**3GPP TSG-SA WG4 Meeting #130 S4-241937**

**Orlando, USA, 18th – 22nd Nov 2024 (revision of S4aI240190)**

|  |
| --- |
| *CR-Form-v12.0* |
| **CHANGE REQUEST** |
|  |
|  | **26**.**804** | **CR** | **0007** | **rev** | **8** | **Current version:** | **18.1.0** |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME |  | Radio Access Network |  | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | **[FS\_AMD] WT#12: Improved QoS support for Media Streaming services** |
|  |  |
| ***Source to WG:*** | Huawei, HiSilicon, Ericsson |
| ***Source to TSG:*** | S4 |
|  |  |
| ***Work item code:*** | FS\_AMD |  | ***Date:*** | 2024-09-25 |
|  |  |  |  |  |
| ***Category:*** | **B** |  | ***Release:*** | Rel-19  |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-10 (Release 10)Rel-11 (Release 11)Rel-12 (Release 12)**Rel-13 (Release 13)Rel-14 (Release 14)Rel-15 (Release 15)Rel-16 (Release 16)* *Rel-17 (Release 17)* *Rel-18 (Release 18)* |
|  |  |
| ***Reason for change:*** | As agreed in SP-240514, how to improve the QoS support for Media Streaming services based on the QoS enhancements and the network information exposure is to be studied. Therefore, this paper proposes the Key Issue of "Improved QoS support for Media Streaming services". |
|  |  |
| ***Summary of change:*** | Proposal of KI#X: Improved QoS support for Media Streaming services. |
|  |  |
| ***Consequences if not approved:*** | SI cannot be completed. |
|  |  |
| ***Clauses affected:*** | 2, 3.3, 5.23(new) |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **X** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **X** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** | S4-240638, S4-240806, S4-240971, S4-241229, S4-241521SA4#129e: S4-241748 merges S4-241746.SA4-e (AH) MBS SWG post 129e: Provide a clean version of new clauses as basis for future work.Merge S4aI240187.Merge S4aI240200. |

\* \* \* \* First change \* \* \* \*

# 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] Akamai Blog, "A QUICk Introduction to HTTP/3", April 2020, <https://developer.akamai.com/blog/2020/04/14/quick-introduction-http3>

[3] Fielding, R., Nottingham, M., and J. Reschke, "HTTP/1.1", Work in Progress, Internet-Draft, draft-ietf-httpbis-messaging-13, 14 December 2020, http://www.ietf.org/internet-drafts/draft-ietf-httpbis-messaging-13.txt

[4] Belshe, M., Peon, R., and M. Thomson, Ed., "Hypertext Transfer Protocol Version 2 (HTTP/2)", RFC 7540, May 2015, https://www.rfc-editor.org/info/rfc7540

[5] RFC 9114, "Hypertext Transfer Protocol Version 3 (HTTP/3)", June 2022

[6] D. Bhat, A. Rizk, and M. Zink, "Not so QUIC: A Performance Study of DASH over QUIC," NOSSDAV'17: Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and VideoJune 2017 Pages 13–18 https://doi.org/10.1145/3083165.3083175

[7] AWS: "Achieving Great Video Quality Without Breaking the Bank", Streaming Media June 2019, [[https://pages.awscloud.com/rs/112-TZM-766/images/GEN elemental-wp-achieving-great-video-quality-without-breaking-the-bank.pdf](https://pages.awscloud.com/rs/112-TZM-766/images/GEN%20elemental-wp-achieving-great-video-quality-without-breaking-the-bank.pdf)](https://pages.awscloud.com/rs/112-TZM-766/images/GEN%20elemental-wp-achieving-great-video-quality-without-breaking-the-bank.pdf)

[8] Netflix, "Optimized shot-based encodes: Now Streaming!", Netflix Blog, May 2018, https://netflixtechblog.com/optimized-shot-based-encodes-now-streaming-4b9464204830

[9] DASH-IF/DVB: "Report on Low-Latency Live Service with DASH", July 2017, available here: <https://dash-industry-forum.github.io/docs/Report%20on%20Low%20Latency%20DASH.pdf>

[10] DASH-IF: "IOP Guidelines v5, Low-latency Modes for DASH", available here: <https://dash-industry-forum.github.io/docs/CR-Low-Latency-Live-r8.pdf>

[11] ISO/IEC 23009-1: "Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats".

[12] IETF RFC 8673: "HTTP Random Access and Live Content".

[13] 3GPP TR 26.939: "Guidelines on the Framework for Live Uplink Streaming (FLUS)".

[14] 3GPP TS 26.238: "Uplink Streaming".

[15] 3GPP TS 26.501: "5G Media Streaming (5GMS); General description and architecture".

[16] 3GPP TS 26.512: "5G Media Streaming (5GMS); Protocols".

[17] ISO/IEC 13818-1:2019: "Information technology — Generic coding of moving pictures and associated audio information — Part 1: Systems".

[18] SCTE 35 2020: "Digital Program Insertion Cueing Message", <https://www.scte.org/pdf-redirect/?url=https://scte-cms-resource-storage.s3.amazonaws.com/SCTE-35-2020_notice-1609861286512.pdf>

[19] ISO/IEC 23000-19:2020: "Information technology — Multimedia application format (MPEG-A) —Part 19: Common media application format (CMAF) for segmented media".

[20] ISO/IEC 23009-1:2019/DAMD1: "Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats — Amendment 1: CMAF support, events processing model and other extensions".

[21] VSF TR-06-01:2020, "RIST Simple Profile", https://vsf.tv/download/technical\_recommendations/VSF\_TR-06-1\_2020\_06\_25.pdf

[22] VSF TR-06-02: "RIST Main Profile", <https://www.videoservicesforum.org/download/technical_recommendations/VSF_TR-06-2_2020_03_24.pdf>

[23] 3GPP TS 23.501: "System architecture for the 5G System (5GS)".

