PROTOCOL DESCRIPTION FOR TRAFFIC DIFFERENTIATION AND OPTIMIZED QOS OF MULTIMODAL IP FLOWS

TECHNICAL FIELD

[0001] The subject matter disclosed herein relates generally to the field of implementing protocol description for traffic differentiation and optimized QoS of multimodal IP flows. This document defines a client for multimodal network traffic differentiation of application data, a processor for multimodal network traffic differentiation of application data, a method performed by a client, the method for multimodal network traffic differentiation of application data.

BACKGROUND

[0002] A wireless communications system may include one or multiple network communication devices, such as base stations, which may support wireless communications for one or multiple user communication devices, which may be otherwise known as user equipment (UE), or other suitable terminology. The wireless communications system may support wireless communications with one or multiple user communication devices by utilizing resources of the wireless communication system (e.g., time resources (e.g., symbols, slots, subframes, frames, or the like) or frequency resources (e.g., subcarriers, carriers, or the like). Additionally, the wireless communications system may support wireless communications across various radio access technologies including third generation (3G) radio access technology, fourth generation (4G) radio access technology, fifth generation (5G) radio access technology, among other suitable radio access technologies beyond 5G (e.g., sixth generation (6G)).

SUMMARY

[0003] An article “a” before an element is unrestricted and understood to refer to “at least one” of those elements or “one or more” of those elements. The terms “a,” “at least one,” “one or more,” and “at least one of one or more” may be interchangeable. As used herein, including in the claims, “or” as used in a list of items (e.g., a list of items prefaced by a phrase such as “at least one of” or “one or more of” or “one or both of”) indicates an

inclusive list such that, for example, a list of at least one of A, B, or C means A or B or C or AB or AC or BC or ABC (i.e., A and B and C). Also, as used herein, the phrase "based on" shall not be construed as a reference to a closed set of conditions. For example, an example step that is described as "based on condition A" may be based on both a condition A and a condition B without departing from the scope of the present disclosure. In other words, as used herein, the phrase "based on" shall be construed in the same manner as the phrase "based at least in part on. Further, as used herein, including in the claims, a "set" may include one or more elements.

[0004] Accordingly, there is provided a client for multimodal network traffic differentiation of application data, the client comprising: at least one memory; and at least one processor coupled with the at least one memory and configured to cause the client to: generate a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements; multiplex the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description; and send a request for an application session with network traffic differentiation to a server, the request comprising: a common multimodal session identifier, a representation of the protocol description, and a QoS profile comprising a set of QoS requirements of the plurality of multiplexed data flows.

[0005] There is further provided a processor for multimodal network traffic differentiation of application data, the processor comprising: at least one controller coupled with at least one memory and configured to cause the processor to: generate a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements; multiplex the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description; and send a request for an application session with network traffic differentiation to a server, the request comprising: a common multimodal session identifier; a representation of the protocol description; and a QoS profile comprising a set of QoS requirements of the plurality of multiplexed data flows.

[0006] There is further still provided a method performed by a client, the method for multimodal network traffic differentiation of application data, the method comprising:

generating a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements; multiplexing the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description; sending a request for an application session with network traffic differentiation to a server, the request comprising: a common multimodal session identifier, a representation of the protocol description, and a QoS profile comprising a set of QoS requirements of the plurality of multiplexed data flows.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Figure 1 illustrates an example of a wireless communications system in accordance with aspects of the present disclosure.

[0008] Figure 2 illustrates an overview of a core network XRM architecture handling of PDU Sets.

[0009] Figure 3 illustrates an example of a one-byte RTP Header Extension for the marking of PDU Set information.

[0010] Figure 4 illustrates an example of a two-byte RTP Header Extension for the marking of PDU Set information.

[0011] Figure 5 illustrates the situation where two PDU sets of different importance and characteristics are mapped to QoS flows and respectively to DRBs.

[0012] Figure 6 provides an overview of the RTP and RTCP stack.

[0013] Figure 7a illustrates a packet format and header information for an RTP packet.

[0014] Similarly, figure 7b illustrates a packet format and header information for an SRTP packet.

[0015] Figure 8 illustrates an overview of the WebRTC stack.

[0016] Figure 9 illustrates RTP/SRTP header extension format and associated syntax.

[0017] Figure 10 shows a TLS 1.3 based encryption and authentication of a QUIC PDU.

[0018] Figure 11 shows an HTTP/3 datagram used as payload to QUIC datagrams.

[0019] Figure 12 illustrates a general demultiplexing solution and mapping of multimodal media content to one or more separate QoS flows for differentiated traffic in downlink with tunnelling over N6.

[0020] Figure 13 shows a general demultiplexing solution and mapping of multimodal media content to one or more separate QoS flows for differentiated traffic in uplink with tunnelling between UE and PSA UPF (showing QUIC as example).

[0021] Figure 14 illustrates UDP-proxying of multimodal QUIC tunnelling for an XR Video application with three data flows, out of which one RTP video stream marked with PDU Set information, one RTP audio stream and one RTCP control message.

[0022] Figure 15 illustrates an example of a user equipment (UE) 1500 in accordance with aspects of the present disclosure.

[0023] Figure 16 illustrates an example of a processor 1600 in accordance with aspects of the present disclosure.

[0024] Figure 17 illustrates an example of a network equipment (NE) 1700 in accordance with aspects of the present disclosure.

[0025] Figure 18 illustrate a flowchart of a method performed by a UE in accordance with aspects of the present disclosure.

DETAILED DESCRIPTION

[0026] The multiplexing of heterogeneous application data flows under a single network application session (i.e., 5-tuple) comes with some challenges especially at the transport layer where a communication path spans across wireless communications networks. A multiplexed flow of multimedia application data cannot readily be transported over such networks in an optimized way without demultiplexing and understanding potential heterogeneous QoS requirements of the multiplexed data flows.

[0027] Traditionally heterogeneous application data has been served over multiple network application sessions in managed wireless networks (e.g., 4G, 5G) with

differentiated QoS treatment. The managed QoS treatment is jointly meant to cater for the heterogeneous transport needs (i.e., delay budgets, packet error rates) of the different data flows and to optimize resource allocation across core and radio access networks while fulfilling overall application QoS requirements. Consequently, there is a need for improved QoS handling of multiplexed multimedia application data in managed wireless communications networks.

[0028] The solutions presented herein comprise at least one identifier being included in a stream. The identifier could be a stream identifier, or alternatively a range of stream identifiers. A UPF may then distinguish and demultiplex streams based on the identifier(s) and so provide mapping of the streams to suitable QoS flows. Such a solution could be applied even if the content of the streams comprises either encrypted or partially encrypted traffic.

[0029] For each QoS flow of each media of the same IP flow the PCF may trigger QoS monitoring in order to monitor the packet delay of each QoS flow. This is required in case packets from a specific QoS flow are delayed more than a configurable value which can result in bad user experience. In case the packet delay of one QoS flow exceeds the threshold, the PCF may adjust the QoS requirements of the QoS flow to ensure the packet delay meets the packet delay criteria.

[0030] Aspects of the present disclosure are described in the context of a wireless communications system.

[0031] Figure 1 illustrates an example of a wireless communications system 100 in accordance with aspects of the present disclosure. The wireless communications system 100 may include one or more NE 102, one or more UE 104, and a core network (CN) 106. The wireless communications system 100 may support various radio access technologies. In some implementations, the wireless communications system 100 may be a 4G network, such as an LTE network or an LTE-Advanced (LTE-A) network. In some other implementations, the wireless communications system 100 may be a NR network, such as a 5G network, a 5G-Advanced (5G-A) network, or a 5G ultrawideband (5G-UWB) network. In other implementations, the wireless communications system 100 may be a combination of a 4G network and a 5G network, or other suitable radio access technology including

Institute of Electrical and Electronics Engineers (IEEE) 802.11 (Wi-Fi), IEEE 802.16 (WiMAX), IEEE 802.20. The wireless communications system 100 may support radio access technologies beyond 5G, for example, 6G. Additionally, the wireless communications system 100 may support technologies, such as time division multiple access (TDMA), frequency division multiple access (FDMA), or code division multiple access (CDMA), etc.

[0032] The one or more NE 102 may be dispersed throughout a geographic region to form the wireless communications system 100. One or more of the NE 102 described herein may be or include or may be referred to as a network node, a base station, a network element, a network function, a network entity, a radio access network (RAN), a NodeB, an eNodeB (eNB), a next-generation NodeB (gNB), or other suitable terminology. An NE 102 and a UE 104 may communicate via a communication link, which may be a wireless or wired connection. For example, an NE 102 and a UE 104 may perform wireless communication (e.g., receive signalling, transmit signalling) over a Uu interface.

[0033] An NE 102 may provide a geographic coverage area for which the NE 102 may support services for one or more UEs 104 within the geographic coverage area. For example, an NE 102 and a UE 104 may support wireless communication of signals related to services (e.g., voice, video, packet data, messaging, broadcast, etc.) according to one or multiple radio access technologies. In some implementations, an NE 102 may be moveable, for example, a satellite associated with a non-terrestrial network (NTN). In some implementations, different geographic coverage areas associated with the same or different radio access technologies may overlap, but the different geographic coverage areas may be associated with different NE 102.

[0034] The one or more UE 104 may be dispersed throughout a geographic region of the wireless communications system 100. A UE 104 may include or may be referred to as a remote unit, a mobile device, a wireless device, a remote device, a subscriber device, a transmitter device, a receiver device, or some other suitable terminology. In some implementations, the UE 104 may be referred to as a unit, a station, a terminal, or a client, among other examples. Additionally, or alternatively, the UE 104 may be referred to as an

Internet-of-Things (IoT) device, an Internet-of-Everything (IoE) device, or machine-type communication (MTC) device, among other examples.

[0035] A UE 104 may be able to support wireless communication directly with other UEs 104 over a communication link. For example, a UE 104 may support wireless communication directly with another UE 104 over a device-to-device (D2D) communication link. In some implementations, such as vehicle-to-vehicle (V2V) deployments, vehicle-to-everything (V2X) deployments, or cellular-V2X deployments, the communication link may be referred to as a sidelink. For example, a UE 104 may support wireless communication directly with another UE 104 over a PC5 interface.

[0036] An NE 102 may support communications with the CN 106, or with another NE 102, or both. For example, an NE 102 may interface with other NE 102 or the CN 106 through one or more backhaul links (e.g., S1, N2, N3, or another network interface). In some implementations, the NE 102 may communicate with each other directly. In some other implementations, the NE 102 may communicate with each other indirectly (e.g., via the CN 106. In some implementations, one or more NE 102 may include subcomponents, such as an access network entity, which may be an example of an access node controller (ANC). An ANC may communicate with the one or more UEs 104 through one or more other access network transmission entities, which may be referred to as a radio heads, smart radio heads, or transmission-reception points (TRPs).

[0037] The CN 106 may support user authentication, access authorization, tracking, connectivity, and other access, routing, or mobility functions. The CN 106 may be an evolved packet core (EPC), or a 5G core (5GC), which may include a control plane entity that manages access and mobility (e.g., a mobility management entity (MME), an access and mobility management functions (AMF)) and a user plane entity that routes packets or interconnects to external networks (e.g., a serving gateway (S-GW), a Packet Data Network (PDN) gateway (P-GW), or a user plane function (UPF)). In some implementations, the control plane entity may manage non-access stratum (NAS) functions, such as mobility, authentication, and bearer management (e.g., data bearers, signal bearers, etc.) for the one or more UEs 104 served by the one or more NE 102 associated with the CN 106.

[0038] The CN 106 may communicate with a packet data network over one or more backhaul links (e.g., via an S1, N2, N3, or another network interface). The packet data network may include an application server. In some implementations, one or more UEs 104 may communicate with the application server. A UE 104 may establish a session (e.g., a protocol data unit (PDU) session, or the like) with the CN 106 via an NE 102. The CN 106 may route traffic (e.g., control information, data, and the like) between the UE 104 and the application server using the established session (e.g., the established PDU session). The PDU session may be an example of a logical connection between the UE 104 and the CN 106 (e.g., one or more network functions of the CN 106).

[0039] In the wireless communications system 100, the NEs 102 and the UEs 104 may use resources of the wireless communications system 100 (e.g., time resources (e.g., symbols, slots, subframes, frames, or the like) or frequency resources (e.g., subcarriers, carriers)) to perform various operations (e.g., wireless communications). In some implementations, the NEs 102 and the UEs 104 may support different resource structures. For example, the NEs 102 and the UEs 104 may support different frame structures. In some implementations, such as in 4G, the NEs 102 and the UEs 104 may support a single frame structure. In some other implementations, such as in 5G and among other suitable radio access technologies, the NEs 102 and the UEs 104 may support various frame structures (i.e., multiple frame structures). The NEs 102 and the UEs 104 may support various frame structures based on one or more numerologies.

[0040] One or more numerologies may be supported in the wireless communications system 100, and a numerology may include a subcarrier spacing and a cyclic prefix. A first numerology (e.g., μ=0) may be associated with a first subcarrier spacing (e.g., 15 kHz) and a normal cyclic prefix. In some implementations, the first numerology (e.g., μ=0) associated with the first subcarrier spacing (e.g., 15 kHz) may utilize one slot per subframe. A second numerology (e.g., μ=1) may be associated with a second subcarrier spacing (e.g., 30 kHz) and a normal cyclic prefix. A third numerology (e.g., μ=2) may be associated with a third subcarrier spacing (e.g., 60 kHz) and a normal cyclic prefix or an extended cyclic prefix. A fourth numerology (e.g., μ=3) may be associated with a fourth subcarrier spacing

(e.g., 120 kHz) and a normal cyclic prefix. A fifth numerology (e.g., μ=4) may be associated with a fifth subcarrier spacing (e.g., 240 kHz) and a normal cyclic prefix.

[0041] A time interval of a resource (e.g., a communication resource) may be organized according to frames (also referred to as radio frames). Each frame may have a duration, for example, a 10 millisecond (ms) duration. In some implementations, each frame may include multiple subframes. For example, each frame may include 10 subframes, and each subframe may have a duration, for example, a 1 ms duration. In some implementations, each frame may have the same duration. In some implementations, each subframe of a frame may have the same duration.

[0042] Additionally or alternatively, a time interval of a resource (e.g., a communication resource) may be organized according to slots. For example, a subframe may include a number (e.g., quantity) of slots. The number of slots in each subframe may also depend on the one or more numerologies supported in the wireless communications system 100. For instance, the first, second, third, fourth, and fifth numerologies (i.e., μ=0, μ=1, μ=2, μ=3, μ=4) associated with respective subcarrier spacings of 15 kHz, 30 kHz, 60 kHz, 120 kHz, and 240 kHz may utilize a single slot per subframe, two slots per subframe, four slots per subframe, eight slots per subframe, and 16 slots per subframe, respectively. Each slot may include a number (e.g., quantity) of symbols (e.g., OFDM symbols). In some implementations, the number (e.g., quantity) of slots for a subframe may depend on a numerology. For a normal cyclic prefix, a slot may include 14 symbols. For an extended cyclic prefix (e.g., applicable for 60 kHz subcarrier spacing), a slot may include 12 symbols. The relationship between the number of symbols per slot, the number of slots per subframe, and the number of slots per frame for a normal cyclic prefix and an extended cyclic prefix may depend on a numerology. It should be understood that reference to a first numerology (e.g., μ=0) associated with a first subcarrier spacing (e.g., 15 kHz) may be used interchangeably between subframes and slots.

[0043] In the wireless communications system 100, an electromagnetic (EM) spectrum may be split, based on frequency or wavelength, into various classes, frequency bands, frequency channels, etc. By way of example, the wireless communications system 100 may support one or multiple operating frequency bands, such as frequency range designations

FR1 (410 MHz – 7.125 GHz), FR2 (24.25 GHz – 52.6 GHz), FR3 (7.125 GHz – 24.25 GHz), FR4 (52.6 GHz – 114.25 GHz), FR4a or FR4-1 (52.6 GHz – 71 GHz), and FR5 (114.25 GHz – 300 GHz). In some implementations, the NEs 102 and the UEs 104 may perform wireless communications over one or more of the operating frequency bands. In some implementations, FR1 may be used by the NEs 102 and the UEs 104, among other equipment or devices for cellular communications traffic (e.g., control information, data). In some implementations, FR2 may be used by the NEs 102 and the UEs 104, among other equipment or devices for short-range, high data rate capabilities.

