|  |
| --- |
| 3GPP TR 26.805 V0.3.0 (2021-08) |
| 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 address3GPP support office address650 Route des Lucioles – Sophia AntipolisValbonne – FRANCETel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16Internet<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 members3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational PartnersLTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational PartnersGSM® 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 9

3.1 Terms 9

3.2 Symbols 9

3.3 Abbreviations 9

4 Review of existing workflows and media protocols 10

4.1 General 10

4.2 Transport Protocols 11

4.2.1 General 11

4.2.2 SMPTE ST 2110 11

4.2.2.1 Introduction 11

4.2.2.2 ST 2110 for audio (ST 2110-30 and ST 2110-31) 12

4.2.2.3 ST 2110 for video (ST 2110-20 and ST-2110-22) 12

4.2.3 Secure Reliable Transport (SRT) 13

4.2.4 Reliable Internet Stream Transport (RIST) 13

4.2.5 Network Device Interface NDI 14

4.2.6 IP Media eXperience (IPMX) 14

4.2.7 Comparison Table 15

4.2.8 Other Protocols 15

4.2.9 Audio Networking Solutions 15

4.3 Codec choice 16

4.5 Review of existing orchestration and control solutions 18

4.5.1 General 18

4.5.2 AMWA Network Media Open Specification (NMOS) 18

4.5.3 Camera control and configuration protocols 20

4.5.3.1 General 20

4.5.3.2 Camera control protocols 20

4.5.4 EMBER+ 21

4.5.5 Other Protocols 21

6 Relevant media production use cases 21

6.1 General 21

6.2 Use-Case X: Audio Visual production 21

6.2.1 Description 21

6.2.2 Wireless camera workflows 22

6.2.2.1 Scenario 1: Wireless cameras within a production workflow 22

6.2.2.2 Scenario 2: Outside broadcast contribution 23

6.2.2.3 Considerations on cloud-based production 24

6.2.2 Collaboration models and deployment architectures 24

6.2.3 Identified 5G System features 24

6.2.4 High level call flows 24

6.2.5 Potential issues 25

6.2.5.1 General 25

6.2.5.2 Utilizing Available Capacity in Multi-Camera Scenarios 25

6.2.5.2.1 QoS requirements – bit rate 25

[6.x Use-Case X 26

6.x.1 Description 26

6.x.2 Collaboration models and deployment architectures 26

6.x.3 Identified 5G System features 26

6.x.4 High level call flows 26

6.x.5 Potential issues 27

7 Candidate Solutions 27

7.1 Issue #1: Utilizing Available Capacity in Multi-Camera Scenarios 27

7.1.1 General 27

7.1.2 Potential solutions 27

8 Summary and Conclusions 27

Annex <X> (informative): Change history 28

# 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 TS 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] VSF: "Reliable Internet Stream Transport (RIST) Activity Group",

[7] VSF TR 06-1: "Reliable Internet Stream Transport (RIST) Protocol Specification – Simple Profile", <https://vsf.tv/download/technical_recommendations/VSF_TR-06-1_2018_10_17.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)".

# 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

Void.

## 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].

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

IPMX IP Media eXperience

mDNS Multicast DNS

NMOS Networked Media Open Specifications

NPN Non-Public Network

PA Public Address

PTP Precision Time Protocol

PTZ Pan, Tilt, Zoom

RIST Reliable Internet Stream Transport

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 has the benefits of being able to use commercial off-the-shelf technologies where previously there have been highly specialised solutions.

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 and often low latency in the network as well as challenging 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 the receiving end.

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

1. *Production:* All of the activity that happens locally on location. This activities often involves multiple sources of audio and video content as well as use alongside of parallel 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 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 evolved. Certain protocols and codecs are better suited to the different tiers of production. Alongside these, different solutions have emerged: some proprietary solutions that meet certain aspects of the workflows and some more open and interoperable. As well as the media transport layer there are also requirements around control and orchestration of specific technologies and again some of these are built on open infrastructures and some 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. ARQ or FEC), different security features, support for packet pacing and/or traffic shaping, and features to allow network address translators (NAT) and 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 [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 [X], 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 [W] 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 (IEC61883-6) 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 the Serial Digital Interface (SDI) [35] [36], 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 High Dynamic Range (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

### 4.2.3 Secure Reliable Transport (SRT)

Secure Reliable Transport (SRT) [5] is an open-source media transport protocol that uses the UDP transport protocol. 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).

