the cameras) while in Tier two and Tier three productions, especially for contribution links, embedded audio is transmitted multiplexed with the video. When the audio is embedded, MPEG-2 Transport Stream might be used over RTP/UDP/IP instead of native RTP carriage. For ST 2110-30 scenarios, six conformance levels are defined [40]. Level A is the only mandatory conformance level to be supported by all compliant equipment and is defined as follows:

- Linear 24-bit PCM encoding.

- 48 kHz sampling frequency (media clock).

- 1 to 8 channels per stream.

- 1 ms packet time (48 audio samples per channel in each packet).

### 6.7.2 Packet-based audio channels

#### 6.7.2.1 Introduction

Audio protocols have evolved towards packet-based technologies to be compliant with standard technologies and infrastructure such as Ethernet and IP. Various protocols have been developed in the industry. Depending on the operating layer one can distinguish between *audio-over-Ethernet* at Layer 2 of the protocol stack (clause 6.7.2.2) and *audio-over-IP* at Layer 3 (clause 6.7.2.3).

#### 6.7.2.2 Audio-over-Ethernet

Layer-2 systems with a proprietary data link protocol. Sometimes compliant with Ethernet framing to utilize existing infrastructure and components, thereby reducing cost. Operated within different topologies, mostly star or tree.

Examples: CobraNet, EtherSound, Soundgrid.

#### 6.7.2.3 Audio-over-IP

Layer-3 protocols based on IP. Often using derivates of UDP at Layer 4 with additional supporting proprietary protocols for solving device discovery, synchronisation, management and configuration, etc.

Examples: DANTE (Digital Audio Networking Through Ethernet) [82], WheatNet, Livewire, Q-LAN, Ravenna.

While DANTE is the most prominent example, native synchronisation depends on underlying PTP-v1, while PTP-v2 support was added later [82]. Payload definition varies per protocol, and so different solution are thus not compatible with each other. Routing and network pathing via the IP layer enables coexistence with other IP-based services, but may depend on managed infrastructure to guarantee QoS.

Interoperability was later introduced via standardization of AES67 [57] as the common denominator of all packet-based audio transport technologies. AES67 defines a profile that all audio-over-IP solutions must provide to be compliant:

- Synchronization via PTPv2, as specified in IEEE 1588-2008 [22].

- 48 kHz audio sampling rate.

- 16- or 24-bit linear PCM encoding ("L16/L24").

- 1 to 8 channels per stream.

- 48 samples per packet.

- IP/UDP/RTP header with unicast and multicast support.

- DiffServ for network QoS using ToS labelling in the IP packet header.

- SDP for announcing sessions.

AES67 is thus compliant with "level A" in ST 2110-30 [30].

### 6.7.3 Solutions

### 6.7.4 Discussion

## 6.8 Key Issue #7: Usage of NPN (SNPN or PNI-NPN)

Starting in Release 16, 3GPP defines the concept of a Non-Public Networks (NPN) to refer to a 5G System (5GS) deployed for private use (e.g. a business-to-business network deployment) and designed to support requirements and services for such scenarios. This may be done by deploying specific features involving physical and/or virtual infrastructure and network services.

The requirements to enable NPNs for video, imaging and audio for professional applications are described in 3GPP TS 22.261 under the following clauses:

- Generic NPN requirements can be found in clause 6.25.

- Requirements on the subscription aspects can be found in clause 6.14.

- Authentication requirements can be found in clause 8.3.

3GPP is addressing such requirements and capabilities for the support of NPNs under different work items involving functional (SA2) and management (SA5) aspects.

3GPP classifies NPNs into two principal categories:

**-** *Standalone NPN (SNPN)* is an NPN whose deployment neither relies on network functions nor on network services provided by a PLMN. The SNPN is operated by an NPN operator which could be the media company itself or a contracted third party. The NPN operator has the capabilities to manage and control the network functions provided by the SNPN.

On the network side, the SNPN is identified by combination of a PLMN ID and Network identifier (NID). At the UE, these two parameters need to be configured to access the SNPN. The PLMN ID may be one assigned in the range of PLMN IDs for private networks (e.g. based on MCC 999, as assigned by the ITU). The PLMN ID of a PLMN that is operating the SNPN may also be reused. The NID could be self-assigned by an individual SNPN or assigned in coordination with other NPN operators.