[24] 3GPP TS 23.502: "Procedures for the 5G System (5GS)".

[25] 3GPP TS 29.517: "5G System; Application Function Event Exposure Service; Stage 3".

[26] 3GPP TS 29.244: "Interface between the Control Plane and the User Plane nodes; Stage 3".

[27] IETF RFC 6733: "Diameter Base Protocol".

[28] 3GPP TS 29.514: "5G System; Policy and Charging Control over Rx reference point; Stage 3".

[29] IETF RFC 7657: "Differentiated Services (Diffserv) and Real-Time Communication", November 1995.

[30] IETF RFC 3168: "The Addition of Explicit Congestion Notification (ECN) to IP", September 2001.

[31] IETF RFC 9204: "QPACK: Field Compression for HTTP/3", June 2022.

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

[33] IETF RFC 9001: "Using TLS to Secure QUIC", May 2021.

[34] IETF RFC 9002: "QUIC Loss Detection and Congestion Control", May 2021.

[35] IETF RFC 5681: "TCP Congestion Control".

[36] M. Kuehlewind and B. Trammell, draft-ietf-quic-manageability-11, "Manageability of the QUIC Transport Protocol", Work in Progress, Internet-Draft, 30 June 2021.

[37] N. Cardwell et. al. "BBR Updates: Internal Deployment, Code, Draft Plans", 9 March 2021, https://datatracker.ietf.org/meeting/110/materials/slides-110-iccrg-bbr-updates-00.pdf

[38] ETSI TS 103 799: "Publicly Available Specification (PAS); DASH-IF Content Protection Information Exchange Format".

[39] ISO/IEC JTC1/SC29/WG11/N19062 23090‑8 FDIS: "MPEG-I: Network-based Media Processing — Network-Based Media Processing Specification".

[40] 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".

[41] 3GPP TS 23.503: "Policy and charging control framework for the 5G System (5GS); Stage 2".

[42] 3GPP TS 29.514: "5G System; Policy Authorization Service; Stage 3".

[43] 3GPP TS 29.522: "5G System; Network Exposure Function Northbound APIs; Stage 3".

[44] 3GPP TS 29.122: "T8 reference point for Northbound APIs".

[45] 3GPP TS 29.512: "5G System; Session Management Policy Control Service; Stage 3".

[46] 3GPP TS 26.803: "5G Media Streaming (5GMS); Architecture extensions".

[47] 3GPP TS 23.558: "Architecture for enabling Edge Applications (EA)".

[48] 3GPP TS 23.288: "Architecture enhancements for 5G System (5GS) to support network data analytics services".

[49] Tdoc S4-210723: "Generic architecture for data collection and reporting", submission from BBC, Dolby Laboratories Inc., LM Ericsson and Qualcomm Incorporated to SA4#114-e, May 19-28, 2021.

[50] Tdoc S2-2103267: "Extension of Naf\_EventExposure for observed service experience data collection from UEs", CR from InterDigital to SA2#144e, Apr 12-16, 2021.

[51] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[52] Tdoc S2-2104496: "Extension of Naf\_EventExposure for observed service experience data collection from UEs", CR from Qualcomm Incorporated to SA2#145e, May 17-28, 2021.

[53] 3GPP TS 26.118: "Virtual Reality (VR) profiles for streaming applications".

[54] 3GPP TS 26.346: "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs".

[55] 3GPP TS 29.554: "Background Data Transfer Policy Control Service; Stage 3".

[56] 3GPP TS 28.530: "Management and orchestration; Concepts, use cases and requirements".

[57] 3GPP TS 28.531: "Management and orchestration; Provisioning".

[58] 3GPP TS 28.532: "Management and orchestration; Generic management services".

[59] 3GPP TS 28.533: "Management and orchestration; Architecture framework".

[60] 3GPP TS 28.540: "Management and orchestration; 5G Network Resource Model (NRM); Stage 1".

[61] 3GPP TS 28.541: "Management and orchestration; 5G Network Resource Model (NRM); Stage 2 and stage 3".

[62] 3GPP TS 28.542: "Management and orchestration of networks and network slicing; 5G Core Network (5GC) Network Resource Model (NRM); Stage 1".

[63] 3GPP TS 28.543: "Management and orchestration of networks and network slicing; 5G Core Network (5GC) Network Resource Model (NRM); Stage 2 and stage 3".

[64] 3GPP TS 28.545: "Management and orchestration; Fault Supervision (FS)".

[65] 3GPP TS 28.546: "Management and orchestration of networks and network slicing; Fault Supervision (FS); Stage 2 and stage 3".

[66] 3GPP TS 28.552: "Management and orchestration; 5G performance measurements".

[67] 3GPP TS 28.554: "Management and orchestration; 5G end to end Key Performance Indicators (KPI)".

[68] 3GPP TS 23.434: " Service Enabler Architecture Layer for Verticals (SEAL); Functional architecture and information flows ".

[69] 3GPP TS 23.700‑99: " Study in Network slice capability exposure for application layer enablement (NSCALE)".

[70] 3GPP TS 29.520: " 5G System; Network Data Analytics Services; Stage 3".

[71] 3GPP TR 23.700-40: "Study on enhancement of network slicing; Phase 2".

[72] 3GPP TS 26.531: “Data Collection and Reporting; General Description and Architecture”.

[73] 3GPP TR 26.802: "Multicast Architecture Enhancement for 5G Media Streaming".

[74] IETF RFC 822: "STANDARD FOR THE FORMAT OF ARPA INTERNET TEXT MESSAGES", August 13, 1982.