[0044] FR1 may be associated with one or multiple numerologies (e.g., at least three numerologies). For example, FR1 may be associated with a first numerology (e.g., μ=0), which includes 15 kHz subcarrier spacing; a second numerology (e.g., μ=1), which includes 30 kHz subcarrier spacing; and a third numerology (e.g., μ=2), which includes 60 kHz subcarrier spacing. FR2 may be associated with one or multiple numerologies (e.g., at least 2 numerologies). For example, FR2 may be associated with a third numerology (e.g., μ=2), which includes 60 kHz subcarrier spacing; and a fourth numerology (e.g., μ=3), which includes 120 kHz subcarrier spacing.

[0045] The transport of multimedia application data (e.g., AR/VR/XR, interactive and streaming services, web applications) often contains multiplexed data flows (e.g., video, audio, control and feedback metadata, application metadata etc.). The multiplexing of multiple data flows of an application comes with deployment advantages. Such deployment advantages may include reducing the usage of network ports, reducing the implementers overhead in managing different sessions for an application and/or providing common web-friendly APIs with targeted functionalities (e.g., WebRTC, WebTransport).

[0046] Nevertheless, the multiplexing of heterogeneous application data flows under a single network application session (i.e., 5-tuple), comes with some challenges especially at the transport layer where a communication path spans across wireless communications networks. A multiplexed flow of multimedia application data cannot readily be transported over such networks in an optimized way without demultiplexing and understanding potential heterogeneous QoS requirements of the multiplexed data flows.

[0047] Traditionally heterogeneous application data has been served over multiple network application sessions in managed wireless networks (e.g., 4G, 5G) with differentiated QoS treatment. The managed QoS treatment is jointly meant to cater for the heterogeneous transport needs (i.e., delay budgets, packet error rates) of the different data flows and to optimize resource allocation across core and radio access networks while fulfilling overall application QoS requirements. Consequently, there is a need for improved QoS handling of multiplexed multimedia application data in managed wireless communications networks.

[0048] A solution described herein comprises using a new multiplexing or protocol description and then demultiplexing multiplexed data flows of an application session for differentiated optimized QoS handling based on the multiplexing or protocol description.

[0049] Another solution described herein comprises providing a protocol description, the protocol description including a listing of multiplexed data flows and their individual, potentially different, QoS requirements as part of an application-common QoS requirement taking into account the 5GS support of a tunnelling protocol either over N6 between the UPF and an Application Server, or between UE and UPF. The tunnelling protocol provides assistance of identifying media multiplexed on the same IP session.

[0050] It should be noted that solutions and embodiments described herein relate interchangeably to both applications serving one or more users as well as network services (e.g., AR/VR immersive phone calling services).

[0051] The domain of eXtended Reality (XR) comprises a plethora of application and services where multimedia flows are multiplexed under a single network application session. Such a network application session may be defined by a 5-tuple containing a source IP address, a destination IP address, a source network port, a destination network port, and a protocol number as identifier. As an example, an XR application based on WebRTC, or alternatively, on RTP/SRTP protocol stack, may contain one or more of multiple video streams and audio streams multiplexed with control and feedback metadata and application metadata (e.g., user pose information, user input actions etc) over a single application data network session.

[0052] It should be noted that while some of the examples presented herein use XR as a reference use case or family of applications for the solutions defined herein, one skilled in the art should realize that the proposed solutions are generally applicable to other multimedia applications and may be embodied by different types of transport and network protocols stacks (e.g., QUIC, WebRTC, WebTransport or alike).

[0053] Hereafter eXtended Reality (XR) is used as an umbrella term for different types of realities, of which Virtual Reality, Augmented Reality, and Mixed Reality are examples.

[0054] Virtual Reality (VR) is a rendered version of a delivered visual and audio scene. The rendering is in this case designed to mimic the visual and audio sensory stimuli of the real world as naturally as possible to an observer or user as they move within the limits defined by the application. Virtual reality usually, but not necessarily, requires a user to wear a head mounted display (HMD), to completely replace the user's field of view with a simulated visual component, and to wear headphones, to provide the user with the accompanying audio. Some form of head and motion tracking of the user in VR is usually also necessary to allow the simulated visual and audio components to be updated to ensure that, from the user's perspective, items and sound sources remain consistent with the user's movements. In some implementations additional means to interact with the virtual reality simulation may be provided but are not strictly necessary.

[0055] Augmented Reality (AR) is when a user is provided with additional information or artificially generated items, or content overlaid upon their current environment. Such additional information or content will usually be visual and/or audible and their observation of their current environment may be direct, with no intermediate sensing, processing, and rendering, or indirect, where their perception of their environment is relayed via sensors and may be enhanced or processed.

[0056] Mixed Reality (MR) is an advanced form of AR where some virtual elements are inserted into the physical scene with the intent to provide the illusion that these elements are part of the real scene.

[0057] XR refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables. It includes representative

forms such as AR, MR and VR and the areas interpolated among them. The levels of virtuality range from partially sensory inputs to fully immersive VR. In some circles, a key aspect of XR is considered to be the extension of human experiences especially relating to the senses of existence (represented by VR) and the acquisition of cognition (represented by AR).

[0058] The XR Media (XRM) feature in 3GPP Release 18 at the core network (CN) level introduced the concept of a PDU Set to handle QoS requirements of XRM applications and streams with a better granularity beyond 5G Rel-17 QoS flow possibilities. As such, according to 3GPP Technical Report TR 23.700-60 (v0.0.3), a PDU Set is composed of one or more PDUs carrying the payload of one unit of information generated at the application level (e.g. a frame or video slice for XRM Services). In some implementations all PDUs in a PDU Set are needed by the application layer to use the corresponding unit of information. In other implementations, the application layer can still recover parts or all of the information unit, when some PDUs are missing.

[0059] In addition, the PDU Set is associated with QoS requirements in terms of delay budget and error rate, which may be defined as PDU Set Delay Budget (PSDB), and/or as a PDU Set Error Rate (PSER) as defined in 3GPP Technical Report TR 23.700-60 (v0.0.3 – May 2022) titled "Study on XR (Extended Reality) and media services" and 3GPP Technical Specification TS 23.501 (v18.1.0 – April 2023) titled "System architecture for the 5G System (5GS)". The PDU Set Delay Budget (PSDB) defines an upper bound for the time that a PDU Set may be delayed between the UE and the N6 termination point at the UPF. PSDB applies to the DL PDU Set received by the UPF over the N6 interface, and to the UL PDU Set sent by the UE, and respectively. The PDU Set Error Rate (PSER) defines an upper bound for the rate of PDU Sets (e.g. set of IP packets constituting a PDU Set) that have been processed by the sender of a link layer protocol (e.g. RLC in RAN of a 3GPP access). The PSER may be used to determine an upper bound for a rate of non-congestion-related packet losses.

[0060] Figure 2 illustrates an overview of a core network (CN) XRM architecture handling of PDU Sets. Figure 2 shows a system 200 comprising an Extended Reality Media Application Function (XRM AF) 210, a Policy and Control Function (PCF) 215, a

Session Management Function (SMF) 220, an Access and Mobility Function (AMF) 225, a Radio Access Network (RAN) 230, a User Equipment (UE) 235, a User Plane Function (UPF) 240, and an Extended Reality Application 245. The UE 235 may comprise a UE 104, 235, 1235, 1335, 1435, or 1500 as described herein. The UPF 240 may comprise a UPF 1240, 1340, or 1440 as described herein. The operation of system 200 will now be described in the example of downlink traffic, a similar process may operate for uplink traffic.

[0061] At 280, the XRM AF 210 determines PDU Set requirements.

[0062] At 281, the XRM Application Function 210 provides QoS requirements for packets of a PDU Set to the PCF 215 and information to identify the application (i.e. 5-tuple or application id). The QoS requirements may comprise PSDB and PSER. The XRM AF 210 may also include an importance parameter for a PDU Set and information for the core network to identify packets belonging to a PDU Set.

[0063] At 282, the PCF 215 derives QoS rules for the XR application and specific QoS requirements for the PDU Set and configures the SMF 220. The QoS rules may use a 5G QoS identifier (5QI) for XR media traffic. The PCF 215 sends the QoS rules to the SMF 220. The PCF 215 may include in the communication to the SMF 220 PCC rules per importance of a PDU Set. The PCC rules may be derived according to information received from the XRM AF 210 or based on an operator configuration.

[0064] At 283, the SMF 220 establishes a QoS flow according to the QoS rules by the PCF 215 and configures the UPF to route packets of the XR application to a QoS flow, and, in addition, to enable PDU Set handling. The SMF 220 also provides the QoS profile containing PDU Set QoS requirements to the RAN 230 via the AMF 225. The AMF 225 may provide the QoS profile containing PDU Set QoS requirements to the RAN 230 in an N2 SM container. Further, the AMF 225 may provide the QoS rules to the UE 235 in an N1 SM container.

[0065] At 284, the UPF 240 inspects the packets and determines packets belonging to a PDU Set. Such a determination may be based on UPF implementation given, for instance by inspecting the RTP packet headers , as described in 3GPP Technical Specification

23.501 v18.2.2 (Jun 2023) titled “System architecture for the 5G System (5GS)”, or based on AS-marked PDU set information transmitted over RTP PDU Set header extensions, as described in 3GPP TS 23.501 v18.2.2 and 3GPP Technical Specification 26.522 v0.1.1 (Sep 2023) titled “5G Real-time Media Transport Protocol Configurations”, using for instance the PDU Set information RTP header extension urn:3gpp:pdu-set-marking:rel-18. The packet inspection may comprise inspecting the RTP packets. When the UPF 240 detects packets of a PDU Set the UPF 240 marks the packets belonging to a PDU Set within a GTP-U header. The GTP-U header information includes a PDU Set sequence number and the size of the PDU Set. The UPF 240 may also determine the importance of the PDU Set either based on UPF 240 implementation means, information provided by the XRM AF 210 or information provided as metadata from an XRM application server. Based on the importance of the PDU Set the UPF 240 may route the traffic to a corresponding QoS flow 1 (according to the rules received from the SMF 220) or include the importance of the PDU Set within a GTP-U header. QoS flow 1 may comprise GTP-U headers, and these may include PDU Set information.

[0066] At 285, the RAN 230 identifies packets belonging to a PDU Set (based on the GTP-U marking) and handles the packets of the PDU Set according to the QoS requirements of the PDU Set provided by the SMF 220. In one implementation the RAN 230 node may use a different radio bearer with higher QoS requirement (according to the PDU Set PSDB/PSER) to guarantee delivery of the packets of the PDU Set, while using a different radio bearer according to the 5QI of the QoS flow for the non-PDU Set packets.

[0067] The above example relates to downlink (DL) traffic. Reciprocal processing is applicable to uplink (UL) traffic wherein the role of UPF 240 packet inspection is taken by the UE 235 which is expected to inspect uplink packets, determine packets belonging to a PDU Set, and signal accordingly the PDU Set to the RAN 230 for scheduling and resource allocation corresponding to an associated DRB capable of fulfilling the PDU Set QoS requirements (i.e., PSDB and PSER). The low-level signalling mechanism associated with the UL UE-to-RAN information passing are up to the specification and implementations of RAN signalling procedures.

[0068] However, in XRM Release 18, e.g. 3GPP Technical Specification TS 23.501 v18.2.2 (Jun 2023) titled "System architecture for the 5G System (5GS)", once the PDU set QoS integrated handling is enabled, the PSA UPF identifies PDUs that belong to PDU sets and determines for each PDU set the PDU set information below sent over to the NG-RAN in the GTP-U header.

[0069] 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.

[0070] The PDU Set information is then used by the NG-RAN for PDU Set based QoS handling as described above.

[0071] The NG-RAN may use Priority Levels as of across QoS Flows and PDU Set Importance within a QoS Flow for PDU Set level packet discarding in presence of congestion. Such Priority Levels are described in 3GPP TS 23.501 v18.2.2 clause 5.7.3.3.

[0072] It is also specified in 3GPP TS 23.501 v18.2.2 that the PSA UPF identifies PDUs that belong to PDU Sets and if the UPF receives a PDU that does not belong to a PDU Set based on Protocol Description for PDU Set identification (e.g., has not been marked with PDU Set information by the AS), then the UPF still maps the PDU to a PDU Set and determines the PDU Set Information as described above. This ensures that for a QoS flow with PDU Set enabled all the PDUs belong to a PDU Set. To this end, if the PSA UPF receives a PDU that does not belong to a PDU Set, it is assumed that the UPF determines the PDU Set Importance value based in some examples on pre-configuration and in other examples on an AS/AF signalled default importance.

[0073] The AS PDU Set information listed above may be provided via a one-byte RTP Header Extension for the marking of PDU Sets, and End of Bursts as defined by 3GPP TS 26.522 v0.1.1 (Sep 2023) titled "5G Real-time Media Transport Protocol Configurations".

[0074] In some implementations, a 5GS comprises the PDU Set marking by the AS and its associated information, as previously detailed, in RTP header extensions according to the IETF RFC 8285. This can be either in one-byte or two-byte formats as displayed in Figures 3 and Figure 4.

[0075] Figure 3 illustrates an example of a one-byte RTP Header Extension 300 for the marking of PDU Set information. Figure 4 illustrates an example of a two-byte RTP Header Extension 400 for the marking of PDU Set information.

[0076] The semantics associated with the PDU Set information fields of the RTP header extension syntaxes outlined in Figure 3 and Figure 4 are as follows.

[0077] "E" 332, 432 is a 1 bit representation of a Boolean flag that shall be set to 1 for the last PDU of the PDU Set and set to 0 for all other PDUs of the PDU Set.

[0078] "EDB" 334, 434 is a 3 bit representation of the end of a Data Burst indication. The 3 bits may encode for example the End of Data Burst indication as per the encoding and guidelines provided in Clause 4.4.2.6.1 of the 3GPP Technical Specification TS 26.522 v0.1.1 (Sep 2023) titled "5G Real-time Media Transport Protocol Configurations".

[0079] "PSI" 336, 436 is a 4 bit representation indicating the importance of a PDU Set compared to other PDU Sets within the same RTP stream, or alternatively, session. Lower values shall indicate a higher importance PDU Set, with the highest importance PDU Set indicated by 0 and the lowest importance PDU Set indicated by 15. This field is applicable for various audio/video codecs supported by a 5GS (e.g., H.264, H.265, HE-AAC etc.).

[0080] "PSSN" 340, 440 comprises a 10 bit representation encoding the sequence number of the PDU Set to which the current PDU belongs acting as a 10-bit numerical identifier for the PDU Set. PSSN may take a value between 0 and 1023, and even though the value may wrap around 1023 in some examples, a receiver (e.g., UPF) may uniquely distinguish between any PDU Sets using a combination of the RTP packet sequence number and the PSSN.

[0081] "PSN" 342, 442 is a 6 bit representation of the sequence number of the current PDU within the PDU Set. The PSN shall be set to 0 for the first PDU in the PDU Set and incremented monotonically for every PDU in the PDU Set in order of transmission from the

sender. A receiver (e.g., UPF, or alternatively, in some examples a gNB) may use the RTP packet sequence number together with the PSN to distinguish between PDUs within a PDU Set that contains more than 64 PDUs.