### 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 published by the VSF in October 2018 and 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 in RFC 4585.

- A Range-based NACK, defined as an application-specific (APP) RTCP packet.

- Bonding of multiple links for load sharing.

- Seamless switching using SMTPE-2022-7.

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

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, 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.

### 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 raterequired 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) 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 lowcompressed 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 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 |
| 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 4.2-2:

- 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.

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

# 5 Relevant media production use cases

## 5.1 General

## 5.2 Use-Case X: Audio Visual production

### 5.2.1 Description

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 that is consuming that content both live and recorded for later distribution. 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 set ups require multiple devices such as cameras, microphones and control surfaces that require extremely close synchronisation to maintain consistency of pictures 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.

### 5.2.2 Wireless camera workflows

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

Different types of network may be deployed depending on how the camera is used. For a single point-to-point (PTP) 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 mesh network with multiple receivers may be set up. This allows the cameras to move freely within the coverage area while maintaining 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 matched against other cameras that are connected directly to the production network by fibre or coax connections. It is important that in this scenario the latency of any radio-connected device is minimised and 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 times 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.

The types of camera 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 5.2.2.1-1: Flows by one camera unit

Figure 5.2.2.1-1 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 else 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 from 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 like the shutter speed, iris, etc 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, Intercom might become more and more independent devices in media production, since intercom typically is setup 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.

#### 5.2.2.2 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 use of 4G networks can bring 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 alternate 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 tend provide a single video link and so if more than one camera is required it either needs multiple units that are often timed differently or people and infrastructure on site to support multiple camera operation. 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 struggle 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.

#### 5.2.2.3 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 matrixes, 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.

### 5.2.2 Collaboration models and deployment architectures

Editor’s Note: No input yet.

<Should we add a Remote Production use-deployment, with an SNPN on-prem and then remote functions?>

### 5.2.3 Potential issues

#### 5.2.3.1 General

#### 5.2.3.2 Utilizing Available Capacity in Multi-Camera Scenarios

##### 5.2.3.2.1 QoS requirements – bit rate

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 5.2.3.2-1: 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 UL10-7 DL | 12 Gbit/s | 20 Mbit/s |
| Uncompressed HD video | 1 | 0 km/h | 1 km2 | 400 ms | 10-9 UL10-7 DL | 3 .2 Gbit/s | 20 Mbit/s |
| Mezzanine compression UHD video | 5 | 0 km/h | 1000 m2 | 1 s | 10-9 UL10-7 DL | 3 Gbit/s | 20 Mbit/s |
| Mezzanine compression HD video | 5 | 0 km/h | 1000 m2 | 1 s | 10-9 UL10-7 DL | 1 Gbit/s | 20 Mbit/s |
| Tier one events UHD | 5 | 0 km/h | 1000 m2 | 1 s | 10-9 UL10-7 DL | 500 Mbit/s | 20 Mbit/s |
| Tier one events HD | 5 | 0 km/h | 1000 m2 | 1 s | 10-8 UL10-7 DL | 200 Mbit/s | 20 Mbit/s |
| Tier two events UHD | 5 | 7 km/h | 1000 m2 | 1 s | 10-8 UL10-7 DL | 100 Mbit/s | 20 Mbit/s |
| Tier two events HD | 5 | 7 km/h | 1000 m2 | 1 s | 10-8 UL10-7 DL | 80 Mbit/s | 20 Mbit/s |
| Tier three events UHD (Note 2) | 5 | 200 km/h | 1000 m2 | 1 s | 10-7 UL10-7 DL | 20 Mbit/s | 10 Mbit/s |
| Tier three events HD (Note 2) | 5 | 200 km/h | 1000 m2 | 1 s | 10-7 UL10-7 DL | 10 Mbit/s | 10 Mbit/s |
| Remote OB | 5 | 7 km/h | 1000 m2 | 6 ms | 10-8 UL10-7 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) $=\frac{10^{-5}}{3x}$, where x $x$ is the data rate in Mbps.NOTE 2: Could use either professional equipment or mobile phone equipped with dedicated newsgathering app  |

Further, Table 5.2.3.2‑1 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.