[75] IETF RFC 1521: "MIME (Multipurpose Internet Mail Extensions)", September 1993.

[76] IETF RFC 2474: "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers".

[77] IETF RFC 2475: "An Architecture for Differentiated Services".

[78] IETF RFC 3246: "An Expedited Forwarding PHB (Per-Hop Behavior)".

[79] IETF RFC 2597: "Assured Forwarding PHB Group".

[80] S. Hurst, draft-hurst-quic-rtp-tunnelling: "QRT: QUIC RTP Tunnelling", Internet-Draft, Work in Progress.

[81] J. Ott and M. Engelbart, draft-engelbart-rtp-over-quic: "RTP over QUIC", Internet-Draft, Work in Progress.

[82] SRT Alliance, “Secure Reliable Transport (SRT) Protocol”, https://github.com/Haivision/srt

[83] M.P. Sharabayko and M.A. Sharabayko, draft-sharabayko-srt-over-quic-00 ,“Tunnelling SRT over QUIC”, Internet-Draft, Work in Progress, 28 July 2021.

[84] Robin Marx, Luca Niccolini, Marten Seemann, draft-ietf-quic-qlog-main-schema-01, "Main logging schema for qlog", Internet-Draft, Work in Progress, 25 October 2021.

[85] Robin Marx, Luca Niccolini, Marten Seemann, draft-ietf-quic-qlog-h3-events-00, "HTTP/3 and QPACK event definitions for qlog", Internet-Draft, Work in Progress, 10 June 2021.

[86] Robin Marx, Luca Niccolini, Marten Seemann, draft-ietf-quic-qlog-quic-events-00, "QUIC event definitions for qlog", Internet-Draft, Work in Progress, 10 June 2021.

[87] Roger Pantos and William May, Jr., "HTTP Live Streaming", RFC 8216, August 2017.

[88] 3GPP TR 26.925: "Typical traffic characteristics of media services on 3GPP networks".

[89] 3GPP TR 26.917: "Multimedia Broadcast Multicast Services (MBMS) and Packet-switchedStreaming Service (PSS) enhancements to support television services".

[90] "DASH-IF WebRTC-based Streaming", https://dashif.org/news/webrtc/

[91] IETF RFC 6749: "The OAuth 2.0 Authorization Framework".

[92] IETF RFC 6750: "The OAuth 2.0 Authorization Framework: Bearer Token Usage".

[93] 3GPP TS 33.501: "Security architecture and procedures for 5G System".

[94] 3GPP TS 26.531: "Data Collection and Reporting; General Description and Architecture".

[95] 3GPP TS 26.532: "Data Collection and Reporting; Protocols and Formats".

[96] 3GPP TS 26.511: "5G Media Streaming (5GMS); Profiles, codecs and formats".

[97] ETSI TS 103 770: "Digital Video Broadcasting (DVB); Service Discovery and Programme Metadata for DVB-I".

[98] Android Developer Documentation: "Handling Android App Links",
<https://developer.android.com/training/app-links>

[99] 3GPP TS 26.347: "Multimedia Broadcast/Multicast Service (MBMS); Application Programming Interface and URL".

[100] ETSI TS 103 769: "Digital Video Broadcasting (DVB); Adaptive media streaming over IP multicast".

[101] 3GPP TS 23.247: "Architectural enhancements for 5G multicast-broadcast services".

[102] 3GPP TS 29.558: "Enabling Edge Applications; Application Programming Interface (API) specification; Stage 3".

[103] IETF RFC 2045: "Multipurpose Internet Mail Extensions Part One: Format of Internet Message Bodies".

[104] IETF RFC 3986: "Uniform Resource Identifier (URI): Generic Syntax".

[X1] IETF RFC 9330:"Low Latency, Low Loss, Scalable Throughput (L4S) Internet Service: Architecture".

[X2] IETF RFC 9331: "Explicit Congestion Notification (ECN) Protocol for Very Low Queuing Delay (L4S)".

[X3] IETF RFC 9332: "Dual-Queue Coupled Active Queue Management (AQM) for Low Latency, Low Loss, and Scalable Throughput (L4S)".

[38300] 3GPP TS 38.300: "NR; NR and NG-RAN Overall Description".

[26522] 3GPP TS 26.522: "5G Real-time Media Transport Protocol Configurations".

[X6] IETF RFC 3168: "The Addition of Explicit Congestion Notification (ECN) to IP".

[29122] 3GPP TS 29.122: "T8 reference point for Northbound APIs".

\* \* \* \* Second change \* \* \* \*

## 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].

CDN Content Delivery Network

DS Differentiated Service

EAS Edge Application Server

ECN Explicit Congestion Notification

EES Edge Enabler Server

FAR Forward Action Rule

L4S Low Latency, Low Loss and Scalable ThroughputMAR Multi-Access Rule

NRF Network Repository Function

PDR Packet Detection Rule

PDU Protocol Data Unit

PFCP Packet Forwarding Control Protocol

PSA PDU Session Anchor

PSDB PDU Set Delay Budget

PSER PDU Set Error Rate

PSIHI PDU Set Integrated Information

QER QoS Enforcement Rule

QLOG QUIC Logging

PHB Per-Hop Behaviour

PFD Packet Flow Description

SDF Service Data Flow

URL Uniform Resource Locator

URR Usage Reporting Rule

\* \* \* \* Third change (all new text) \* \* \* \*

## 5.23 Improved QoS support for Media Streaming services

### 5.23.1 Description

#### 5.23.1.1 General

QoS support for Media Streaming services was introduced in Release 16. For example, the dynamic policy feature allows a 5GMSd Client to request specific QoS handling of its application flows, and the network assistance feature allows it to find out about the current network status. New QoS enhancements and the network information exposure have been introduced in recent releases, which could be useful for Media Streaming services.