[0082] "PSSize" 344, 444 is a 24 bit representation indicating the total size of all PDUs of the PDU Set to which this PDU belongs. This field is optional and subject to an SDP signalling offer/answer negotiation, where the AS may indicate whether it will be able to provide the size of the PDU Set for an RTP stream, or alternatively, an RTP session. If in some embodiments it is not enabled, the field will not be present. However, if in some embodiments is enabled, but the AS is not able to determine the PDU Size for a particular PDU Set, the AS will set the value to 0 in all PDUs of that PDU Set. The PSSize indicates in some implementations the size of a PDU Set including RTP/UDP/IP header encapsulation overhead of its corresponding PDUs. Alternatively, the PSSize indicates the sum of RTP payload sizes of all PDUs present in a PDU Set. The PSSize is expressed in bytes. As this field is optionally appended to the RTP header extensions, its presence is negotiated and signalled via SDP offer/answer procedure by means of the "pdu-set-size" extension attribute, as per example:

```
a=extmap:1 sendonly urn:3gpp:pdu-set-marking:rel-18 pdu-set-size
```

[0083] Reciprocal processing is applicable to UL whereas the role of UPF packet inspection is taken by the UE which is expected to inspect packets, determine packets belonging to a PDU set, and signal accordingly the PDU set to the RAN for scheduling and resource allocation corresponding to an associated DRB capable of fulfilling the PDU set QoS requirements (i.e., PSDB and PSER). The low-level signalling mechanism associated with the UL UE-to-RAN information passing are up to the specification and implementations of RAN signalling procedures.

[0084] Depending on the QoS flow mappings and RAN procedures, several alternative PDU set to QoS flow to DRB mappings are possible given two distinct PDU sets with different PDU set attributes, such as PDU set importance. Figure 5 illustrates this where two PDU sets of different importance and characteristics are being mapped to QoS flows and respectively to DRBs. Consider in this example PDU set 1 to be of high importance

with strict QoS requirements (i.e., PSDB, PSER etc.) and PDU set 2 to be of low importance with potentially lower QoS requirements (i.e., PSDB, PSER etc.) than PDU set 1. As illustrated in Figure 5, the PDU set to QoS flow to DRB can take the following instantiations depending on QoS flow policies and Layer 2 RAN procedures.

[0085] Figures 5a to 5d illustrate 5GS PDU Set-aware QoS handling framework description of PDU Set to QoS flow to DRB mappings. Depending on the QoS flow mappings and RAN procedures, several alternative PDU Set to QoS flow to DRB mappings are possible given two distinct PDU Sets with different PDU Set attributes, such as PDU Set importance. Figure 5 illustrates some options where two PDU Sets 510 of different importance and characteristics are mapped to QoS flows 520 and respectively to Data Radio Bearers (DRBs) 530. Consider in this example PDU Set 1 to be of high importance with strict QoS requirements (i.e., PSDB, PSER etc.) and PDU Set 2 to be of low importance with potentially lower QoS requirements (i.e., PSDB, PSER etc.) than PDU Set 1. As illustrated in Figure 5, the PDU Set 510 to QoS flow 520 to DRB 530 can take the following instantiations depending on QoS flow policies and Layer 2 RAN procedures:

[0086] Figure 5a illustrates 1-to-1-to-1 mapping: whereby the separation of QoS flows 520 and DRBs 530 is complete between high and low importance PDU Sets 510 optimizing finely the radio and network resources on a per PDU Set basis.

[0087] Figure 5b illustrates M-to-M-to-1 mapping: whereby the separation between high and low importance PDU Sets 510 is performed only at QoS flow level, whereas the same DRB 530 is used for the over-the-air transmission of both PDU Sets 510, which may lead to overprovisioning of radio resources for low importance PDU Sets 510 yet require a lower overhead of RAN complexity and management.

[0088] Figure 5c illustrates M-to-1-to-1 mapping: whereby there is no separation between the QoS flows 520 and DRBs 530 of different importance PDU Sets 510 and the higher importance PDU Set QoS requirements are prioritized in handling the QoS management across both CN and RAN; this may lead to overprovisioning of resources for low importance PDU Sets 510 in both CN and RAN implementations but requires lower overhead and control within the 5GS QoS framework.

[0089] Figure 5d illustrates M-to-1-to-M mapping: whereby there is no separation across the QoS flows 520 between PDU Set importance levels, yet distinct DRBs 530 are used to cater for the individual requirements of the distinct importance levels; this compromises the QoS flow management complexity and uses PDU Set information to filter the PDU Sets 510 on different DRBs 530 in order to better match the QoS requirements at RAN level and optimize resource allocation according to individual PDU Set needs.

[0090] The transport of multiplexed multimedia flows is largely addressed by WebRTC, or alternatively, RTP/SRTP, and QUIC.

[0091] In 3GPP Release 17, SA4 analysed the XR traffic model, as described in 3GPP Technical Report TR 26.926 (v1.3.0 – Jan 2023) titled "Traffic Models and Quality Evaluation Methods for Media and XR Services in 5G Systems", and concluded the QoS requirements in terms of delay budget, data rate and error rate necessary for a satisfactory experience at the application level. These led to 4 additional 5QIs for the 5GS XR QoS flows as delay-critical GBR 5QIs valued 87-90. These are described at 3GPP Technical Specification TS 23.501 (v18.1.0 – April 2023) titled "System architecture for the 5G System (5GS)" and in particular Table 5.7.4-1. The latter are applicable to XR video streams and control metadata necessary to provide immersive and interactive XR experiences.

[0092] XR video traffic is mainly composed of multiple DL/UL video streams of high resolution (e.g., at least 1080p dual-eye buffer usually), frames-per-second (e.g., 60+ fps) and high bandwidth (e.g., usually at least 20-30 Mbps) which needs to be transmitted across a network with minimal delay (typically upper bounded by 15-20 ms) to maintain a reduced end-to-end application round-trip interaction delay. The latter requirements are of critical importance given the XR application dependency on cloud/edge processing (e.g., content downloading, viewport generation and configuration, viewport update, viewport rendering, media encoding/transcoding etc.). Furthermore, XR traffic also includes one or more audio streams, as well as user interaction metadata (e.g., user actions, user inputs, user pose), or alternatively, feedback and control metadata (actions, XR space metadata, as well as transport control and feedback messages over RTCP for instance).

[0093] The traffic of immersive and interactive XR applications as the ones described above require often real-time suited transport architectures and protocols. As part of the latter, the state of art is represented by the Real-time Transport Protocol (RTP, as defined in IETF standard RFC 3550 - RTP: A Transport Protocol for Real-Time Applications), its securely provisioned Secure Real-time Transport Protocol (SRTP, as defined in IETF standard RFC 3711 - The Secure Real-time Transport Protocol), and its web-targeted stack Web Real-Time Communications WebRTC (defined by w3.org in WebRTC 1.0: Real-Time Communication Between Browsers), respectively.

[0094] RTP is a media codec agnostic network protocol with application-layer framing used to deliver multimedia (e.g., audio, video etc.) data in real-time over IP networks. It is used in conjunction with a sister protocol for control, i.e., Real-time Transport Control Protocol (RTCP), to provide end-to-end features such as jitter compensation, packet loss and out-of-order delivery detection, synchronization, and source streams multiplexing.

[0095] Figure 6 provides an overview of the RTP and RTCP stack. An IP layer 605 carries signalling from the data plane 610 and from the control plane 650. The data plane 610 stack comprises functions for a User Datagram Protocol (UDP) 612, RTP 616, RTCP 614, Media codecs 620 and quality control 622. The control plane 650 stack comprises functions for UDP 652, Transmission Control Protocol (TCP) 654, Session Initiation Protocol (SIP) 662 and Session Description Protocol (SDP) 664.

[0096] SRTP is a secured version of RTP, and is defined by the IETF in RFC 3711 “The Secure Real-time Transport Protocol (SRTP)”. SRTP provides encryption (mainly by means of payload confidentiality), message authentication and integrity protection (by means of PDU, i.e., headers and payload, signing), as well as replay attack protection. Similarly to RTP, the SRTP sister protocol is SRTCP. This provides the same functions to its RTCP counterpart. As such, in vanilla SRTP versions, the RTP header information is still accessible but non-modifiable, whereas the payload is encrypted. These security provisions are illustrated in part in Figure 7b. Furthermore, the key exchange and additional security parameters necessary to use SRTP are based upon the Datagram Transport Layer Security (DTLS) key exchange procedure. SRTP is used for these reasons

as the transport protocol for media in the WebRTC stack which ensures secure RTC multimedia communications over web browser interfaces.

[0097] Figure 7a illustrates a packet format and header information for an RTP packet 730 and Figure 7b illustrates a packet format and header information for an SRTP packet 760. The individual fixed header information and complete header information (including header extensions) is discussed for the RTP/SRTP packets further below. It can be noted that the Fixed Header Info comprises "V" 732, 762, "P" 733, 763, "X" 734, 764, "CC" 736, 766, "M" 738, 768, "PT" 740, 770, "Sequence number" 742, 772, "Timestamp" 744, 774, "Synchronization Source (SSRC) identifier" 746, 776, and "Contributing Source (CSRC) identifier" 748, 778.

[0098] Figure 8 illustrates an overview of the WebRTC stack. As illustrated, an IP layer 805 carries signalling from the data plane 810 and the control plane 850. The data plane stack 810 comprises functions for User Datagram Protocol (UDP) 812, Interactive Connectivity Establishment (ICE) 824, Datagram Transport Layer Security (DTLS) 826, SRTP 817, SRTCP 815, media codecs 820, Quality Control 822 and SCTP 828. ICE 824 may use the Session Traversal Utilities for NAT (STUN) protocol and Traversal Using Relays around NAT (TURN) to address real-time media content delivery across heterogeneous networks and NAT rules and firewalls. The SCTP data plane 728 is mainly dedicated as an application data channel and may be non-time critical. The SRTP based stack 817 and elements of control, i.e., SRTCP 815, encoding, i.e., media codecs, and Quality of Service (QoS), i.e., Quality Control, are dedicated to time-critical transport. The control plane 850 stack comprises functions for Transmission Control Protocol (TCP) 854, Transport Layer Security (TLS) 856, Hypertext Transfer Protocol (HTTP) 858, WebSocket 866, Session Initiation Protocol (SIP) 862, Session Description Protocol (SDP) 864, Server-sent Events (SSE) 868 and Extensible Messaging and Presence Protocol (XMPP) 870.

[0099] The data plane is usually established over a single application session (i.e., a 5-tuple) multiplexing the transport of media (i.e., video, audio, haptic media codec payloads), of user plane quality control and feedback based on RTCP messages and of WebRTC data channels (i.e., application data such as user chat messages, notifications, user inputs, user actions, user pose etc.) over SCTP/DTLS.

[0100] The RTP/SRTP fixed header information and complete header information (including header extensions) are defined in both RFC 3550 - RTP: A Transport Protocol for Real-Time Applications (ietf.org) and RFC 3711 - The Secure Real-time Transport Protocol (SRTP) (ietf.org) as follows.

[0101] Fixed Header Info comprises “V” 732, 762, “P” 733, 763, “X” 734, 764, “CC” 736, 766, “M” 738, 768, “PT” 740, 770, “Sequence number” 742, 772, “Timestamp” 744, 774, “Synchronization Source (SSRC) identifier” 746, 776, and “Contributing Source (CSRC) identifier” 748, 778.

[0102] “V” 732, 762 is 2 bits indicating the protocol version used.

[0103] “P” 733, 763 is a 1 bit field indicating that one or more zero-padded octets at the end of the payload are present, whereby, among others, the padding may be necessary for fixed-sized encrypted blocks or for carrying multiple RTP/SRTP packets over lower layer protocols.

[0104] “X” 734, 764 is a 1 bit indicating that the standard fixed RTP/SRTP header will be followed by an RTP header extension usually associated with a particular data/profile that will carry more information about the data (e.g., the frame marking RTP header extension for video data (as defined in IETF RFC 3711 - The Secure Real-time Transport Protocol (SRTP)), or generic RTP header extensions such as the RTP/SRTP extended protocol (as defined by w3.org in WebRTC 1.0: Real-Time Communication Between Browsers)).

[0105] “CC” 736, 766 is a 4 bit field indicating number of contributing media sources (CSRC) that follow the fixed header.

[0106] “M” 738, 768 is 1 bit intended to mark an information frame boundary in the packet stream, whose behaviour is exactly specified by RTP profiles (e.g., H.264, H.265, H.266, AV1 etc.).

[0107] “PT” 740, 770 is 7 bits indicating the payload type, which in case of video profiles is dynamic and negotiated by means of SDP (e.g., 96 for H.264, 97 for H.265, 98 for AV1 etc.).

[0108] "Sequence number" 742, 772 is 16 bits indicating the sequence number which increments by one with each RTP data packet sent over a session.

[0109] "Timestamp" 744, 774 is 32 bits indicating timestamp in ticks of the payload type clock reflecting the sampling instant of the first octet of the RTP data packet (associated for video stream with a video frame), whereas the first timestamp of the first RTP packet is selected at random.

[0110] "Synchronization Source (SSRC) identifier" 746, 776 is a 32 bit field indicating a random identifier for the source of a stream of RTP packets forming a part of the same timing and sequence number space, such that a receiver may group packets based on synchronization source for playback.

[0111] "Contributing Source (CSRC) identifier" 748, 778 is a list of up to 16 CSRC items of 32 bits each given the amount of CSRC mixed by RTP mixers within the current payload as signalled by the CC bits; the list identifies the contributing sources for the payload contained in this packet given the SSRC identifiers of the contributing sources.

[0112] Complete Header Information (incl. header extensions) comprises "RTP header extension" 750, 780.

[0113] "RTP header extension" 750, 780 is a variable length field present if the X bit is marked; the header extension is appended to the RTP fixed header information after the CSRC list if present; the RTP header extension is 32-bit aligned and formed of the following fields:

- A 16-bit extension identifier defined by a profile and usually negotiated and determined via the Session Description Protocol (SDP) signalling mechanism;
- A 16-bit length field describing the extension header length in 32-bits multiples excluding the first 32 bits corresponding to the 16 bits extension identifier and the 16 bits length fields itself; and
- A 32-bit aligned header extension raw data field formatted according to some RTP header extension identifier specified format.

[0114] Figure 9 illustrates RTP/SRTP header extension format and syntax 900. The RTP header extension format and syntax are like the ones of SRTP. A sketch of these is

thus commonly provided as shown in Figure 9. In addition, in both RTP and SRTP only one RTP extension header may be appended to the fixed header information, as defined in IETF RFC 3550 - RTP: A Transport Protocol for Real-Time Applications. However, for both RTP and SRTP extensions to the base protocols exist to allow for multiple RTP header extensions of predetermined types to be appended to the fixed header information of the protocols, as per RFC 8285: A General Mechanism for RTP Header Extensions (rfc-editor.org).

[0115] In some embodiments, RTP header extensions produced at the source may be ignored by the destination endpoints that do not have the knowledge to interpret and process the RTP header extensions transmitted by the source endpoint.

[0116] QUIC is defined in RFC 9000 - QUIC: A UDP-Based Multiplexed and Secure Transport (ietf.org), and is a new transport protocol meant to replace TCP. Originally, QUIC was designed to benefit HTTP/3 applications by avoiding application-layer and transport-layer head-of-line blocking directed at improving the performance of HTTP/2 over TCP. QUIC is a multiplexed and framed protocol implemented in the user space of OSes running atop of UDP transport layer. QUIC is designed to be a connection-based stream-oriented transport with reliable transmission based on stream-aware acknowledgements. A list of basic QUIC frame types registered with IANA is provided by the RFC 9000, Clause 12.4, and replicated for convenience below in Table 1.