**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 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.

#### 5.2.3.3 Key Issue #2: Media Protocols on 5G: Using QoS for traffic segregation

##### 5.2.3.3.1 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 5.2.2.1-1) – 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.

Editor’s Note: Solutions should consider multiple combinations of application flows. Input is needed on the prioritization between application flows, e.g. when audio is present with the program video. It will be studied how audio or video is dropped in contentious situations.

Editor’s Note: Evaluation of this Key Issue can allow protocol consideration and recommendations on network usage, e.g. flow separation, etc.

##### 5.2.3.3.2 Usage of RIST Simple Profile

Editor’s Note: This section aims to describe the usage of RIST Simple profile [7] features on 5G (NPN) Systems. Here, the various flows (uplink and downlink) should be separated & prioritized using 3GPP QoS framework. (Media and Non-Media like RC & telematics)

##### 5.2.3.3.3 Usage of RIST Main Profile

Editor’s Note: Same as previous subclause, but with RIST Main Profile [8] feature.

##### 5.2.3.3.4 Usage of SRT

Editor’s Note: Same as previous subclause, but with SRT [5] features.

##### 5.2.3.3.5 Summary

#### 5.2.3.4 Key Issue #3: Media Protocols on 5G: Using Network Slices or Multiple PDU Sessions for traffic segregation

##### 5.2.3.4.1 General

This clause focuses in the same set of issues (i.e. media protocol usage) as described in clause 5.2.3.3, with the difference of using Network Slices or multiple PDU Sessions for traffic separation. Traffic separation is needed when multiple concurrent productions are sharing the same NPN facility (SNPN or PNI-NPN) at an outside broadcast venue. This Key Issue aims to examine the pros and cons of network slices versus different PDU sessions for supporting such Use Cases. It is assumed that each PDU session contains only a single QoS flow with a default QoS PCC rule.

Example realizations:

- Program Video and Audio are carried by a separate Network Slice or PDU Session from other Media Production traffic, i.e. audio and video in the same Network Slice or PDU Session.

- Return Video is carried is carried by a separate Network Slice or PDU Session from Program media and other media.

##### 5.2.3.4.2 Usage of RIST Simple Profile

##### 5.2.3.4.3 Usage of RIST Main Profile

##### 5.2.3.4.4 Usage of SRT

##### 5.2.3.4.5 Summary

#### 5.2.3.5 Key Issue #4: Remote camera configuration and remote control

Editor’s Note: This clause should study the needs for (remote) camera configuration and camera control. It is not the intend to promote the definition of a new application, instead, the (remote) camera configuration and camera control application aspects are defined by other organizations like NMOS. Camera configuration refers to procedures and parameters to configure a camera e.g. encoders and/or decoders and media protocols (IP addresses, ports, transport protocol, etc). Camera Control refers to procedures to change setting during capturing, e.g. pan–tilt–zoom, iris, etc.

Editor’s Note: Existing NMOS standard extensively uses the HTTP REST model. For camera configuration (as example device), IS-05 requires that the camera exposes HTTP REST APIs and hosts an HTTP server. For camera control using IS-07, the camera can either expose an HTTP REST API or receive the messages via WebSockets or MQTT,

Outcome: Recommendations on 5G System features, which are beneficial for (remote) camera configuration protocol options and features.

#### 5.2.3.6 Key Issue #5: Different bit rates for Standby vs Program Cameras

Editor’s Note: This clause should describe implications on protocol usage, when only the program camera(s) send a high quality stream. Standby cameras only send a video stream with preview quality or no data.

#### 5.2.3.7 Key Issue #6: Dynamic bit rate adaptation

##### 5.2.3.7.1 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 bitrate 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], etc.).

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

- Trade-off 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, bitrate boundaries, such as accaptable minimal bitrate, etc needed from media producer side.