This Key Issue proposes to study whether and how to integrate the new features of 5GS to improve the QoS support for Media Streaming services.

#### 5.23.1.2 QoS enhancements and network information exposure in 5GS

##### 5.23.1.2.1 Support of ECN marking for L4S

As described in RFC 9330 [X1], RFC 9331 [X2] and RFC 9332 [X3], the purpose of ECN marking for L4S (Low Latency, Low Loss and Scalable Throughput) is to inform a recipient host at the earliest opportunity that an IP packet has experienced network congestion at some point in its routing path. It exposes congestion information by marking ECN bits in the IP header of the user IP packets between the UE and the application server. This early notification may be used by the receiving application to report the congestion to its sending peer using a suitable Layer 4 feedback mechanism. Based on this feedback, the sender should reduce the sending bit rate. In the context of adaptive segmented media delivery, application layer rate adaptation may be needed in addition. For example, a media player consuming a media presentation that receives an ECN-marked downlink packet from a streaming media server may induce a reduction in the sender’s bit rate by switching to a representation of a lower bit rate. To support this functionality, the recipient host needs to support L4S feedback as described in RFC 9330 [X1].

L4S is based on the idea that delay is mainly caused by the classic congestion control algorithms introduced with TCP. L4S replaces these traditional congestion control algorithms with a class of scalable congestion control algorithms.

The L4S architecture relies on three components to operate:

- A scalable congestion control algorithm,

- A modified ECN marking behaviour,

- An active queue management algorithm that isolates L4S traffic

An example of a scalable congestion control algorithm that is widely deployed today is TCP Prague. In TCP Prague the congestion window is adjusted proportionally to the probability of receiving an ECN mark.

$cwnd= \frac{1}{p^{α}}$ , where p is the probability of receiving an ECN mark, and $α$ is a constant

TCP Prague further adjust the congestion window to implement an RTT independence, which is crucial to not overly react to RTT variations. The additive increase part of the algorithm is also adjusted to ensure that flows with short RTT are not penalized by a slow increase.

Another popular scalable congestion control algorithm is BBRv2, which stands for Bottleneck Bandwidth and Round-trip propagation time. BBRv2 continuously estimates the bottleneck bandwidth of the connection and the RTT. It then uses the two parameters to adjust its congestion window using the formula $cwnd=BtlBW\*RTprop$ , where BtlBW is the estimated bottleneck bandwidth and RTprop is the estimated minimum round-trip propagation time.

The algorithm’s reaction to packet loss and ECN marks is also modified to be less aggressive, compared to classic congestion control algorithms. To accurately estimate the bottleneck bandwidth, BBRv2 periodically sends data at a high rate to probe the network for the maximum throughput. BBRv2 also tracks the data that is sent and received during a RTT cycle.

The ECN mechanism is also modified in L4S. The default reaction to an ECN mark is to treat it as an equivalent to a packet loss. This is no more the case in L4S, which allows the network to signal the start of congestion more frequently without the risk of messing up the connection’s bandwidth. This change requires the identification of L4S connections by the network. The ECT code point is used for this purpose. When set to 1, the network identifies the sender as L4S capable and deploys the fine-tuned congestion notification mechanism.

Finally, the queue management approach closes the loop for L4S. RFC9332 [X3] defines one such management approach, the Dual Queue Coupled Active Queue Management, which in essence separates the L4S and non-L4S traffic into two separate queues for differentiated latency treatment and ECN marking. The available bandwidth is still shared among both queues though through the coupling mechanism.

According to clause 6.1.3.22 of TS 23.503 [41], an Application Function may provide an explicit indication that the uplink and/or downlink path of a service data flow supports ECN marking for L4S by means of the Nnef\_AfsessionWithQoS service at reference point N33 or the Npcf\_PolicyAuthorization service at reference point N5. The indication is supported at both MediaComponent and MediaSubComponent levels, which provides a high degree of flexibility on its usage. Based on AF input and/or local configuration, the PCF indicates to the SMF that ECN marking for L4S is enabled for that service data flow. The SMF accordingly configures ECN marking for the corresponding QoS Flow in the uplink and/or downlink direction. ECN marking for L4S in the IP header is supported in either the NG-RAN (see clause 5.37.3.2 and TS 38.300 [38300]), or in the PDU Session Anchor (PSA) UPF (see clause 5.37.3.3 of TS 23.501[23]).

In the case of ECN marking for L4S by the PSA UPF, the NG-RAN is instructed to perform congestion information monitoring and report to the PSA UPF the congestion information of the QoS Flow in the uplink and/or downlink directions via GTP-U header extension. Accordingly, the PSA UPF may mark the uplink and/or downlink direction packets as congested.

##### 5.23.1.2.2 Support of PDU Set handling

A PDU Set is comprised of one or more PDUs carrying an application layer payload that together form a logical access unit such as a video frame or a slice of a video frame.

The AF may provide a Protocol Description and PDU Set QoS Parameters to the 5GC (i.e. PCF) by means of the Nnef\_AfsessionWithQoS service at reference point N33 or the Npcf\_PolicyAuthorization service at reference point N5.

- The Protocol Description is used to assist UPF/UE in identifying PDUs that belong to a PDU Set. This may involve deep packet inspection of the PDU payload by the UPF (downlink PDUs) or by the UE (uplink PDUs).

- The PDU Set QoS parameters, including a PDU Set Integrated Handling Information (PSIHI), PDU Set Delay Budget (PSDB) and PDU Set Error Rate (PSER), are used to instruct the PDU Set based handling in NG-RAN.