Value (Frame Type ID)	Frame Type Name	Specification
0x00	PADDING	[RFC9000, Section 19.1]
0x01	PING	[RFC9000, Section 19.2]
0x02-0x03	ACK	[RFC9000, Section 19.3]
0x04	RESET_STREAM	[RFC9000, Section 19.4]
0x05	STOP_SENDING	[RFC9000, Section 19.5]
0x06	CRYPTO	[RFC9000, Section 19.6]
0x07	NEW_TOKEN	[RFC9000, Section 19.7]
0x08-0x0f	STREAM	[RFC9000, Section 19.8]
0x10	MAX_DATA	[RFC9000, Section 19.9]
0x11	MAX_STREAM_DATA	[RFC9000, Section 19.10]
0x12-0x13	MAX_STREAMS	[RFC9000, Section 19.11]
0x14	DATA_BLOCKED	[RFC9000, Section 19.12]
0x15	STREAM_DATA_BLOCKED	[RFC9000, Section 19.13]
0x16-0x17	STREAMS_BLOCKED	[RFC9000, Section 19.14]
0x18	NEW_CONNECTION_ID	[RFC9000, Section 19.15]

0x19	RETIRE_CONNECTION_ID	[RFC9000, Section 19.16]
0x1a	PATH_CHALLENGE	[RFC9000, Section 19.17]
0x1b	PATH_RESPONSE	[RFC9000, Section 19.18]
0x1c-0x1d	CONNECTION_CLOSE	[RFC9000, Section 19.19]
0x1e	HANDSHAKE_DONE	[RFC9000, Section 19.20]
0x30-0x31	DATAGRAM	[RFC9221]

Table 1: QUIC frame types according to RFC 9000 and RFC 9221.

[0117] Furthermore, QUIC has connection- and stream-aware flow control, and congestion control transport procedures for paced Internet transport capable of establishing and exploiting the available bandwidth over a connection. QUIC also relies on TLS 1.3 as per RFC 9001 - Using TLS to Secure QUIC (ietf.org), to encrypt all non-essential protocol elements, including large portions of QUIC headers and QUIC frames (including QUIC frames' headers), and respectively, to authenticate through a QUIC authentication message (QUIC mac) the critical QUIC headers (i.e., the header form, fixed bit, spin bit, and the QUIC destination connection ID header fields) alongside the encrypted content. A perspective of the encrypted and authenticated content of a QUIC packet is displayed in Figure 10, whereby the QUIC payload is embodied to carry an ACK frame, a flow control frame, and a stream frame corresponding to HTTP data.

[0118] Figure 10 shows a TLS 1.3 based encryption and authentication of a QUIC PDU 1000. The PDU 1000 comprises A UDP packet header comprising a source port 1002, a destination port 1004, a checksum and length 1006. The PDU 1000 further comprises a UDP payload 1008. The UDP payload 1008 comprises a QUIC packet header comprising flags 1010, a connection ID 1012, and a packet number 1014. The UDP payload 1008 further comprises a QUIC payload 1016 and a QUIC mac 1018. The QUIC payload 1016 comprises an ACK frame 1020, a flow control frame 1022, a stream frame header 1024 and a stream frame payload 1026 which may include HTTP data 1028.

[0119] Besides the reliable transport mode based on stream frames, QUIC also offers a datagram frame extension, i.e., 0x30-0x31 frame type defined in RFC 9221. The QUIC datagrams may be IDed and mapped to different logic flows but this is up to the application layers rather than the QUIC protocol. In essence, datagrams are unreliable data containers with minimum overhead, i.e., a header indicating just the frame type (0x30, 0x31) and

optionally the frame length (i.e., for datagrams of type 0x31) to allow for datagrams multiplexing with other QUIC frames. Furthermore, datagrams frames count towards flow-control, yet a receiver may drop them in case it cannot accommodate a sender's rate, as datagrams frames are inherently unreliable. Nonetheless, datagrams are acknowledged yet their ACKs do not require retransmission by a sender in case of lost, or alternatively, dropped QUIC packets and as such multiplexing datagrams and stream frames over the same QUIC packet is generally not a good sender strategy. However, datagrams are object to congestion control by the sender and as such they may be impacted by a sender reacting to congestion events which may impact delaying datagram sending up to congestion controller decision or to an optional maximum sending time window upon which the expiration of a datagram may happen as configured by an application exercising paced transmissions.

[0120] Albeit QUIC was developed at start to benefit HTTP application-layer protocol, it has evolved over the last 5 years into a fully-fledged general transport protocol that can serve different applications and associated protocols besides the web. For instance, QUIC has been regarded as an efficient, multiplexed transport for multimedia applications and proxying. To this end, QUIC is studied and evolved further in IETF in the working groups of Media over QUIC (MoQ) and Multiplexed Application Substrate over QUIC Encryption (MASQUE). QUIC based HTTP/3 is seen as a replacement of HTTP/2 for DASH and adaptive low-latency streaming, but also as an end-to-end secure transport mechanism for RTP in RTP over QUIC (RoQ), as per draft-ietf-avtcore-rtp-over-quic-05**.**

[0121] QUIC's multiplexing-aware transport and its datagram frames extension makes it desirable to deploy in proxying scenarios whereby QUIC is used to multiplex heterogeneous flows over a single connection (i.e., corresponding to a single 5-tuple) in an encrypted manner end-to-end. This allows endpoints to cater for traffic multiplexing and potential prioritization, including QoS flow control without any intermediary middleboxes interventions given the end-to-end integrity and confidentiality protection offered out-of-the-box by QUIC.

[0122] QUIC-based proxying relies on the well-established CONNECT HTTP procedure existent for TCP sockets. In MASQUE IETF recently HTTP/3 extensions to

CONNECT over QUIC and UDP has been proposed to allow for UDP proxying as per RFC 9298: Proxying UDP in HTTP (rfc-editor.org), and for IP proxying as per draft-ietf-masque-connect-ip-13 - Proxying IP in HTTP**.** To this end the HTTP/3 capsule framing has been defined to match the concept of datagrams in QUIC as per RFC 9297: HTTP Datagrams and the Capsule Protocol (rfc-editor.org). Defining a HTTP/3 datagram capsule framing allows HTTP/3 datagrams to be encapsulated natively within both QUIC datagrams for high-performance unreliable proxying but also over QUIC streams (in case QUIC datagrams are not available or desirable to a proxying implementation) for reliable proxying, respectively. Coupled with the CONNECT HTTP extensions this enables end-to-end encrypted and multiplexed HTTP/3 CONNECT proxying over QUIC and UDP of TCP, UDP and IP based protocols between any proxy server and a client.

[0123] The HTTP/3 datagram used as payload to QUIC datagrams are outlined in Figure 11. A HTTP/3 datagram has a low overhead consisting of a Quarter Stream ID. The concept of a Stream ID from QUIC streams is reused to be able to differentiate and multiplex among different types of QUIC datagrams and expose multiplexing control to the QUIC endpoints's applications. As such the Quarter Stream ID is a variable-length integer that contains the value of the client-initiated bidirectional flow that this datagram is associated with divided by four. The division by four stems from the fact that HTTP requests are sent on client-initiated bidirectional streams, which would have in an embodiment Stream IDs that are divisible by four. In this embodiment, the largest legal QUIC Stream ID value is 2^62-1, so the largest legal value of the Quarter Stream ID field is 2^60-1 as per RFC 9297: HTTP Datagrams and the Capsule Protocol (rfc-editor.org).

[0124] Figure 11 illustrates a representation of HTTP/3 datagram as payload to QUIC datagram frame in a UDP proxying configuration running HTTP/3 extended CONNECT proxying over QUIC and UDP. In particular, Figure 11 shows a UDP datagram 1100 containing a QUIC packet 1102. The QUIC packet 1102 in turn contains a QUIC DATAGRAM frame 1104. The QUIC DATAGRAM frame 1104 in turn contains an HTTP DATAGRAM 1106. The HTTP DATAGRAM 1106 in turn comprises a Quarter Stream ID 1112 and a UDP datagram payload 1114.

[0125] In some embodiments, as QUIC datagrams may be unavailable as a transport protocol, RFC 9297: HTTP Datagrams and the Capsule Protocol (rfc-editor.org), defines a capsule protocol that can be used. The Capsule Protocol is a sequence of type-length-value tuples out of which HTTP upgrade tokens can choose to use. It allows endpoints to reliably communicate request-related information end-to-end on HTTP request streams, even in the presence of HTTP intermediaries. As such, in some embodiments the Capsule Protocol can be used to exchange HTTP/3 datagrams (i.e., the payload carried in Figure 11), which is necessary when HTTP is running over a transport layer that does not support the QUIC datagram frame. In some embodiments, this can run atop of QUIC stream frames to provide a reliable proxying as HTTP/3 extensions CONNECT proxying over QUIC streams, as per RFC 9297: HTTP Datagrams and the Capsule Protocol (rfc-editor.org).

[0126] Multimodality support in 5GS and AF/AS interface support. Up to end of Release 18 Stage 2 normative work for XRM in 3GPP TS 23.501 v18.2.2 (Jun 2023) titled "System architecture for the 5G System (5GS)" the 5GS specify minimal support for multimodal flows in terms of providing a common multimodal service ID for the PCF to correlate QoS policy and rules for multiple flows over a common PDU Session as per 3GPP TS 23.501 Clause 5.37.2. Yet, no architectural features are in the CN (e.g., at the UPF) or in the RAN for dedicated support of multimodal flows and differentiated traffic support are available.

[0127] Furthermore, from the perspective of 5G multimedia, interactive and immersive architecture, the interface between the AS and the AF, i.e., M3 interface in the context of 5G Media Streaming TS 26.501, and respectively, RTC-3 interface in the context of 5G Real-time Communications TS 26.506, is up to the implementation and out of scope of 5GS within the timeline of current Release 18. Yet, this is set to change given the extension of AS-driven provisioning of 5GS service support regarding dynamic policies, service adaptation, as well as requests for optimized QoS handling of multimedia related applications.

[0128] Multimodal flow support within the 5GS falls also in the latter category whereby AS-assistance information towards the 5GS, such as protocol description of multimodal services and/or applications, is further needed. Currently, minimal

specification support is offered over N5 interface at the PCF via the *Npcf_PolicyAuthorization* service (cf. 3GPP TS 29.514 v18.3.0 (Sep 2023) titled "5G System; Policy Authorization Service; Stage 3"), or alternatively, over N33 interface at the NEF via the *Nnef_AFSessionWithQoS* service (cf. 3GPP TS 29.122 v18.3.0 (Sep 2023) titled "T8 reference point for Northbound APIs") to provide a multimodal service ID, a protocol description limited only to PDU Set marked flows and a list of QoS requirements for multiple IP data flows associated with the same multimodal service ID.

[0129] Gaps are therefore present in supporting multimodal flows served over the same flow whereby at least of the multimodal protocol data flows are encrypted and/or not marked with PDU Set information by the AS.

[0130] Current capabilities: The current behaviour of the 5GS in Release-18 as outlined in Figure 2 does not allow for traffic differentiation, i.e., identification of media sent over the same IP flow where an IP flow corresponds to a 5-tuple containing source and destinations IP address and port numbers. In fact, a QoS flow enabled with the PDU Set feature requires in all embodiments that all PDUs of a session (IP flow) mapped to said QoS flow are mapped to a PDU Set and marked with PDU Set information. As such, all of the PDUs and PDU Sets are treated uniformly based on the QoS flow PSER, PSDB and PSIHI QoS rules configuration given the application requirements.

[0131] Problem of current procedures: This behaviour is simple yet in many cases of multimodal traffic (e.g., XR, cloud gaming or generic interactive media applications) it treats the same heterogeneous flows with different QoS requirements. For instance, in some embodiments, audio, video and metadata of a media session are multiplexed over the same IP flow. Currently the granularity of QoS flow in the 3GPP system is per IP flow where an IP flow corresponds to a 5-tuple containing source and destinations IP address and port numbers. This results in such media sessions to be routed over the 5GS system over the same QoS flow enabled with PDU Set and a delay-critical guaranteed-bit rate satisfying in some examples 10^-4 PSER and 20 ms PSDB for a default maximum data burst of 63kBytes as per 3GPP TS 23.501 v18.2.2 Table 5.7.4-1. While this may be appropriate for video media content, it is an overprovisioning of resources for audio and metadata content which may be better mapped to specific QoS flows. Furthermore, in

some embodiments different media sources may each have their own sampling frequencies (e.g., video 60fps, audio 20ms, haptics/pose 10ms), jitter associated with sampling and encoding. In addition, the media capture devices (e.g., microphones, cameras, haptic gloves, HMDs accelerometers/SLAMs) may not be in sync and as such the generated packets to be transmitted may overlap in a non-periodic manner at a multiplexed inter-flow level, whilst individual data flows (e.g., video, audio) may behave quasi-periodic. To enable the network to take advantage of these multimodal traffic characteristics differentiated traffic treatment within 5GS CN of multiplexed traffic is necessary and preferrable. An additional challenge to differentiated traffic treatment is in many embodiments the end-to-end added security, confidentiality and integrity protection of modern multiplexing protocols and media transport stacks (e.g., SCTP/DTLS and SRTP, SRTCP in WebRTC, TLS1.3-based QUIC encryption).

[0132] Figure 12 illustrates a general demultiplexing solution and mapping of multimodal media content to one or more separate QoS flows for differentiated traffic in DL with tunnelling over N6. This may comprise identifying multiplexed media over the same IP flow via N6, i.e., in a DL direction. Figure 12 shows a system 1200 comprising an application function (AF) 1210, a network exposure function (NEF) 1212, a Policy Control Function (PCF) 1215, a Session Management Function (SMF) 1220, an Access and Mobility Function (AMF) 1225, a Radio Access Network (RAN) 1230, a User Equipment (UE) 1235, a User Plane Function (UPF) 1240, and an Application Server (AS) 1245. The AS 1245 may comprise an Extended Reality Application Server.

[0133] At 1281, the AS 1245 provides a protocol description to an AF 1210 and a request for a specific QoS profile based on the protocol description. The protocol description is a representation of the protocol used on the wire to transport the AS application data. The protocol description includes at least a mapping of each individual media content (e.g., video, audio, haptics, application metadata) to a “stream identifier” (or a range of stream identifiers) corresponding to specific type of media of the IP flow. In one embodiment the identifier may correspond to individual streams encapsulating the media content data over the wire. The specific QoS profile corresponding to the protocol description contains associated QoS requirements for at least one of an application-level

(i.e., a stream-common PDB), and stream-level (i.e., per media content/media type, e.g. PDU Set marked video @60 fps with QoS requirements of PSER at 10^-4, PSDB of at 20 ms and PDU Set Integrated Handling ON) QoS parameters.

[0134] At 1282, the AS 1245 establishes over N6 a tunnelling session allowing the AS to provide information on each multiplexed media within the IP flow. The AS ensures that media of particular media type (based on the protocol description) correspond to the mapping of protocol description to stream identifier (or a range of stream identifiers) as provided in step 1281.

[0135] The AF 1210 optionally validates the AS protocol description and further requests from the PCF 5GS via the NEF, or alternatively directly to the PCF, for AFs in the Trusted Domain, an AS Session with the AS requested QoS profile.

[0136] At 1284, the AF 1210 establishes an AF session towards the 5GS (via NEF 1212). The AF includes in the request information providing a mapping of an identifier (stream identifier) (or range of identifiers) to specific media type where the identifier is associated to specific media information provided over the N6 tunnel.

[0137] The PCF 1215 validates based on available SLAs and capabilities the AF request for creation of an AS session with the QoS requirements. Upon validation, the PCF responds to the AF request with the appropriate status and/or error code (e.g., an HTTP code 200 to indicate success of the request and appropriate AS session configuration by the CN and 5GS). The PCF further derives one or more QoS rules each QoS rule corresponding to one or more multiplexed media within the IP flow based on the information provided by the AF. Each QoS rule, comprises packet detection rules which comprises one or more of 5-tuple information of the media flow, protocol description of the media flow, 5-tuple information of the tunnelled session between the UPF and AS and a stream identifier of the tunnelled session over N6 (e.g. using one or more 5QIs for multimodal media traffic) and stream-specific, or alternatively, application-specific QoS requirements. In some cases the PCF may include a default QoS rules for packets that do not contain specific media type. The default QoS rule may also contain a range of stream identifiers where it is known (based on AF information) that such media/traffic do not need specific QoS treatment (e.g. PDU Set QoS handling). In some embodiments, the

associated QoS flows are related at the CN level by a single multimodal service ID, application ID or 5-tuple descriptor. In some embodiments the PCF may determine QoS requirements for one or more QoS flows that rely on the PDU Set feature with QoS parameters of PSER, PSDB and PSIHI, whereas in other embodiments the PCF may determine QoS requirements for one or more QoS flows based on legacy (i.e., non-PDU Set) PER and PDB rules. The rest of the PCF processing follows regular 5GS QoS flow establishment operations, i.e., SMF configuration procedures with the determined QoS and PCC rules.