Note that a UE connected to an SNPN may also be able to access services from a PLMN. In such case, the UE is required to authenticate in both networks. Release 16 specifications do not include support for roaming, handover between SNPNs not interworking with Evolved Packet Core (EPC). Emergency services are not supported in SNPNs.

Editor’s Note: What if the NPN operator uses DNNs or Network Slicing (i.e. PNI-NPN technologies) to offer network services to media producers?

**-** *Public Network Integrated NPN (PNI-NPN)* is an NPN deployed with the support of at least one PLMN. This model may involve a contract between the NPN user (e.g. media company) and the PLMN providing the network resources (including radio access and core network) to support the media company requirements. Two deployment solutions are normative:

- *PNI-NPN deployment by means of dedicated Data Network Names (DNNs).* The DNN defines a dedicated gateway (UPF) in the PLMN to/from which NPN traffic is conveyed and dispatched to the NPN local area network.

*- PNI-NPN deployment by means of network slicing.* The PLMN provisions a dedicated slice of the PLMN comprising a set of resources allocated for the exclusive use of the NPN. Such a network slice may define specific network functions or features to be used for the NPN including, for instance, UE onboarding and authentication, Time Sensitive Networking (TSN) [39] integration, etc, i.e. features can typically always be provided by an SNPN.

For both of these deployment models, the PLMN ID is used to access the PNI-NPN. Therefore, UEs must already have a subscription to a PLMN. In order to control the service area of the NPN, a list of subscribers who are allowed ~~to~~ access the cells associated with the PNI-NPN can optionally be provided by means of a Closed Access Group (CAG). When PNI-NPN is provisioned by network slicing, a UE may be preconfigured with Single Network Slice Selection Assistance Information (S-NSSAI) to access certain slices.

The NPN architecture has been enhanced in Release 17, including for instance:

- Enable support for SNPN along with subscription/credentials owned by an entity separate from the SNPN operator.

- Support UE onboarding and provisioning for NPNs.

- Support audio–visual content production service requirements, e.g. for service continuity

- Support voice/IMS emergency services for SNPN.

Depending on the considered application, the NPN can also be enriched with other complementary functionalities, including Wi-Fi access and TSN technologies.

## 6.9 Key Issue #8: Usage of mmWAVE for low-latency television production

As introduced in clause 5.5, a lightweight video compression codec is typically used to achieve low latency in television production and is currently transmitted via IP or Ethernet cabling. 5G mmWAVE is capable of providing the sufficient large data rate, but the wireless channel fluctuation may challenge service quality. In particular, mmWAVE spectrum is prone and sensitive to propagation and penetration loss, as well as to blockage by objects and human bodies. To better utilize the wireless channel reflection and scattering, 3GPP defines advanced beam management as well as multi-carrier operation [59].

Nevertheless, to operate production scenarios at several hundred megabits per second over a mmWAVE radio link, content delivery protocols and codecs are needed that can compensate the mmWAVE channel characteristics, support high reliability and achieve the high bit rate as shown in Figure 6.9-1.

A picture containing timeline

Description automatically generated

Figure 6.9-1: Protocol architecture – Mobile telestudio

At least the following techniques should be considered for mmWAVE based content delivery protocols:

1. *Adaptive bit rate encoding.* Dynamically adjusting the encoder bit rate allows the encoded picture quality to be increased or decreased according to variation in the performance of the wireless channel. (See clause 6.6 for more details.)

2. *Packet reordering.* Wireless channels change rapidly and the mobile system usual employs Hybrid Automated Repeat Query (HARQ) at the physical layer to retransmit data blocks that fail to arrive at the receiver. The packets sent earlier may arrive after the packets sent later due to different retransmission times. This might lead to wrong orders of packets and further errors at the decoder. The receiver could buffer the received packets and reorder them in case of missing original packets, although doing so incurs a latency penalty.

3. *Error correction.* AL-FEC, ARQ, or Frame repair are usually used to compensate for reception errors.