To support QoS handling of PDU Sets in the downlink direction, the PSA UPF identifies PDUs that belong to PDU Sets based on a protocol description (e.g. the RTP Header Extension defined in TS 26.522 [26522]) if available or else in an implementation-specific way), and determines the following PDU Set Information which it sends to the NG-RAN in the GTP-U header. The PDU Set information is used by the NG-RAN for QoS handling of PDU Sets as described above.

The PDU Set Information comprises:

- PDU Set Sequence Number.

- Indication of End PDU of the PDU Set.

- PDU Sequence Number within a PDU Set.

- PDU Set Size in bytes.

- PDU Set Importance, which identifies the relative importance of a PDU Set compared to other PDU Sets within a QoS Flow.

Based on the PDU Set QoS parameters provided by the 5GC and the PDU Set Information carried over the GTP-U header of downlink packets, the NG-RAN applies PDU Set QoS handling accordingly.

In the uplink direction, based on the PDU Set QoS parameters, the RAN configures the UE to apply QoS handling to PDU Sets. Uplink PDU Sets are identified by the UE based on the protocol description or else in an implementation-specific way.

##### 5.23.1.2.3 Support of QoS monitoring

QoS monitoring comprises of measurements of QoS monitoring parameters and reports of the measurement result for a service data flow (i.e., QoS Flow) and can be enabled based on third-party application requests and/or operator policies configured in the 5GC (i.e. PCF).

The AF may request measurements and subscribe to the event for one or more of the following QoS monitoring parameters by means of the Nnef\_AfsessionWithQoS service at reference point N33 or the Npcf\_PolicyAuthorization service at reference point N5, which may trigger QoS monitoring for service data flow(s):

- Uplink packet delay, downlink packet delay and round-trip packet delay for a service data flow (see clause 5.45.2 of TS 23.501 [23]).

- Congestion (see clause 5.45.3 of TS 23.501 [23]).

- Data Rate (see clause 5.45.4 of TS 23.501 [23]).

- Packet Delay Variation (see clause 5.37.7 of TS 23.501 [23]).

- Round-trip packet delay considering the uplink path of one service data flow and the downlink path of another service data flow (see clause 5.37.4 of TS 23.501 [23]).

Using the QoS monitoring mechanisms of the 5G Core, the above parameters can be derived and further exposed to the AF via the PCF or the UPF (directly or further via NEF) as requested.

#### 5.23.1.3 Key Issue objectives

Regarding the features described in clause 5.23.1.2, it is proposed to study:

- Whether these features of the 5G System can be beneficial and valid for the Media Delivery System in the context of segmented media delivery (i.e., 5G Media Streaming):

- Whether ECN marking for L4S can be beneficial and valid.

- Whether PDU Set handling can be beneficial and valid.

- Whether QoS monitoring can be beneficial and valid.

- How to apply these features to the Media Delivery System:

- How to integrate the ECN marking for L4S feature into the Media Delivery System.

- How to integrate the PDU Set handling feature into the Media Delivery System.

- How to integrate the QoS monitoring feature into the Media Delivery System.

### 5.23.2 Collaboration scenarios

#### 5.23.2.1 General

Collaboration scenarios 2–11 and 13–15 from TS 26.501 [15] are potential points of departure for improved QoS handling support with the following additions:

1. Similar to the Network Assistance feature in TS 26.501 [15], the network status of the 5G System may be exposed to media delivery sessions using the *QoS monitoring* feature and the *ECN marking for L4S* feature. The network status, including the data rate, latency, congestion, etc. may be used by the Media Delivery System for bit rate adaptation and/or congestion control.

 The PDU Set handling feature may be used to label PDUs belonging to a video frame or video slice as members of the same PDU Set.

NOTE: Whether the concept of PDU Set is applicable to media segments in a segment-based streaming service is not clear.

2. In the case of network congestion, the NG-RAN may consider the PDU Set Importance when decideding whether to discard packets at the level of a PDU Set level. This is not expected for delivery of media segments over reliable transports such as TCP or a stream-based QUIC.

NOTE: Whether the PDU Set feature is beneficial for segment-based Media Streaming services is for future study.

#### 5.23.2.2 Collaboration scenarios for L4S ECN marking

Collaboration scenarios for L4S ECN marking are depicted below. Both the Media AS and the 5GMS Client make use of an L4S-enabled protocol stack. Figure 5.23.2.2-1 assumes that the Media AS resides within the external DN, while figure 5.23.2.2-2 assumes the Media AS within the Trusted DN.



Figure 5.23.2.2-1: Media AS in External DN



Figure 5.23.2.2-1: Media AS in Trusted DN

### 5.23.3 Architecture mapping

Not applicable to this Key Issue.

### 5.23.4 High-level call flows

#### 5.23.4.1 Integrating QoS monitoring and/or ECN marking for L4S

The high-level call flow for integrating the QoS monitoring and/or ECN marking for L4S is shown below as well as the corresponding procedures.

It is assumed that the MNO and the 5GMS Application Provider have negotiated a Service Level Agreement that allows the 5GMS Application Provider to enable the ECN marking for L4S and QoS monitoring in the 5G System for media delivery.



Figure 5.23.4.1-1: Potential call flow for improved QoS handling support

Prerequisites:

- The 5GMS Application Provider has agreed an SLA with the Network Operator to allow the usage of network assistance for Media Streaming service.

Steps:

1. The 5GMS Application Provider provisions the 5G Media Streaming System configures content ingest. **A Network Assistance configuration is provided to allow the usage of ECN marking for L4S and/or QoS monitoring to notify the 5GMS Client of the latest network status.**

NOTE: Whether to introduce the feature of "Improved QoS Support" or reuse the "Network Assistance" feature is for further study.

2. When the 5GMS-Aware Application starts, the Media Session Handler retrieves the Service Access Informaiton via M5 or M8. The 5GMS AF address that offers the network assistance is provided in the Service Access Information **and the options for QoS monitoring and/or ECN marking are also present**.