[0138] The SMF 1220 establishes a QoS flow according to the QoS rules by the PCF. In some embodiments, the SMF will establish a set of one or more associated QoS flows as per the PCF QoS rules and provides N4 rules to the UPF where the N4 rules include information to route packets to a QoS flow based on the packet detection rules in the QoS flow. In addition, in some arrangements, the SMF will enable PDU set handling for one or more of the associated QoS flows as per the protocol description derived QoS rules. The SMF also provides the QoS profile of the one or more associated QoS flows and their individual QoS requirements to the RAN via the AMF.

[0139] The UPF 1240 applies the N4 rules to demultiplex the AS session corresponding to multimodal media content, identify the target recipient of the media flow by inspecting the information within the tunnelled session (e.g. by inspecting target address included in the QUIC control plane information) and map its streams further onto QoS flows as per the SMF instructions to fulfil the AS original request for an AS Session with specific QoS and differentiated traffic treatment. In one implementation, the UPF may assign a PDU Set media stream (e.g., corresponding to video content that is marked by the AS with PDU Set information) to one QoS flow with PDU Set integrated handling, and one or more streams (with similar PER and/or PDB requirements) to another QoS flow with legacy (i.e., non-PDU Set) treatment. In such an implementation the two QoS flows for a set of associated QoS flows serving the same QoS profile of an application with multimodal media content over a single AS session.

[0140] The RAN 1230 processing of QoS flow mapping to DRBs is similar to the legacy processing. Concretely, for PDU Set marked QoS flows of the one or more

associated QoS flows, the RAN identifies packets belonging to a PDU set (based on the GTP-U marking) and handles the packets of the PDU Set according to the QoS requirements of the PDU Set provided by the SMF. In one implementation, as illustrated in Figure 12, the RAN node may use a different radio bearer with higher QoS requirements (according to the PDU set PSDB/PSER requirements) to guarantee delivery of the packets marked with PDU Set information corresponding to one of the associated QoS flows, while using a different radio bearer according to a 5QI for non-PDU Set marked packets corresponding to another associated QoS flow of the set of one or more associated QoS flows.

[0141] The solution described herein assumes that the UPF can distinguish and demultiplex streams based on an identifier (e.g., a stream identifier, or alternatively a range of stream identifiers) and a provided mapping to QoS flows even in the context of encrypted or partially encrypted traffic. Detailed implementation and embodiments describing this behaviour are described below.

[0142] For each QoS flow of each media of the same IP flow the PCF may trigger QoS monitoring in order to monitor the packet delay of each QoS flow. This is required in case packets from a specific QoS flow are delayed more than a configurable value which can result in bad user experience. In case the packet delay of one QoS flow exceeds the threshold, the PCF may adjust the QoS requirements of the QoS flow to ensure the packet delay meets the packet delay criteria.

[0143] Accordingly, there is a need to identify multiplexed media over the same IP flow via a user plane protocol between UE and PSA UPF, i.e., in UL direction. A solution proposed herein is illustrated in Figure 13. Figure 13 shows a general demultiplexing solution and mapping of multimodal media content to one or more separate QoS flows for differentiated traffic in UL with tunnelling between UE and PSA UPF (showing QUIC as example).

[0144] Figure 13 shows a system 1300 comprising an Extended Reality Media Application Function (XRM AF) 1310, a Policy and Control Function (PCF) 1315, a Session Management Function (SMF) 1320, an Access and Mobility Function (AMF) 1325, a Radio Access Network (RAN) 1330, a User Equipment (UE) 1335, a User Plane

Function (UPF) 1340, and an Extended Reality Video Application 1345. The UE 1335 comprises a QUIC proxy client, which is an example of a tunnelling client. The UPF 1340 comprises a QUIC proxy server, which is an example of a tunnelling server. The UE 1335 runs a local XR video application 1337.

[0145] In Figure 13 video data is illustrated with cross hatching; audio data is indicated by diagonal lines from top left to bottom right; and metadata is illustrated with dappling or dots. A first QoS flow between the UPF 1340 and the RAN 1330 has PSDB/PSER requirements. A second QoS flow between the UPF 1340 and the RAN 1330 has legacy QoS requirements, which may comprise PDB/PER requirements.

[0146] At 1381, the AF 1310 provides configuration information to the UE 1335 where the configuration information includes "QUIC MASQUE rules" that contain an application identifier and rules to map media of an IP flow of the application to a specific stream over a tunnelling protocol (e.g. using protocols defined in IETF MASQUE group) between the UE and the PSA UPF. The rules may include information to map media corresponding to a specific protocol over a separate stream. For example, media of video stream requiring PDU Set handling should be sent over a separate stream compared to other media. The rules may also include a range of stream identifiers that should be allocated for the stream corresponding to the protocol description. The rules may also include a default stream identifier for any packets where the protocol cannot be identified (default stream). The configuration information may comprise: a Protocol Description (listing of multiplexed data flows and corresponding protocols); a Mapping of Quarter Stream ID (to data flows such as payload type, subprotocol); and/or a Mapping of Quarter Stream ID (or Quarter Stream ID range) to data flows QoS requirements.

[0147] At 1382, at the same time the AF 1310 has provided towards the 5GS (via NEF and PCF 1315) information providing for each media in the IP flow one or more of the following: a protocol description, QoS requirements, range of stream identifiers. These may comprise PDU Set requirements and may be defined by a stream ID range with respective QoS requirements.

[0148] At 1383, the PCF 1315 based on SLA and subscription information derives one or more QoS rules each QoS rule corresponding to one or more of the multiplexed media

within the IP flow based on the information provided by the AF. By way of example, the QoS rules may include PDU Set related QoS requirements for a 5-tuple of the protocol session between the UE and UPF. Each QoS rule comprises packet detection rules with one or more of 5-tuple information of the media flow within the tunnelled session, protocol description of the media flow, 5-tuple information of the tunnelled session between the UE and UPF and a stream identifier (i.e. using one or more 5QIs for multimodal media traffic) and stream-specific, or alternatively, application-specific QoS requirements. By way of example, the associated QoS flows are related at the CN level by a single service ID, application ID or 5-tuple descriptor. The PCF may determine QoS requirements for one or more QoS flows that rely on the PDU Set feature with QoS parameters of PSER, PSDB and PSIHI. The PCF may determine QoS requirements for one or more QoS flows based on legacy (i.e., non-PDU Set) PER and PDB rules. The rest of the PCF processing follows regular 5GS QoS flow establishment operations, i.e., SMF configuration procedures with the determined QoS and PCC rules.

[0149] At 1383, the SMF 1320 establishes a QoS flow according to the QoS rules provided by the PCF 1315. The SMF 1320 may provide QoS rules to the UE where the QoS rule may include packet detection information on how to route uplink packets to QoS flows based on protocol description of the media and/or stream identifiers. For the downlink the SMF will establish a set of one or more associated QoS flows as per the PCF QoS rules and provides N4 rules to the UPF where the N4 rules include information to route packets to a QoS flow based on 5-tuple and protocol description. The UPF 1340 may also receive QUIC rules that are provided separately or within the N4 rules the QUIC rules indicating how to route the packet (i.e. over which stream identifier) to the UE in the downlink based on the protocol description and stream identifier. If the UPF cannot identify the protocol the UPF routes the packet over the default stream if the UE has established a default stream. For example, the SMF 1320 may provide a QoS profile to the AMF 1325. The QoS profile may define QoS requirements in relation to QoS Flow 1 and QoS Flow 2, and may be associated QoS flows with the same Application Identity or Service Identity.

[0150] When the UE 1335 registers to 5GS the UE 1335 establishes a QUIC session with a UPF 1340 supporting a QUIC proxy. When the local XR video application 1337 requests a network connection, the UE 1335 identifies media of the local XR video application 1337 according to the "QUIC MASQUE rules" and sends the media over separate streams to the UPF 1340. The UE 1335 encapsulates the media within a QUIC connection and allocates a stream identifier based on the QUIC MASQUE rules. The UE uses the uplink QoS rules to convey the QUIC stream over a QoS flow to the UPF corresponding to the QoS rule. The UPF de-encapsulates the QUIC packet and routes the contents to the end server based on the target address included in the QUIC control plane information.

[0151] At 1385, the UPF 1340 upon reception of downlink packets checks the N4 rules. If the N4 rule include protocol description and MASQUE handling rules (or there are associated MASQUE handling rules) the UPF routes the downlink packet to the stream according to the protocol description and/or stream identifier configuration. The processing at the UPF 1340 may be based on Quarter Stream ID in QUIC tunnel id data flow #1 as video traffic and send over QoS flow #1 with PDU Set QoS requirements.

[0152] The RAN 1330 processing of QoS flow mapping to DRBs is similar to the legacy processing. Concretely, for PDU Set marked QoS flows of the one or more associated QoS flows, the RAN identifies packets belonging to a PDU set (based on the GTP-U marking) and handles the packets of the PDU Set according to the QoS requirements of the PDU Set provided by the SMF. In one implementation, as illustrated in Figure 12, the RAN node may use a different radio bearer with higher QoS requirements (according to the PDU set PSDB/PSER requirements) to guarantee delivery of the packets marked with PDU Set information corresponding to one of the associated QoS flows, while using a different radio bearer according to a 5QI for non-PDU Set marked packets corresponding to another associated QoS flow of the set of one or more associated QoS flows.

[0153] The identification of multiplexed media over a single IP flow may be combined to comprise both proposed solutions illustrated in Figure 12 and Figure 13, i.e., in both UL and DL directions.

[0154] QUIC tunnelling may be enabled between the AS and the UPF. In such arrangements, the UPF acts as a QUIC proxy tunnelling server and the AS acts as a QUIC proxy tunnelling client.

[0155] QUIC tunnelling may be enabled between the UE and the UPF. In such arrangement, the UPF acts as a QUIC proxy tunnelling server and the UE acts as a QUIC proxy tunnelling client.

[0156] The solutions described herein are applicable to both AS-to-UPF QUIC tunnelling, or alternatively, UE-to-UPF QUIC tunnelling, may rely on HTTP extended CONNECT over QUIC and UDP proxying and as such in one implementation the QUIC tunnelling may be established to proxy a UDP connection as per RFC 9298 based on "connect-udp" protocol switching, in another implementation the QUIC tunnelling may be established to proxy a TCP connection as for instance per RFC 2817 and in some other implementation the QUIC tunnelling may be established to proxy an IP connection as per IETF draft specification draft-ietf-masque-connect-ip-13 titled "Proxying IP in HTTP" based on "connect-ip" protocol switching. The tunnel establishment via CONNECT HTTP method may rely in some embodiments on legacy CONNECT method of HTTP/1.1, cf. Section 9.3.6. of RFC 9110, enhanced optionally with upgrade headers, cf. Section 7.8 of RFC 9110, for support of UDP and IP tunnelling. The tunnel establishment may rely on extended CONNECT of HTTP/2, cf. RFC 8441, or it may rely on extended CONNECT of HTTP/3, cf. RFC 9220.

[0157] The HTTP extended CONNECT, e.g., for HTTP/2 or HTTP/3, over QUIC and UDP proxying request from the AS (i.e., the client) to a 3GPP UPF proxy server implementation is for example listed below and establishes a UDP tunnelling to a remote UE endpoint with target host name 100.188.67.134 and target port 443 via the connect-udp protocol as per RFC 9298.

```
HEADERS
:method = CONNECT
:protocol = connect-udp
:scheme = https
```

```
:path = /.well-known/masque/udp/100.188.67.134/443/
:authority = 3gpp-upf-proxy.example.org
capsule-protocol = ?1
```

[0158] The host name may be referenced by name instead of an IPv4 or IPv6 address and this will require an UPF proxy server to perform a DNS resolution. In some embodiments, the AS, or alternatively, in other embodiments the UE, may resort to using a reverse DNS lookup provided by application-level controlled logic to obtain the IPv4 or IPv6 of target host UE, or alternatively, AS. Once the optional DNS lookup is performed and the path to the target host is resolved, the UPF proxy server may reply with success in some embodiments or with an error message in other embodiments depending on the requested configuration and available UPF proxy capabilities. The reply status is determined by the server based on the client request and server capabilities. For instance, a client (an AS, or alternatively, a UE) may request (e.g., via HTTP extended CONNECT headers) a server (a PSA UPF) for a network configuration for the client-server tunnel (e.g., QUIC tunnel). The network configuration may comprise of a flow-control configuration (e.g., enforcing stream-level/datagram-level QUIC flow-control) may be based on a QoS profile or QoS requirements corresponding to client multimodal application data (the multiplexed single data flows comprising a multimodal IP flow of, e.g., audio, video, haptics, application metadata etc.). In some arrangements, such client requests may comprise the “QUIC MASQUE rules”.

[0159] For instance, in one implementation, the UPF proxy server replies with a HTTP success status 200 to indicate to the request of an AS proxy client as listed below.

```
HEADERS
:status = 200
capsule-protocol = ?1
```

[0160] Where tunnelled transport is not possible over QUIC datagram frames, (e.g., HTTP1/1, HTTP/2 or HTTP/3 without QUIC datagram frame support), the Capsule

Protocol of RFC 9297 is used to transport the tunnelled UDP payloads in HTTP Datagram Capsules. Alternatively, the Capsule Protocol of RFC 9297 may be used to further transport other types of capsules, e.g., as for IP tunnelling capsule types of ADDRESS_REQUEST, ADDRESS_ASSIGN, ROUTE_ADVERTISEMENT for instance. The Capsule Protocol provides to this end a lightweight serialized stream format of Capsules whereby each Capsule has a type identifier, a length, and a value (i.e., the payload) whose semantics depend on the Capsule type. For instance, a HTTP/3 Datagram capsule format (i.e., the semantics of the Capsule value field of type HTTP/3 Datagram) is listed below as per RFC 9297 and represented graphically in Figure 11.

```
HTTP/3 Datagram {
  Quarter Stream ID (i),
  HTTP Datagram Payload (..),
}
```

[0161] Where HTTP/3 proxying over QUIC datagrams is implemented, the HTTP/3 Datagram Capsules fill as a result the QUIC datagram frame type and an explicit encapsulation of the Capsule type, length is not needed anymore as the QUIC datagram syntax and semantics are reused as outlined in Figure 11.

[0162] As a consequence of the above, by usage of the Capsule Protocol, or alternatively, of QUIC datagram frames for tunnelling proxied traffic, multiplexing of heterogeneous data flows is possible over the same tunnel, whereby the individual data flows are identifiable either by protocol negotiated identifiers, e.g., Quarter Stream ID as based on QUIC datagram payloads, by Quarter Stream ID as based on QUIC streams, or alternatively by Quarter Stream ID as based on Capsule Protocol HTTP Datagram Capsules as per RFC 9297.

[0163] QUIC tunnelling may be used as described above based on HTTP extended CONNECT to proxy UDP traffic between an AS and UE, whereby the QUIC tunnelling is between the AS proxy client and the UPF proxy server. As a result, the QUIC protocol may be used to multiplex multimodal data flows of an AS over a single application session

(i.e., a 5-tuple, or alternatively, IP flow). In addition, the UPF proxy terminating the QUIC tunnel may identify each data flow given their mapping to Quarter Stream IDs. This mapping may be provided by the AS to the AF as a protocol description together with a QoS requirements of the one or more data flows multiplexed by the AS. The AF requests then on the behalf of the AS to the 5GS an application session with differentiated QoS traffic management via control plane signalling. The communication between the AF and 5GS happens over the control plane and in some arrangements, it is between the AF and the NEF, whereas in some other arrangements, where the AF is deployed in the Trusted Domain of a MNO, it is between the AF and the PCF, respectively.