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

#### 5.2.3.8 Key Issue #7: Configurable Audio Channels

Editor’s Note: This clause should describe implications on protocol usage, when a predefined number of audio channels (as in MADI or SDI) is allocated, independently on its needs. In SDI, always 32 audio channels are allocated. Unused audio channels are “muted”. See ST 299 for more details. (https://tech.ebu.ch/docs/techreports/tr002.pdf)

* 1. Are muted audio channels used for other purposes in SDI / MADI, which should be considered for 5G deployments?
	2. Is it needed to send audio frames with “many null payload bytes“? What is the practice in ST 2110, which also supports separated A & V?
	3. Would all audio channel perceive same quality/QoS? Or can some audio channels require low latency while other audio channels are “embedded with video”?

Editor’s Note: This clause should describe the possibility of configuring audio channels on a need basis.

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%22%20%5Co%20%22Serial%20transmission) over [coaxial cable](https://en.wikipedia.org/wiki/Coaxial_cable%22%20%5Co%20%22Audio%20bit%20depth) or optical [fibre](https://en.wikipedia.org/wiki/Fibre-optic%22%20%5Co%20%22Fibre-optic) lines of 28, 56, 32, or 64 channels; and [sampling rates](https://en.wikipedia.org/wiki/Sampling_rate%22%20%5Co%20%22Sampling%20rate) to 96 kHz and beyond with an [audio bit depth](https://en.wikipedia.org/wiki/Audio_bit_depth%22%20%5Co%20%22) 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 atomospherhic or spot effects.

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 ultra high definition 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.

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).

#### 5.2.3.9 Key Issue #8: Usage of NPN (SNPN or PNI-NPN)

Editor’s Note: SA2 is studying NPN evolutions and results are documented in TR 23.700-07. It is unclear whether additional considerations are needed, e.g. to integrate the NPN and the NPN devices into a Media Production network (e.g. NMOS authorization, etc.). It is expected that credentials for accessing the NPN (establish IP connectivity) and for accessing the Media Production network (access to NMOS applications and devices) are kept separate.

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 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.x Use-Case X

### 6.x.1 Description

Editor’s Note: (text From the SID) To identify the relevant media production use cases (professional, semi-professional, production, contribution), based on existing use-cases from TR 22.287 as well as requirements from TS 22.163, that may benefit from 5G System functionalities. This includes collaboration use cases between media producers and 5G System operators.

<Use-cases from TR 22.827 are preferably broken down into smaller use-cases such as

Multi-camera aspects like synachronization

Usage and purpose of different per-camera flows (like return video)

>

State of the art (current issues in content production)

o Focus on multiple cameras for live video production controlled remotely

o Focus on multiple microphone for live audio production

Workflows/architectures/deployment scenarios

o Live video

o Live Audio

### 6.x.2 Collaboration models and deployment architectures

Editor’s Note: (text from the SID) 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.

### 6.x.3 Identified 5G System features

Editor’s Note: (text from the SID) 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.).

Editor’s Note: (text from the SID) 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.

< e.g. TSN in future 3GPP releases, QoS, Network Slicing>

### 6.x.4 High level call flows

Editor’s Note: (text from the SID) 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.

### 6.x.5 Potential issues

]

# 7 Candidate Solutions

< this section should describe, how identified 5G features are used in context of media production>

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

### 7.1.1 General

As highlighted in clause 6.2.2.3, there is in several scenarios a need to dynamically and proactively control media rates such that not all cameras use the maximum 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.

### 7.1.2 Potential solutions

[TBD]

# 8 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 reportS4-210641: 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 productionS4aI211165: Overview of NMOS functionality | 0.1.1 |
| May 2021 | SA4#114 | S4-210939 |  |  |  | S4-210919: FS\_NPN4AVProd: Utilizing Available Capacity in Multi-Camera ScenariosS4-210913: 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 ProductionS4-211242: [FS\_NPN5AVProd] Proposal of Media Protocol related Key IssuesS4-211243: [FS\_NPN5AVProd] Proposal of a Remote Camera Configuration Key IssueS4-211244: [FS\_NPN5AVProd] Proposal of two bitrate adaptation related Key IssuesS4-211245: [FS\_NPN5AVProd] Proposal of a Key Issue around configurable audio channelsS4-211246: [FS\_NPN5AVProd] Proposal of a new NPN usage related Key Issue |  |