3. The Media Session Handler invokes the **Enhanced** Network Assistance API on the 5GMS AF **to find out about the latest network status. For instance, the 5GMS Media Session Handler may subscribe to the periodic congestion status report from 5GMS AF.**

4. The 5GMS AF interacts with the PCF or NEF to enable QoS monitoring and/or ECN marking for L4S in the 5G System via the Npcf\_PolicyAuthorization service at reference point N5 or the Nnef\_AFsessionWithQoS service at reference point N33.

5. In the case of QoS monitoring, the 5GMS AF can receive the notifications from PCF or NEF via the Npcf\_PolicyAuthorization\_Notify at reference point N5 or the Nnef\_AFsessionWithQoS\_Notify at reference point N33.

6. Alternatively, in the case of QoS monitoring, the 5GMS AF may receive the notifications directly from the UPF via the Nupf\_EventExposure\_Notify at reference point N5 or from NEF the Nnef\_EventExposure\_Notify service at reference point N33. This is beneficial when the 5GMS AF is deployed in the Edge DN and the SMF/PCF is generally deployed centrally.

7. The 5GMS AF further sends the notifications exposed by the network to the Media Session Handler using the MQTT notification channel for the Provisioning Session. The 5GMS Client may take this into account for rate adaptaion, congestion/flow control.

#### 5.23.4.2 QoS monitoring for media streaming



Figure 5.23.4.2-1: High-level call flow for QoS monitoring for Media Streaming

1. 5GMS Application Provider provisions the 5GMS AF with the **Network Assistance configuration** as described in step 1 of clause 5.23.4.1 The **Network Assistance configuration** contains the configuration of QoS monitoring, including the parameters to be monitored, reporting frequency (event triggered, periodic), optionally target entity of reporting and optionally the notification via UPF.

NOTE: In case the 5GMS AS is deployed as an EAS instance in the Edge DN, a local UPF can also be inserted for local access to the 5GMS EAS. In order to reduce the latency used for exposure of the QoS monitoring results, the local UPF is expected to provide the notifications of network status directly to the 5GMS AF and 5GMS AS, or via a locally deployed NEF as defined in clause 5.8.2.17 of TS 23.501 [23].

2. The Media Session Handler retrieves Service Access Information with the configuration of QoS monitoring provided inside the client Network Assistance configuration.

3. If the Media Session Handler is interested in understanding the network status (e.g., congestion status, packet latency) it creates an enhanced Network Assistance Session **that includes the requested QoS montoring configuration** on the 5GMS AF at reference point M5.

4. Based on the QoS monitoring configuration received in the previous step, , or based on the provisioning from the 5GMS Application Providerrequiring 5GMS AS traffic control, e.g. congestion control, bit rate adaptation for progressive download, the 5GMS AF recognises the need to enable QoS monitoring.

5. The 5GMS AF invokes the Npcf\_PolicyAuthorization service or the Nnef\_AFsessionWithQoS service **with the requested QoS monitoring configurations**. in this step

6. The PCF accepts the request and enables QoS monitoring within the 5G System, i.e., by configuring the RAN and/or the UPF and/or the local UPF for monitoring and reporting of target QoS parameters.

7. Following the QoS monitoring request(s), the PCF exposes the QoS monitoring results to the 5GMS AF periocially or by event triggers.

**8. Alternatively, the QoS monitoring results can be exposed to the 5GMS AF by the UPF directly using the Nupf\_EventExposure\_Notify service or via a locally deployed NEF using the Nnef\_EventExposure\_Notifyservice at reference point N33.**

9. If QoS monitoring was requested by the Media Session Handler, **the 5GMS AF sends the notifications of the QoS monitoring results to the Media Session Handler** via the MQTT notification channel at reference point M5 associated with the Network Assistance Session.

**10. The Media Session Handler further provides the QoS monitoring results to the Media Stream Handler at reference point M11.**

**11. The Media Stream Handler may use the notified QoS monitoring results to modify its behaviour.**

For example, in the case of downlink media streaming, the Media Player may use the monitored packet latency to determine when to request the next media segment, and/or to change the bit rate of the next media segemtn based on the monitored congestion status.

NOTE: Whether notification of network status by the 5GMS AF to the 5GMS AS at reference point M3 is practical, useful and desirable is for further study. This could, for example, use the same asynchronous MQTT-based notification mechanism available at reference point M5.

#### 5.23.4.3 L4S-on-request for downlink media streaming

An Application Function may request L4S support from the 5G Network for a certain QoS Flow, e.g. by invoking the Nnef\_AfsessionWithQoS service. The concept of this solution is that an application only requests L4S support from the network when the application layer provides support. The activation leverages the existing 5GMS Dynamic Policy invocation API, allowing the 5GMS-Aware Application to request L4S support as and when it is needed.

A high-level call flow for downlink media streaming is sketched in figure 5.23.4.3-1 below. The following is assumed:

- The service here is a unicast downlink media streaming service with dynamic policy support, as described in clause 5.7 of TS 26.501 [15].

- The Layer 4 protocol used for application flows is TCP and the TCP stack used supports L4S.

- The network supports L4S packet marking.

- The application has specifically requested ECN marking for its media delivery session.

- NG-RAN manipulaties the ECN bits (per clause 5.37.3.2 of TS 23.501 [23]). It is equally possible that the PSA-UPF manipulates the ECN bits (per clause 5.37.3.3 of [23]).



Figure 5.23.4.3-1: Downlink media streaming call flow for L4S on request

The steps are as follows:

0: *Policy Template Provisioning.* A Policy Template is provisioned **with the requirement for L4S capability indicated by setting a flag**.