[0164] In some implementations QoS requirements of the multiplexed AS session tunnelled over a QUIC proxy are specific to each data flow and list the QoS PER, PDB requirements of individual data flows. For example, one or more data flows may be each enabled with PDU Set marking by the AS provided for instance over RTP header extensions for PDU Set as per 3GPP TS 26.522 v0.1.1. Thus, the PDU Set integrated QoS treatment is enabled and the QoS requirements are expressed relative to the PSER, PSDB and PSIHI requirements of the individual data flows enabled with PDU Set marking. In some arrangements, besides the individual, data flow-specific QoS requirements, the application flow multiplexing multimodal data flows may further comprise a data flow-common set of QoS requirements, related particularly to a maximum tolerable PER and PDB across all multiplexed data flows. The application may use the maximum tolerable PER and PDB as the processing window in resynchronization of the multimodal streams at a receiver's end to preserve a desired QoE. The specific details of the synchronization processing at the application level are specific to an application domain (e.g., XR streaming/conversational, cloud gaming) and out of the herein.

[0165] The PCF processes the QoS requirements based on available PCC and OAM configuration and determines the set of QoS rules associated with the QoS profile requested by the AF. The QoS rules may be sent to the SMF which determines the corresponding QoS flows and notifies the UPF of the QoS rules and available QoS flows including the mapping of tunnelled data flows to specific QoS flows. The mapping may be based on the

available QoS rules and the data flow identifiers, e.g., Quarter Stream IDs for QUIC datagrams, derived from the protocol description.

[0166] The UPF may terminate the QUIC tunnel and proxies the data flows further to the UE remote endpoint based on the QoS rules and their mapping to data flow identifiers as provided via the N4 interface SM container.

[0167] The AS QUIC tunnelling client connects to the UPF QUIC tunnelling proxy to via a HTTP/3 extended CONNECT over QUIC and UDP. The AS is a multimedia XR application which multiplexes different data flows over QUIC. In the example illustrated in Figure 14 the QUIC tunnel carries at least 3 different data flows corresponding to 3 RTP/RTCP streams, wherein 1 RTP stream marked with PDU Set information carries video payload type (e.g., H.264, H.265, H.266 or alike), 1 RTP stream carries audio (e.g., OPUS, HE-AAC, 3GPP EVS/IVAS etc.) and a the RTCP stream carries RTCP control messages related to the other two RTP streams multiplexed. These streams are identified by Quarter Stream IDs, e.g., QSI = 0 for video, QSI = 1 for audio and QSI = 2 for RTCP control messages. Thus, the UPF maps the RTP video stream to a PDU Set enabled QoS flow with PSDB and PSER video-specific requirements, e.g., PER = 10^-4, PDB = 20 ms and maximum data burst of 120kBytes, whereas the RTP audio and RTCP streams are mapped together on a QoS flow with common PER and PDB rules covering both the requirements of audio and RTCP control messages, e.g., PER = 10^-3, PDB = 10 ms and low-sized maximum data bursts of up to 500 Bytes. The UPF proxies then at the UDP layer the RTP and RTCP data flows to a remote UE with traffic differentiation based on demultiplexing of data flows and their mapping to QoS flows according to corresponding QoS rules matching the requirements of each data flow.

[0168] Figure 14 illustrates UDP-proxying of multimodal QUIC tunnelling for an XR Video application with three data flows, out of which one RTP (or alternatively, SRTP) video stream marked with PDU Set information, one RTP (or alternatively, SRTP) audio stream and one RTCP (or alternatively, SRTCP) control messages.

[0169] Figure 14 shows a system 1400 comprising an Extended Reality Application Function (XR AF) 1410, a Policy and Control Function (PCF) 1415, a Session Management Function (SMF) 1420, an Access and Mobility Function (AMF) 1425, a

Radio Access Network (RAN) 1430, a User Equipment (UE) 1435, a User Plane Function (UPF) 1440, and an Extended Reality Video Application 1445. The XR video application 1445 comprises a QUIC client, which is an example of a tunnelling client. The UPF 1440 comprises a QUIC server, which is an example of a tunnelling server.

[0170] In Figure 14 video data is illustrated with cross hatching; audio data is indicated by diagonal lines from top left to bottom right; and metadata is illustrated with dappling or dots. A first QoS flow between the UPF 1440 and the RAN 1430 has PSDB/PSER requirements. A second QoS flow between the UPF 1440 and the RAN 1430 has legacy QoS requirements, which may comprise PDB/PER requirements.

[0171] The XR video application 1445 sends a protocol description to the XR AF 1410, the protocol description comprising a listing of multiplexed data flows and corresponding protocols. This may include the mapping of Quarter Stream ID to data flows (e.g., payload type, subprotocol), and/or the mapping of a Quarter Stream ID or a Quarter Stream ID range to data flows QoS requirements. Similarly, the SMF 1420 may send an N4 QoS configuration rules to the AMF 1425. The QoS configuration rules may define QoS mapping in relation to QoS Flow 1 and QoS Flow 2 based on a XR video application requested QoS profile. QoS Flow 1 and QoS Flow 2 may be associated QoS flows with the same Application Identity or Service Identity. The SMF 1420 sends to the UPF 1440 stream ID to QoS Flow Indicator mapping information.

[0172] Similar to the above examples in the DL direction, in the UL direction QUIC tunnelling may be used to proxy UDP traffic between the UE 1435 and the AS, in this case the XR video application 1445, whereby the QUIC tunnelling is between the UE proxy client and the UPF proxy server. As a result, the QUIC protocol may be used to multiplex multimodal data flows of a UE over a single application session (i.e., a 5-tuple, or alternatively, IP flow). In these examples, the UE is configured by an AF 1410 with "QUIC MASQUE rules" that contain an application identifier and rules to map media of an IP flow of the application to a specific stream over a tunnelling protocol (e.g., using protocols defined in IETF MASQUE group) between the UE 1435 and the PSA UPF. The configuration rules may include information to map media corresponding to a specific protocol over a separate stream (e.g., media of a video stream with PDU Set handling

should be sent over a separate stream compared to other media), a range of stream identifiers that should be allocated for the stream corresponding to the application's protocol description, and a default stream identifier for any packets where the protocol cannot be identified (i.e., a default stream).

[0173] Additionally, a UE configured with "MASQUE rules" for an application may establish upon registration to 5GS a QUIC tunnelled session with a UPF supporting a QUIC proxy. When the application requests a network connection the UE identifies media of the application according to the "QUIC MASQUE rules" and sends the media over separate streams to the UPF. The UE encapsulates the media within a QUIC connection and allocates a stream identifier (or alternatively a Quarter Stream ID) based on the QUIC MASQUE rules. The UE uses the uplink QoS rules to convey each QUIC stream over a specific QoS flow to the UPF corresponding to each stream associated QoS rule as per SMF configuration. The UPF de-encapsulates the QUIC packet and routes the contents to the end AS, or alternatively, a peer UE, based on the target address included in the QUIC control plane information, e.g., the HTTP extended CONNECT QUIC over UDP proxy establishment.

[0174] The UE may request (e.g., by means of RESTful APIs over HTTP/S) the AF for a specific "MASQUE rules" configuration for an application. The AF processes the request and based on application logic, SLAs and/or ASP policies validates the UE request. Upon successful validation (e.g., by means of an HTTP 2xx status), the UE applies the requested "MASQUE rules".

[0175] The tunnelled data flows as depicted in Figure 14 may implement WebRTC using SRTP instead of RTP, SRTCP instead of RTCP, and providing an additional data channel based on SCTP DTLS over DTLS as part of a single multiplexed application session. In such arrangements, despite using the WebRTC stack, the individual streams can still be mapped to a Quarter Stream ID, or alternatively a Capsule Protocol Context ID.

[0176] In one example, the following mapping may be applied:

- SRTP video stream maps to Quarter Stream ID = 0 which is further allocated by UPF to QoS flow #1 given the QoS rules;

- SRTP audio stream maps to Quarter Stream ID = 1 which is further allocated by UPF to QoS flow #2 given the QoS rules;
- SRTCP control messages map to Quarter Stream ID = 2 which is further allocated by UPF to QoS flow #2 given the QoS rules; and
- SCTP/DTLS WebRTC data channel maps to Quarter Stream ID = 3 which is further allocated by UPF to QoS flow #2 given the QoS rules.

[0177] In other arrangements a WebRTC application session may not be tunnelled over QUIC and instead be sent over the wire without additional encapsulation. In such arrangements the UPF will perform deep packet inspection based on the protocol description representation and available associated QoS rules. To this end, in some implementations the UPF may apply a combination of at least one of the underlying WebRTC protocols, or alternatively referred to as "subprotocols" (e.g., SRTP for media, SRTCP for media session control, SCTP/DTLS for data channel), and the media payload type transported to identify and demultiplex WebRTC data flows and subsequently map them to associated QoS flows with different QoS requirements. In some arrangements, these associated QoS flows will share at least one of a same service ID, application ID, or 5-tuple descriptor across the 5GS QoS flow session management procedures.

[0178] In one example, consider a WebRTC session comprising of:

- two SRTP video stream one of which corresponds to a HD H.264 media source with profile-level-id=64001f and RTP payload type 96, and another one which corresponds to a 720p H.264 media source with profile-level-id=42001f and RTP payload type 97 with H.264,
- a SRTP audio stream with payload type 113,
- one or more SRTCP streams, and
- a SCTP/DTLS data channel.

[0179] The payload type of RTP/SRTP streams may be exposed by the AS to the UPF via control plane signalling. For instance, the AS communicates a protocol description to AF which requests an AS session with QoS requirements and shares the protocol description with the PCF/NEF such that QoS rules are derived to include associated RTP payload types and subprotocol types. In the implementation example above then, the UPF

may first perform deep packet inspection to detect the type of subprotocol encapsulation (e.g., SRTP), and second use the payload type information (e.g., 96 for HD H.264 first source, 97 for 720p H.264 second source, and 113 for audio) to identify and potentially separate the three streams on up to three separate QoS flows given the derived QoS rules. Similarly, by means of deep packet inspection the UPF can identify the other WebRTC subprotocols, i.e., SRTCP and DTLS to separate and treat differently at the QoS flow level the SRTCP and WebRTC data channels as per the application's QoS requirements and PCF derived QoS rules.

[0180] Similar treatment is possible for other multiplexed data flows over a single application session, with or without encryption of payload, or alternatively, payload and partially of header elements, such as:

- RTP and RTCP multiplexed streams (e.g., according to the AVP or AVPF RTP profiles)
- SRTP and SRTCP multiplexed streams (e.g., according to the SAVP or SAVPF RTP profiles)
- SRTP and SRTCP multiplexed streams with additionally encrypted RTP header extensions as defined by either CRYPTEX, i.e., RFC 9335, or RFC 6904.

[0181] By way of further example, it should be noted that the IP flow may be encapsulated within GTP-U. In such an arrangement, the AS establishes a GTP-U tunnel with the UPF over N6. The AS on per packet basis includes within a GTP-U header an identifier corresponding to the media information contained within the packet. As per Figure 12, if the packet corresponds to video stream the AS includes within the GTP-U header ID#0 whereas if the packet corresponds to audio stream the GTP-U header is marked with ID#1.

[0182] The UPF based on the N4 rules received from the SMF routes each received packet to a QoS flow according to 5-tuple and stream identifier information received from the GTP-U packet over N6.

[0183] In yet a further example, to enable the 5GS and the UPF to identify and demultiplex different data flows transported over a common application session for differentiated traffic and QoS management, the AS, or alternatively, an Application Service

Provider provisioning network function may need to expose the components of the underlying transport protocol of the multimodal application data.

[0184] In some arrangements the minimum necessary information to enable an AF to request the establishment of an AF session with QoS support for traffic differentiation for a multimodal protocol flow over a single AS session can based in 5GS at least on:

- a first part of information corresponding to a common multimodal service identifier, or alternatively in other embodiments, a common application/session identifier or a multimodal protocol (e.g., WebRTC) 5-tuple descriptor corresponding to a source IP address, a destination IP address, a source network port, a destination network port, and a protocol flow identifier (e.g., "WebRTC" string identifier).
- a second part of information corresponding to a protocol description containing an indexed list of mappings corresponding to the single modal data flows of the multimodal protocol flow, and
- a third part of information corresponding to an associated indexed list of one or more QoS requirements corresponding to the single modal data flows of the multimodal protocol flow. The third part of information may include a stream identifier (or a range of stream identifiers) corresponding to each media multiplexed over the same IP flow.

[0185] In some embodiments, the protocol description contains the indexed list of mappings between single modal data flows and identifiable protocol elements, wherein the latter may for example be:

- Stream IDs, or respectively, Quarter Stream IDs corresponding to sets of QUIC stream frames and QUIC datagram frames in QUIC based protocols (e.g., connect-udp tunnelling protocols between AS and UPF), or
- (S)RTP payload types, e.g., AVC/HEVC/audio codec profiles, and subprotocol PDUs, e.g., (S)RTCP encapsulated control messages or DTLS encapsulated and encrypted SCTP messages as is the case in WebRTC protocols stack.

[0186] The AS, or alternatively, an ASP, may request over N33 interface to the *Nnef_AFSessionWithQoS* the establishment of the multimodal session with traffic differentiation and QoS optimized treatment. In some embodiments this is based on the

AS/ASP providing the above information as part of the HTTP request with JSON *AsSessionWithQoSSubscription* data type comprising the common multimodal service identifier, and respectively, an indexed list (or alternatively, a hashmap) of *AsSessionMediaComponent* JSON objects. Each of the mapped *AsSessionMediaComponent* JSON objects includes further an integer index, e.g., *medCompN*, (or alternatively a hasmap key), and a protocol description comprising at least one of a MediaType (e.g., as video, audio, haptics, application, text, control, or other string encoded media identifier), a *MediaProtocol* (e.g., (S)RTP, (S)RTPC, RTCP, DTLS, connect-udp/QUIC etc.), a *PayloadType* (e.g., AVC/96, HEVC/97, OPUS/98, AV1/100 etc.), and an array of one or more Stream IDs (e.g., QUIC Stream ID, Quarter Stream ID etc.). Each of the *AsSessionMediaComponent* may further include QoS requirements associated with the correspondingly indexed data flow media component. In some implementations, the QoS requirements may be formed of at least one of a PER, PDB, 5G QoS flow identifier, and an alternative QoS identifier. Alternatively, the QoS requirements may be further based on the PDU Set feature and include PSER, PSDB, PSIHI requirements.

[0187] An AF deployed in the trusted domain requests directly over N5 interface to the PCT *Npcf_PolicyAuthorization* service the same establishment of the multimodal session with traffic differentiation and QoS optimized treatment. In such an arrangement the policy authorization request will contain an indexed list (or alternatively, a hashmap) of *MediaComponent* JSON objects commonly referenced inside the 5GS by a common multimodal service identifier. Furthermore, each of the *MediaComponent* JSON objects includes in an embodiment an integer index, i.e., a *medCompN* (or alternatively a hasmap key), and a protocol description comprising at least one of a MediaType (e.g., as video, audio, haptics, application, text, control, or other string encoded media identifier), a *MediaProtocol* (e.g., (S)RTP, (S)RTPC, RTCP, DTLS, connect-udp/QUIC etc.), a *PayloadType* (e.g., AVC/96, HEVC/97, OPUS/98, AV1/100 etc.), and an array of one or more Stream IDs (e.g., QUIC Stream ID, Quarter Stream ID etc.). In other arrangements, each of the *MediaComponent* may further include QoS requirements associated with the correspondingly indexed data flow media component. In some implementations, the QoS requirements may be formed of at least one of a PER, PDB, 5G QoS flow identifier, and an

alternative QoS identifier. In other implementations, the QoS requirements may be further based on the PDU Set feature and include PSER, PSDB, PSIHI requirements.