- *Application Layer Forward Error/Erasure Correction (AL‑FEC)* techniques transmit additional redundant coded bits or packets in order to make the bit stream or packet stream more robust. Extra channel bandwidth is required to accommodate the AL‑FEC overhead and additional memory is required at the sender and at the receiver to process the AL‑FEC information. AL‑FEC also includes a variable time penalty, which may be undesirable in low-latency production.

- *Automatic Repeat Query (ARQ)* involves retransmitting data that was not received at the request of the receiver, signalled by means of explicit acknowledgement packets back to the sender. This retransmission introduces a variable delay which is undesirable for low-latency production.

- The *Frame repair* technique is an error concealment technique and compensates for undecodable video frames that result from lost packets by interpolating between neighbouring successfully decoded frames. If the receiver interpolates only from (successfully received) past frames, this neither requires additional bandwidth, nor does it add any latency. If future frames are used, the interpolation process does incur some additional latency.

Among these three error correction techniques, if the errors happen occasionally, Frame repair can maintain a similar bit rate and latency performance as Ethernet.

# 7 Summary and Conclusions

# Annex <X> (informative): Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change history** | | | | | | | |
| **Date** | **Meeting** | **Tdoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
| Apr 2021 | SA4#113 | S4-210519 |  |  |  | Initial version | 0.0.1 |
| Apr 2021 | SA4#113 | S4-210678 |  |  |  | S4-210527: Structure of the technical report  S4-210641: Description of existing media protocols in media production | 0.1.0 |
| May 2021 | Post SA4#113 | S4-210726 |  |  |  | S4aI211164: Description of camera media flows in a Multi-Camera production  S4aI211165: Overview of NMOS functionality | 0.1.1 |
| May 2021 | SA4#114 | S4-210939 |  |  |  | S4-210919: FS\_NPN4AVProd: Utilizing Available Capacity in Multi-Camera Scenarios  S4-210913: Addition of different production types and addition of more information about existing workflows. | 0.2.0 |
| Aug 2021 | SA4#115 | S4-211267 |  |  |  | S4-211241: [FS\_NPN5AVProd] Clarification of Cloud vs Remote Production  S4-211242: [FS\_NPN5AVProd] Proposal of Media Protocol related Key Issues  S4-211243: [FS\_NPN5AVProd] Proposal of a Remote Camera Configuration Key Issue  S4-211244: [FS\_NPN5AVProd] Proposal of two bit rate adaptation related Key Issues  S4-211245: [FS\_NPN5AVProd] Proposal of a Key Issue around configurable audio channels  S4-211246: [FS\_NPN5AVProd] Proposal of a new NPN usage related Key Issue | 0.3.0 |
| Nov 2021 | SA4#116 | S4-211600 |  |  |  | S4aI211249: [FS\_NPN4AVProd] Update of SRT and RIST description  S4-211601: [FS\_NPN4AVProd] QoS Separation | 0.4.0 |
| 2021-12 | SP#94-e | SP-211341 |  |  |  | Presentation to SA plenary for information | 1.0.0 |
| 2021-12 | Post SA4#116 | S4aI211275 |  |  |  | S4aI211265: Structure update for TR 26.805 | 1.0.1 |
| 2022-02 | Post SA4#116 | S4-220136 |  |  |  | Editorial Corrections  S4aI221294: [FS\_NPN4AVProd] SMPTE audio metadata | 1.0.2 |
| 2022-02 | Post SA4#117e | S4-220279 |  |  |  | S4-220029: [FS\_NPN4AVProd] mmWAVE for Media Production  S4-220059: Tunnelling RTP media sessions over QUIC  S4-220142: [FS\_NPN4AVProd]: Definition of Collaboration Scenarios  S4-220143: [FS\_NPN4AVProd]: Introduction to Candidate Solutions and updates to KI#2 | 1.1.0 |
| 2022-04 | SA4#118e | S4-220548 |  |  |  | S4-220467: Description of KI#4 (Standby and Program Cameras), incl solutions  S4-220561: KI6 Interfacing Audio Channels  S4-220562: Device On Boarding | 1.2.0 |
| 2022-05 | SA4#119e | S4-220687 |  |  |  | Editorial Corrections | 1.2.1 |