1: *Dynamic Policy activation.* The Media Session Handler within the 5GMSd Client obtains Service Access Information and triggers a dynamic policy activation. A Policy Template Binding is present within the Service Access Information for each provisioned Policy Template. **Policy Template Bindings suitable for L4S are indicated by an L4S capability requirement flag being set. The 5GMSd Client detects that an L4S-capable media transport stack is present and in use. The selected Policy Template is one configured with L4S capability.**

2: *QoS request.* The 5GMSd AF requests QoS handling using e.g. the Nnef\_AfSessionWithQoS service or the Npcf\_PolicyAuthorization service. **If the L4S capability requirement flag is set in the selected Policy Template, this indicates that the new QoS flow is required to be L4S-enabled.** The new QoS flow with the L4S indication setting propagates through the 5G System.

3: **If the L4S capability requirement flag is set in the Policy Template Binding for the selected Policy Template, the 5GMSd Client selects/enables the L4S capability of the used transport protocol.**

NOTE: This step may happen implicitly by selecting an L4S-supporting transport protocol stack.

4: The Media Player within the 5GMSd Client triggeres the establishment of a TCP connection. The ECT(1) codepoint is set in the IP header, indicating an L4S-Capable Transport, and the SDAP entity ensures that the packet is forwarded via the matching QoS flow.

5: The 5GMSd AS responds to the TCP connection establishment request. The 5GMSd AS sets ECT(1) in the IP headers, indicating an L4S-Capable Transport.

6: The UPF finds the matching QoS flow identifier for the downlink packet and sends the packet via the according QoS flow to the UE. TCP Connection setup continues, with one ECT bit set in all packets.

7. When the RAN detects an upcoming congestion (based on continuous congestion monitoring), the 5G System sets the CE (Congestion Experienced) codepoint in the IP header of the downlink packet.

8. The TCP protocol stack used by the Media Player in the 5GMSd Client reflects the Early Congestion Notification to the TCP sender by setting the ECN-Echo (ECE) flag in the TCP header of an uplink PDU of the same TCP connnection. The TCP sender reacts to the ECN-Echo accordingly (i.e., by reducing its sending congestion window).

NOTE 1: The ECN-Echo flag is also acknowledged by the TCP sender setting the Congestion Window Reduced (CWR) flag in an outgong TCP frame, but this acknowledgement is not illustrated in this call flow.

NOTE 2: Classic ECN as specified in RFC 3168 [X6] requires an ECN signal to be treated as equivalent to a packet drop. L4S as specified in RFC 9330 [X1] specifies a more fine-grained response and an early congestion signal triggers a less severe reaction. How a TCP sender behaves “accordingly” is not in scope of the specification.

9. Based on the CE indication received in step 7, or by detecting a reduced bit rate in the downlink application flow, the Media Player in the 5GMSd Client reacts by, for example, changing the requested representation.

#### 5.23.4.4 L4S-on-request for uplink media streaming

Support for uplink media streaming is very similar to that for downlink media streaming.

A high-level call flow for uplink media streaming is sketched in figure 5.23.4.3-1 below. The following is assumed:

- The service here is a unicast uplink media streaming service with dynamic policy support, as described in clause 6.9 of TS 26.501 [15].

- The Layer 4 protocol used for application flows is TCP and the TCP stack used supports L4S.

- The network supports L4S packet marking.

- The application has specifically requested ECN marking for its media delivery session.

- NG-RAN manipulaties the ECN bits (per clause 5.37.3.2 of TS 23.501 [23]). It is equally possible that the PSA-UPF manipulates the ECN bits (per clause 5.37.3.3 of [23]).



Figure 5.23.4.4-1: Uplink media streaming call flow for L4S on request

The steps are as follows:

0: *Policy Template Provisioning.* A Policy Template is provisioned **with the requirement for L4S capability indicated by setting a flag**.

1: *Dynamic Policy activation.* The Media Session Handler within the 5GMSu Client obtains Service Access Information and triggers a dynamic policy activation. A Policy Template Binding is present within the Service Access Information for each provisioned Policy Template. **Policy Template Bindings suitable for L4S are indicated by an L4S capability requirement flag being set. The 5GMSu Client detects that an L4S-capable media transport stack is present and in use. The selected Policy Template is one configured with L4S capability.**

2: *QoS request.* The 5GMSu AF requests QoS handling using e.g. the Nnef\_AfSessionWithQoS service or the Npcf\_PolicyAuthorization service. **If the L4S capability requirement flag is set in the selected Policy Template, this indicates that the new QoS flow is required to be L4S-enabled.** The new QoS flow with the L4S indication setting propagates through the 5G System.

3: **If the L4S capability requirement flag is set in the Policy Template Binding for the selected Policy Template, the 5GMSu Client selects/enables the L4S capability of the used transport protocol.**

NOTE: This step may happen implicitly by selecting an L4S-supporting transport protocol stack.

4: The Media Streamer within the 5GMSu Client triggeres the establishment of a TCP connection. The ECT(1) codepoint is set in the IP header, indicating an L4S-Capable Transport, and the SDAP entity ensures that the packet is forwarded via the matching QoS flow.

5: The 5GMSu AS responds to the TCP connection establishment request. The 5GMSu AS sets ECT(1) in the IP headers, indicating an L4S-Capable Transport.

6: The UPF finds the matching QoS flow identifier for the downlink packet and sends the packet via the according QoS flow to the UE. TCP Connection setup continues, with one ECT bit set in all packets.

7. When the RAN detects an upcoming congestion (based on continuous congestion monitoring), the 5G System sets the CE (Congestion Experienced) codepoint in the IP header of the uplink packet.