[0188] The JSON objects, nodes or arrays referenced in the prequel (i.e., *AsSessionWithQoSSubscription*, *MediaComponent*/*AsSessionMediaComponent*, MediaType, MediaProtocol, PayloadType, StreamID etc.) may be replaced by a different alternative encoding such as XML.

[0189] In a still further example, a UE configuration may be provided for handling traffic over a MASQUE connection between the UE and the UPF. The configuration rules (or QUIC MASQUE rules) may include one or more of the following information:

- Application ID or Traffic Category of the type of traffic initiated by an application;
- A Masque handling indicator instructing the UE to route the packet over a QUIC connection to the PSA UPF;
- Protocol Description corresponding to the type of media (e.g. H.264, audio, metadata); and/or
- A range of stream identifiers to route the media of the QUIC connection.

[0190] The UL QoS rules provided by the PCF to the UE (via the SMF) may include one or more of the following information:

[0191] Packet Detection Rule associating uplink traffic of a QUIC connection containing to a QoS flow. The packet detection rule may include the stream identifier (or range of stream identifiers).

[0192] Accordingly, there is provided a client for multimodal network traffic differentiation of application data, the client comprising: at least one memory; and at least one processor coupled with the at least one memory and configured to cause the client to: generate a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements; multiplex the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description; and send a request for an application session with network traffic differentiation to a server, the request comprising: a common multimodal session identifier,

a representation of the protocol description, and a QoS profile comprising a set of QoS requirements of the plurality of multiplexed data flows.

[0193] Such a client facilitates multimodal network traffic differentiation of application data by way of the protocol description which includes a listing of multiplexed data flows and their individual, potentially different, QoS requirements. The QoS requirements may be part of an application-common set of QoS requirements for differentiated traffic treatment corresponding to a multimodal service over, for example, the 5GS.

[0194] By way of example, the server may comprise an Application Function (AF). The client may comprise a client device. The client may comprise a Application Server (AS) or a User Equipment (UE). The AS/UE cannot interact directly with the UPF. The AS communicates with the UPF according to the following signalling path: AS -> AF -> NEF/PCF -> SMF -> UPF. Alternatively, the UE communicates with the UPF according to the following signalling path: UE <-> AF -> NEF/PCF -> SMF -> UPF.

[0195] The interface between AS -> AF is typically an implementation detail, but in the case of multimedia applications (such as XR and streaming as described herein), at least one media service enabler is likely to be used which will define its own interface in 5G/6G.

[0196] The mapping of a plurality of data flows to identifiable protocol elements may comprise an indexed listing.

[0197] The set of QoS requirements of the plurality of multiplexed data flows may comprise QoS requirements for each of a subset of the plurality of multiplexed data flows, e.g., video data flows. The set of QoS requirements of the plurality of multiplexed data flows may comprise QoS requirements for each of the plurality of multiplexed data flows.

[0198] The identifiable protocol elements may comprise at least one of: a stream protocol frame identifiable by at least one Stream ID; a datagram protocol frame identifiable by at least one Stream ID; a subprotocol identifiable by at least one of a protocol type and a protocol number; and a subprotocol carrying an identifiable media payload type.

[0199] The stream protocol frame identifiable by at least one Stream ID may comprise a QUIC stream frame identifiable by means of a Stream ID, or alternatively, a range/array of Stream IDs.

[0200] The datagram protocol frame identifiable by at least one Stream ID may comprise a QUIC datagram frame identifiable by means of a Quarter Stream ID. The datagram protocol frame identifiable by at least one Stream ID may comprise a range/array of Stream IDs.

[0201] The subprotocol identifiable by at least one of a protocol type and a protocol number may comprise a protocol identifiable according to IANA protocol IDs and protocol syntax applied to the different subprotocols composing the application protocol flow. By way of example, any combination of WebRTC RTP/SRTP, RTCP/SRTCP, DTLS may be multiplexed together on the application WebRTC flow.

[0202] In RTP/SRTP media payload types are visible to the UPF and can be inspected, identified, and mapped to a specific video/audio/haptic codec profile configuration and payload, such as, for example video H.264 baseline profile, H.264 extended profile or H.265 main profile RTP payload types.

[0203] The client may comprise one of an Application Server (AS) and a User Equipment (UE).

[0204] The client may dynamically update at least one part of the request for an application session with network traffic differentiation based on at least one of: a change in the protocol description; and a change in the QoS profile and the set of QoS requirements of the one or more multiplexed multimodal data flows.

[0205] The change in the protocol description may be based on the dynamic termination and reallocation of at least one of the stream identifiers, the protocol identifiers and the RTP media type identifiers.

[0206] The server may comprise an Application Function (AF). Where the client is an AS, the server will be the AF.

[0207] The AF may communicate the request for the application session with network traffic differentiation with at least one of: a Network Exposure Function (NEF), and a Policy Control Function (PCF).

[0208] There is further provided a processor for multimodal network traffic differentiation of application data, the processor comprising: at least one controller coupled with at least one memory and configured to cause the processor to: generate a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements; multiplex the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description; and send a request for an application session with network traffic differentiation to a server, the request comprising: a common multimodal session identifier; a representation of the protocol description; and a QoS profile comprising a set of QoS requirements of the plurality of multiplexed data flows.

[0209] There is further still provided a method performed by a client, the method for multimodal network traffic differentiation of application data, the method comprising: generating a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements; multiplexing the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description; sending a request for an application session with network traffic differentiation to a server, the request comprising: a common multimodal session identifier, a representation of the protocol description, and a QoS profile comprising a set of QoS requirements of the plurality of multiplexed data flows.

[0210] As such, multimodal network traffic differentiation of application data is facilitated by the protocol description which includes a listing of multiplexed data flows and their individual, potentially different, QoS requirements. The QoS requirements may be part of an application-common set of QoS requirements for differentiated traffic treatment corresponding to a multimodal service over, for example, the 5GS.

[0211] By way of example, the server may comprise an Application Function (AF). The client may comprise a client device. The client may comprise a Application Server (AS) or a User Equipment (UE). The AS/UE cannot interact directly with the UPF. The

AS communicates with the UPF according to the following signalling path: AS -> AF -> NEF/PCF -> SMF -> UPF. Alternatively, the UE communicates with the UPF according to the following signalling path: UE <-> AF -> NEF/PCF -> SMF -> UPF.

[0212] The interface between AS -> AF is typically an implementation detail, but in the case of multimedia applications (such as XR and streaming as described herein), at least one media service enabler is likely to be used which will define its own interface in 5G/6G.

[0213] The mapping of a plurality of data flows to identifiable protocol elements may comprise an indexed listing.

[0214] The set of QoS requirements of the plurality of multiplexed data flows may comprise QoS requirements for each of a subset of the plurality of multiplexed data flows, e.g., video data flows. The set of QoS requirements of the plurality of multiplexed data flows may comprise QoS requirements for each of the plurality of multiplexed data flows.

[0215] The identifiable protocol elements comprise at least one of: a stream protocol frame identifiable by at least one Stream ID; a datagram protocol frame identifiable by at least one Stream ID; a subprotocol identifiable by at least one of a protocol type and a protocol number; and a subprotocol carrying an identifiable media payload type.

[0216] The stream protocol frame identifiable by at least one Stream ID may comprise a QUIC stream frame identifiable by means of a Stream ID, or alternatively, a range/array of Stream IDs.

[0217] The datagram protocol frame identifiable by at least one Stream ID may comprise a QUIC datagram frame identifiable by means of a Quarter Stream ID. The datagram protocol frame identifiable by at least one Stream ID may comprise a range/array of Stream IDs.

[0218] The subprotocol identifiable by at least one of a protocol type and a protocol number may comprise a protocol identifiable according to IANA protocol IDs and protocol syntax applied to the different subprotocols composing the application protocol flow. By way of example, any combination of WebRTC RTP/SRTP, RTCP/SRTCP, DTLS may be multiplexed together on the application WebRTC flow.

[0219] In RTP/SRTP media payload types are visible to the UPF and can be inspected, identified, and mapped to a specific video/audio/haptic codec profile configuration and payload, such as, for example video H.264 baseline profile, H.264 extended profile or H.265 main profile RTP payload types.

[0220] The client may comprise one of an Application Server (AS) and a User Equipment (UE).

[0221] The client may dynamically update at least one part of the request for an application session with network traffic differentiation based on at least one of: a change in the protocol description; and a change in the QoS profile and the set of QoS requirements of the one or more multiplexed multimodal data flows.

[0222] The change in the protocol description may be based on the dynamic termination and reallocation of at least one of the stream identifiers, the protocol identifiers, and the RTP media type identifiers.

[0223] The server may comprise an Application Function (AF). Where the client is an AS, the server will be the AF.

[0224] The AF may communicate the request for the application session with network traffic differentiation with at least one of: a Network Exposure Function (NEF); and a Policy Control Function (PCF).

[0225] Figure 15 illustrates an example of a UE 1500 in accordance with aspects of the present disclosure. The UE 1500 may include a processor 1502, a memory 1504, a controller 1506, and a transceiver 1508. The processor 1502, the memory 1504, the controller 1506, or the transceiver 1508, or various components thereof may be examples of means for performing various aspects of the present disclosure as described herein. These components may be coupled (e.g., operatively, communicatively, functionally, electronically, electrically) via one or more interfaces.

[0226] The processor 1502, the memory 1504, the controller 1506, or the transceiver 1508, or various combinations or components thereof may be implemented in hardware (e.g., circuitry). The hardware may include a processor, a digital signal processor (DSP), an application-specific integrated circuit (ASIC), or other programmable logic device, or

any combination thereof configured as or otherwise supporting a means for performing the functions described in the present disclosure.

[0227] The processor 1502 may include an intelligent hardware device (e.g., a general-purpose processor, a DSP, a CPU, an ASIC, an FPGA, or any combination thereof). In some implementations, the processor 1502 may be configured to operate the memory 1504. In some other implementations, the memory 1504 may be integrated into the processor 1502. The processor 1502 may be configured to execute computer-readable instructions stored in the memory 1504 to cause the UE 1500 to perform various functions of the present disclosure.

[0228] The memory 1504 may include volatile or non-volatile memory. The memory 1504 may store computer-readable, computer-executable code including instructions when executed by the processor 1502 cause the UE 1500 to perform various functions described herein. The code may be stored in a non-transitory computer-readable medium such as the memory 1504 or another type of memory. Computer-readable media includes both non-transitory computer storage media and communication media including any medium that facilitates transfer of a computer program from one place to another. A non-transitory storage medium may be any available medium that may be accessed by a general-purpose or special-purpose computer.

[0229] In some implementations, the processor 1502 and the memory 1504 coupled with the processor 1502 may be configured to cause the UE 1500 to perform one or more of the functions described herein (e.g., executing, by the processor 1502, instructions stored in the memory 1504). For example, the processor 1502 may support wireless communication at the UE 1500 in accordance with examples as disclosed herein. The UE 1500 may be configured to support a means for multimodal network traffic differentiation of application data as described herein.

[0230] The controller 1506 may manage input and output signals for the UE 1500. The controller 1506 may also manage peripherals not integrated into the UE 1500. In some implementations, the controller 1506 may utilize an operating system such as iOS®, ANDROID®, WINDOWS®, or other operating systems. In some implementations, the controller 1506 may be implemented as part of the processor 1502.

[0231] In some implementations, the UE 1500 may include at least one transceiver 1508. In some other implementations, the UE 1500 may have more than one transceiver 1508. The transceiver 1508 may represent a wireless transceiver. The transceiver 1508 may include one or more receiver chains 1510, one or more transmitter chains 1512, or a combination thereof.

[0232] A receiver chain 1510 may be configured to receive signals (e.g., control information, data, packets) over a wireless medium. For example, the receiver chain 1510 may include one or more antennas for receiving the signal over the air or wireless medium. The receiver chain 1510 may include at least one amplifier (e.g., a low-noise amplifier (LNA)) configured to amplify the received signal. The receiver chain 1510 may include at least one demodulator configured to demodulate the received signal and obtain the transmitted data by reversing the modulation technique applied during transmission of the signal. The receiver chain 1510 may include at least one decoder for decoding the demodulated signal to receive the transmitted data.

[0233] A transmitter chain 1512 may be configured to generate and transmit signals (e.g., control information, data, packets). The transmitter chain 1512 may include at least one modulator for modulating data onto a carrier signal, preparing the signal for transmission over a wireless medium. The at least one modulator may be configured to support one or more techniques such as amplitude modulation (AM), frequency modulation (FM), or digital modulation schemes like phase-shift keying (PSK) or quadrature amplitude modulation (QAM). The transmitter chain 1512 may also include at least one power amplifier configured to amplify the modulated signal to an appropriate power level suitable for transmission over the wireless medium. The transmitter chain 1512 may also include one or more antennas for transmitting the amplified signal into the air or wireless medium.

[0234] Figure 16 illustrates an example of a processor 1600 in accordance with aspects of the present disclosure. The processor 1600 may be an example of a processor configured to perform various operations in accordance with examples as described herein. The processor 1600 may include a controller 1602 configured to perform various operations in accordance with examples as described herein. The processor 1600 may optionally include at least one memory 1604, which may be, for example, an L1/L2/L3 cache. Additionally,

or alternatively, the processor 1600 may optionally include one or more arithmetic-logic units (ALUs) 1606. One or more of these components may be in electronic communication or otherwise coupled (e.g., operatively, communicatively, functionally, electronically, electrically) via one or more interfaces (e.g., buses).

[0235] The processor 1600 may be a processor chipset and include a protocol stack (e.g., a software stack) executed by the processor chipset to perform various operations (e.g., receiving, obtaining, retrieving, transmitting, outputting, forwarding, storing, determining, identifying, accessing, writing, reading) in accordance with examples as described herein. The processor chipset may include one or more cores, one or more caches (e.g., memory local to or included in the processor chipset (e.g., the processor 1600) or other memory (e.g., random access memory (RAM), read-only memory (ROM), dynamic RAM (DRAM), synchronous dynamic RAM (SDRAM), static RAM (SRAM), ferroelectric RAM (FeRAM), magnetic RAM (MRAM), resistive RAM (RRAM), flash memory, phase change memory (PCM), and others).

[0236] The controller 1602 may be configured to manage and coordinate various operations (e.g., signalling, receiving, obtaining, retrieving, transmitting, outputting, forwarding, storing, determining, identifying, accessing, writing, reading) of the processor 1600 to cause the processor 1600 to support various operations in accordance with examples as described herein. For example, the controller 1602 may operate as a control unit of the processor 1600, generating control signals that manage the operation of various components of the processor 1600. These control signals include enabling or disabling functional units, selecting data paths, initiating memory access, and coordinating timing of operations.

[0237] The controller 1602 may be configured to fetch (e.g., obtain, retrieve, receive) instructions from the memory 1604 and determine subsequent instruction(s) to be executed to cause the processor 1600 to support various operations in accordance with examples as described herein. The controller 1602 may be configured to track memory address of instructions associated with the memory 1604. The controller 1602 may be configured to decode instructions to determine the operation to be performed and the operands involved. For example, the controller 1602 may be configured to interpret the instruction and

determine control signals to be output to other components of the processor 1600 to cause the processor 1600 to support various operations in accordance with examples as described herein. Additionally, or alternatively, the controller 1602 may be configured to manage flow of data within the processor 1600. The controller 1602 may be configured to control transfer of data between registers, arithmetic logic units (ALUs), and other functional units of the processor 1600.

[0238] The memory 1604 may include one or more caches (e.g., memory local to or included in the processor 1600 or other memory, such as RAM, ROM, DRAM, SDRAM, SRAM, MRAM, flash memory, etc. In some implementations, the memory 1604 may reside within or on a processor chipset (e.g., local to the processor 1600). In some other implementations, the memory 1604 may reside external to the processor chipset (e.g., remote to the processor 1600).