8. The TCP protocol stack used by the 5GMSu AS reflects the Early Congestion Notification to the TCP sender by setting the ECN-Echo (ECE) flag in the TCP header of a downlink PDU of the same TCP connnection. The TCP sender reacts to the ECN-Echo accordingly (i.e., by reducing its sending congestion window).

NOTE 1: The ECN-Echo flag is also acknowledged by the TCP sender setting the Congestion Window Reduced (CWR) flag in an outgong TCP frame, but this acknowledgement is not illustrated in this call flow.

NOTE 2: Classic ECN as specified in RFC 3168 [X6] requires an ECN signal to be treated as equivalent to a packet drop. L4S as specified in RFC 9330 [X1] specifies a more fine-grained response and an early congestion signal triggers a less severe reaction. How a TCP sender behaves “accordingly” is not in scope of the specification.

9. Based on the CE indication received in step 7, or by detecting a reduced bit rate in the uplink application flow, the Media Streamer in the 5GMSu Client reacts by, for example, changing the requested representation.

### 5.23.5 Gap analysis and requirements

#### 5.23.5.1 Integrating QoS monitoring and/or ECN marking for L4S

#### 5.23.5.2 QoS monitoring for media streaming

Based on the call flow in clause 5.23.4.3, the following observations are made:

- The 5GMS AF needs to explicitly request QoS monitoring by the 5G System for specific parameters (i.e., congestion information, packet latency, data rate and Packet Delay Variation) by interacting with the PCF at reference point N5 (or else via the NEF at reference poiont N33).

- The Policy Template resource structure at reference point M1 needs to be extended to include the QoS monitoring configuration, including the parameters to be monitored, reporting frequency (event triggered, periodic), optionally the target entity of reporting and optionally the notification via UPF.

- The Policy Template Binding data structure carried in the Service Access Information resource at reference point M5 needs to be extended to reflect the QoS monitoring configuration in the corresponding Policy Template.

- QoS monitoring results need to be exposed to the 5GMS AF, either directly at reference point N5 via the Nupf\_EventExposure\_Notifyservice, or else via a NEF using the Nnef\_EventExposure\_Notifyservice at reference point N33.

- To expose QoS monitoring results to the Media Session Handler in the 5GMS Client, notification events relating to Dynamic Policies at reference point M5 need to be extended to include the QoS monitoring results.

- The QoS monitoring results need to be further provided to the Media Stream Handler by the Media Session Handler at reference point M11.

#### 5.23.5.3 L4S-on-request for downlink and uplink media streaming

Based on the call flow in clause 5.23.4.3, the following observations are made:

- L4S/ECN does not require modifications to the Media Player or a TCP based Media Streamer.

- The 5GMS AF needs to explicitly request L4S handling of packets by the 5G System by interacting with the PCF at reference point N5 (or else via the NEF at reference poiont N33).

- The Policy Template resource structure at reference point M1 needs to be extended to include an L4S capability requirement flag.

- The Policy Template Binding data structure carried in the Service Access Information resource at reference point M5 needs to be extended to reflect the value of the L4S capability requirement flag in the corresponding Policy Template.

- An L4S-capable transport protocol stack is required in both the 5GMS Client and at the 5GMS AS.

NOTE: When the transport protocol stack used on the UE or the Application Server does not support ECN marking, the ECT flags are set accordingly to explicitly indicate lack of support.

- Depending on the transport stack implementation, an explicit L4S activation is required at session start.

### 5.23.6 Candidate solutions

#### 5.23.6.1 Integrating QoS monitoring and/or ECN marking for L4S

Candidate solutions addressing this topic are for future study.

#### 5.23.6.2 QoS monitoring for media streaming

Provisioning information is provided by the 5GMS Application Provider at reference point M1 to declare that a Policy Template requires QoS monitoring. The Policy Template structure is enhanced to provide a QoS monitoring configuration (qosMonInfoand directNotifInd as described in clause 5.14.2.1.2 of TS 29.122 [29122]). This QoS monitoring configuration is also provided to the Media Session Handler in the Policy Template binding exposed in Service Access Information.

In this candidate solution, two Policy Templates may be provisioned by the 5GMS Application Provider, one with QoS monitoring configuration and one without. The Media Session Handler in the 5GMS Client then instantiates the appropriate Policy Template depending on its requirements.

When the QoS monitoring configuration is included in the instantiated Policy Template, the 5GMS AF requests QoS monitoring by the 5G System and the 5G System enables the QoS monitoring as requested.

The DynamicPolicy resource is extended to include the QoS monitoring results. When the 5G System reports the QoS monitoring results to the 5GMS AF as requested, the 5GMS AF further provides the notification of the QoS monitoring results to the Media Session Handler via the asynchronous MQTT notification channel as an Application Message conveyed as the payload of an MQTT PUBLISH message.

The Dynamic Policy client API is extended to support the notification events relating to Dynamic Policy. When the QoS monitoring results are received by the Media Session Handler, it further exposes the QoS monitoring results to the Media Stream Handler to react accordingly.

#### 5.23.6.3 L4S-on-request for downlink and uplink media streaming

Provisioning information is provided by the 5GMS Application Provider at reference point M1 to declare that a Policy Template requires L4S support. The Policy Template structure is enhanced to offer a L4S enablement flag. This flag is also exposed to the Media Session Handler in the Policy Template binding exposed in Service Access Information.

In this candidate solution, two Policy Templates may be provisioned by the 5GMS Application Provider, one with L4S enabled and one without. The Media Session Handler in the 5GMS Client then instantiates the appropriate Policy Template depending on its requirements.

When the L4S flag is set in the instantiated Policy Template, the 5GMS AF requests L4S handling by the 5G System and the 5G System assumes that the traffic is L4S enabled.

### 5.23.7 Summary and conclusions

\* \* \* \* End of changes \* \* \* \*