[0239] The memory 1604 may store computer-readable, computer-executable code including instructions that, when executed by the processor 1600, cause the processor 1600 to perform various functions described herein. The code may be stored in a non-transitory computer-readable medium such as system memory or another type of memory. The controller 1602 and/or the processor 1600 may be configured to execute computer-readable instructions stored in the memory 1604 to cause the processor 1600 to perform various functions. For example, the processor 1600 and/or the controller 1602 may be coupled with or to the memory 1604, the processor 1600, the controller 1602, and the memory 1604 may be configured to perform various functions described herein. In some examples, the processor 1600 may include multiple processors and the memory 1604 may include multiple memories. One or more of the multiple processors may be coupled with one or more of the multiple memories, which may, individually or collectively, be configured to perform various functions herein.

[0240] The one or more ALUs 1606 may be configured to support various operations in accordance with examples as described herein. In some implementations, the one or more ALUs 1606 may reside within or on a processor chipset (e.g., the processor 1600). In some other implementations, the one or more ALUs 1606 may reside external to the processor chipset (e.g., the processor 1600). One or more ALUs 1606 may perform one or more

computations such as addition, subtraction, multiplication, and division on data. For example, one or more ALUs 1606 may receive input operands and an operation code, which determines an operation to be executed. One or more ALUs 1606 be configured with a variety of logical and arithmetic circuits, including adders, subtractors, shifters, and logic gates, to process and manipulate the data according to the operation. Additionally, or alternatively, the one or more ALUs 1606 may support logical operations such as AND, OR, exclusive-OR (XOR), not-OR (NOR), and not-AND (NAND), enabling the one or more ALUs 1606 to handle conditional operations, comparisons, and bitwise operations.

[0241] The processor 1600 may support wireless communication in accordance with examples as disclosed herein. The processor 1600 may be configured to or operable to support a means for generating a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements; multiplexing the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description; sending a request for an application session with network traffic differentiation to a server, the request comprising: a common multimodal session identifier, a representation of the protocol description, and a QoS profile comprising a set of QoS requirements of the plurality of multiplexed data flows.

[0242] Figure 17 illustrates an example of a NE 1700 in accordance with aspects of the present disclosure. The NE 1700 may include a processor 1702, a memory 1704, a controller 1706, and a transceiver 1708. The processor 1702, the memory 1704, the controller 1706, or the transceiver 1708, or various combinations thereof or various components thereof may be examples of means for performing various aspects of the present disclosure as described herein. These components may be coupled (e.g., operatively, communicatively, functionally, electronically, electrically) via one or more interfaces.

[0243] The processor 1702, the memory 1704, the controller 1706, or the transceiver 1708, or various combinations or components thereof may be implemented in hardware (e.g., circuitry). The hardware may include a processor, a digital signal processor (DSP), an application-specific integrated circuit (ASIC), or other programmable logic device, or

any combination thereof configured as or otherwise supporting a means for performing the functions described in the present disclosure.

[0244] The processor 1702 may include an intelligent hardware device (e.g., a general-purpose processor, a DSP, a CPU, an ASIC, an FPGA, or any combination thereof). In some implementations, the processor 1702 may be configured to operate the memory 1704. In some other implementations, the memory 1704 may be integrated into the processor 1702. The processor 1702 may be configured to execute computer-readable instructions stored in the memory 1704 to cause the NE 1700 to perform various functions of the present disclosure.

[0245] The memory 1704 may include volatile or non-volatile memory. The memory 1704 may store computer-readable, computer-executable code including instructions when executed by the processor 1702 cause the NE 1700 to perform various functions described herein. The code may be stored in a non-transitory computer-readable medium such as the memory 1704 or another type of memory. Computer-readable media includes both non-transitory computer storage media and communication media including any medium that facilitates transfer of a computer program from one place to another. A non-transitory storage medium may be any available medium that may be accessed by a general-purpose or special-purpose computer.

[0246] In some implementations, the processor 1702 and the memory 1704 coupled with the processor 1702 may be configured to cause the NE 1700 to perform one or more of the functions described herein (e.g., executing, by the processor 1702, instructions stored in the memory 1704). For example, the processor 1702 may support wireless communication at the NE 1700 in accordance with examples as disclosed herein. The NE 1700 may be configured to support a means for generating a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements; multiplexing the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description; sending a request for an application session with network traffic differentiation to a server, the request comprising: a common multimodal session identifier, a representation of the protocol description, and a

QoS profile comprising a set of QoS requirements of the plurality of multiplexed data flows.

[0247] The controller 1706 may manage input and output signals for the NE 1700. The controller 1706 may also manage peripherals not integrated into the NE 1700. In some implementations, the controller 1706 may utilize an operating system such as iOS®, ANDROID®, WINDOWS®, or other operating systems. In some implementations, the controller 1706 may be implemented as part of the processor 1702.

[0248] In some implementations, the NE 1700 may include at least one transceiver 1708. In some other implementations, the NE 1700 may have more than one transceiver 1708. The transceiver 1708 may represent a wireless transceiver. The transceiver 1708 may include one or more receiver chains 1710, one or more transmitter chains 1712, or a combination thereof.

[0249] A receiver chain 1710 may be configured to receive signals (e.g., control information, data, packets) over a wireless medium. For example, the receiver chain 1710 may include one or more antennas for receiving the signal over the air or wireless medium. The receiver chain 1710 may include at least one amplifier (e.g., a low-noise amplifier (LNA)) configured to amplify the received signal. The receiver chain 1710 may include at least one demodulator configured to demodulate the received signal and obtain the transmitted data by reversing the modulation technique applied during transmission of the signal. The receiver chain 1710 may include at least one decoder for decoding and processing the demodulated signal to receive the transmitted data.

[0250] A transmitter chain 1712 may be configured to generate and transmit signals (e.g., control information, data, packets). The transmitter chain 1712 may include at least one modulator for modulating data onto a carrier signal, preparing the signal for transmission over a wireless medium. The at least one modulator may be configured to support one or more techniques such as amplitude modulation (AM), frequency modulation (FM), or digital modulation schemes like phase-shift keying (PSK) or quadrature amplitude modulation (QAM). The transmitter chain 1712 may also include at least one power amplifier configured to amplify the modulated signal to an appropriate power level suitable

for transmission over the wireless medium. The transmitter chain 1712 may also include one or more antennas for transmitting the amplified signal into the air or wireless medium.

[0251] Figure 18 illustrates a flowchart of a method in accordance with aspects of the present disclosure. The operations of the method may be implemented by a NE as described herein. In some implementations, the NE may execute a set of instructions to control the function elements of the NE to perform the described functions.

[0252] At 1802, the method may include generating a protocol description wherein the protocol description contains a mapping of a plurality of data flows to identifiable protocol elements. The operations of 1802 may be performed in accordance with examples as described herein. In some implementations, aspects of the operations of 1802 may be performed by a UE as described with reference to Figure 15.

[0253] At 1804, the method may include multiplexing the plurality of data flows into an application multimodal protocol flow, the multiplexing based on the protocol description. The operations of 1804 may be performed in accordance with examples as described herein. In some implementations, aspects of the operations of 1804 may be performed by a UE as described with reference to Figure 15.

[0254] At 1806, the method may include sending a request for an application session with network traffic differentiation to a server. The operations of 1806 may be performed in accordance with examples as described herein. In some implementations, aspects of the operations of 1806 may be performed by a UE as described with reference to Figure 15.

[0255] It should be noted that the method described herein describes A possible implementation, and that the operations and the steps may be rearranged or otherwise modified and that other implementations are possible.

[0256] The description herein is provided to enable a person having ordinary skill in the art to make or use the disclosure. Various modifications to the disclosure will be apparent to a person having ordinary skill in the art, and the generic principles defined herein may be applied to other variations without departing from the scope of the disclosure. Thus, the disclosure is not limited to the examples and designs described herein but is to be accorded the broadest scope consistent with the principles and novel features disclosed herein.

[0257] There are provided herein arrangements that demultiplex multiplexed data flows of an application session corresponding to a multimodal service/application over 5GS for differentiated optimized QoS handling based on the inherent multimodality and its corresponding protocol description. Focus is placed on encrypted multiplexing capable protocols (QUIC, WebRTC, SRTP/SRTCP + Cryptex for RTP header encryption).

[0258] There is also provided a protocol description including a listing of multiplexed data flows and their individual, potentially different, QoS requirements as part of an application-common set of QoS requirements for differentiated traffic treatment corresponding to a multimodal service over 5GS. Differentiated handling of traffic is performed when a tunnelled session is established between UE and PSA UPF, or alternatively, between AS and PSA UPF. Further provided are system and policy control configuration and corresponding signalling procedures.

[0259] Certain aspects described herein relate to differentiated traffic treatment of multimodal tunnelled and encrypted traffic.

[0260] Accordingly, there is provided a method between a client and a server for multimodal network traffic differentiation of application data comprising of the client multiplexing one or more data flows of the application on a multimodal protocol flow; the client providing a protocol description determining the one or more multiplexed data flows of the multimodal protocol flow and a corresponding QoS profile of a set of QoS requirements of the one or more multiplexed data flows; the server receiving the multimodal protocol flow, and a request for an application session with traffic differentiation comprising of at least a representation of the protocol description and the corresponding QoS profile of the multimodal protocol flow; the server demultiplexing the one or more data flows based at least on the protocol description, the corresponding QoS profile and a network QoS policy; and the server mapping each of the one or more data flows corresponding to the multimodal protocol flow to one or more QoS flows for differentiated traffic handling within the network based at least on the corresponding QoS profile and the network QoS policy.

[0261] The multimodal protocol flow may be at least in part encrypted. The multimodal protocol flow may comprise at least one of: a QUIC tunnelling protocol; a

WebRTC protocol; a combination of at least one secure real-time protocol (SRTP) media stream and at least one of a real-time protocol (RTP) media stream, a real-time control protocol (RTCP) flow, and a secure real-time control protocol (SRTCP) flow; a combination of at least one SRTP media stream with additionally encrypted RTP header extensions and at least one of a real-time protocol (RTP) media stream, a RTCP flow and a SRTCP flow.

[0262] The QUIC tunnelling protocol may comprise an HTTP extended CONNECT over QUIC and UDP proxying one of: a UDP connection; a TCP connection; and an IP connection to a remote endpoint.

[0263] The client may be an Application Server (AS). The server may be a User Plane Function (UPF).

[0264] The representation of the protocol description and the corresponding QoS profile of the multimodal protocol flow may comprise QoS rules provided by a Session Management Function (SMF).

[0265] The network QoS policy may be based at least on one of: a Policy Control Function (PCF) within the network; a Service Level Agreement (SLA); and a Mobile Network Operator (MNO) Operations, Administration and Maintenance (OAM) network configuration.

[0266] The server mapping of the one or more data flows to one or more QoS flows may correspond to each of the one or more data flows being mapped to exactly one QoS flow.

[0267] The one or more QoS flows may be associated to the application session with network traffic differentiation based on at least one of: a multimodal application identifier; a multimodal session identifier; a multimodal service identifier; and a multimodal protocol [e.g., WebRTC] 5-tuple descriptor corresponding to a source IP address, a destination IP address, a source network port, a destination network port, and a protocol flow identifier.

[0268] A further aspect of the present document relates to a protocol description of a client multimedia traffic flow. This aspect is likely embodied in sender logic comprised in a AS/UE.

[0269] Accordingly, there is provided a method at a client for multimodal network traffic differentiation of the client application data comprising: the client generating a protocol description wherein the protocol description contains a mapping of one or more data flows to identifiable protocol elements as an indexed listing; the client multiplexing based on the protocol description one or more data flows into an application multimodal protocol flow; and the client sharing with a server a request for an application session with network traffic differentiation, the request comprising of at least a first part corresponding to a common multimodal session identifier, a second part corresponding to a representation of the protocol description and a third part corresponding to a QoS profile comprising a set of QoS requirements of the one or more multiplexed data flows.

[0270] The identifiable protocol elements may correspond to at least one of: a stream protocol frame identifiable by at least one Stream ID [e.g., a QUIC stream frame identifiable by means of a stream ID, or alternatively, a range/array of Stream IDs]; a datagram protocol frame identifiable by at least one Stream ID [e.g., a QUIC datagram frame identifiable by means of a quarter stream ID, or alternatively a range/array of Stream IDs]; a subprotocol identifiable by at least one of a protocol type and a protocol number [e.g., a protocol identifiable according to IANA protocol IDs and protocol syntax applied to the different subprotocols composing the application protocol flow, e.g., for WebRTC RTP/SRTP, RTCP/SRTCP, DTLS are multiplexed together on the application WebRTC flow]; and a subprotocol carrying an identifiable media payload type [e.g., in RTP/SRTP media payload types are visible to the UPF and can be inspected, identified, and mapped to a specific video/audio/haptic codec profile configuration and payload, e.g., video H.264 baseline profile, H.264 extended profile or H.265 main profile RTP payload types].

[0271] The client may be comprised within one of an Application Server (AS) and a User Equipment (UE). The server may be comprised within an Application Function (AF). The AF communicating the request for the application session with network traffic differentiation with at least one of a Network Exposure Function (NEF), and a Policy Control Function (PCF).

[0272] The client may dynamically update at least one part of its request for the application session with network traffic differentiation based on at least one of: a change in

the protocol description; and a change in the QoS profile and the set of QoS requirements of the one or more multiplexed multimodal data flows.

[0273] The change in the protocol description based on the dynamic termination and reallocation of at least one of the stream identifiers, the protocol identifiers and the RTP media type identifiers.

[0274] A still further aspect of the present document concerns a system and policy configuration for differentiating media on the same traffic flow.

[0275] There is provided a method for a policy control network function comprising: receiving from a first network function a request to establish an IP flow session to a UE, the request including 5-tuple information, and for each media within IP flow information, QoS requirements, protocol description of the media and a range of identifiers; determining QoS rules for each media within IP flow information and associating the QoS rule to a range of identifiers; sending the QoS rules to a second network function (SMF) each QoS rules associated to the UE in the first request.

[0276] There is further provided a method for a SMF comprising: receiving from a first network function one or more QoS rules for a UE each QoS rules associated with a range of identifier; determining configuration rules for a second network function [UPF] wherein the rules provide assistance to the second network function to route received packets [over N6] to a QoS flow towards RAN based on the identifier; and determining UL QoS rules for the UE wherein the UL QoS rules provide assistance to the UE to route uplink packet to a QoS flow towards RAN based on the identifier.

[0277] There is further provided a method for a UPF (with tunnel at N6) comprising: receiving a downlink packet [over N6] where the downlink packet is received via a tunnelled connection from a first server function; decapsulating the downlink packet from the tunnelled connection and processing the identifier (stream identifier) associated to the received packet in the tunnelled connection; determining the QoS flow to route the packet based on configuration rules received from the SMF; and routing the packet over a corresponding QoS flow based on the configuration rules.

[0278] The configuration rule may comprise a protocol description to a stream identifier. The configuration rule may comprise a protocol description to a QoS flow. The configuration rule may comprise a stream identifier mapping to a QoS flow.

[0279] There is further provided a method for a UPF comprising: receiving a downlink packet [over N6]; determining to route the packet over a stream connection to the first UE based on first configuration rules received; determining the stream identifier of the stream connection based on first configuration rules; encapsulating the received packet within a stream connection corresponding to the stream identifier; and routing the stream packet over a corresponding QoS flow based on second configuration rules received from the SMF.

[0280] The first configuration rule may comprise a protocol description to a stream connection and a stream identifier. The second configuration rule may comprise a stream identifier mapping to a QoS flow.

[0281] A yet further aspect of the present document concerns configuration information for the UE.

[0282] There is provided a method for a UE, the method comprising: determining the traffic category and/or identifier of the application for traffic received from the application layer in the UE; and encapsulating and routing the packet over a QUIC stream connection associated with a stream identifier towards a second network function according to first configuration rules where the QUIC stream packet is sent over a QoS flow according to second configuration rules.

[0283] The first configuration rule may comprise one or more of the following: application ID or Traffic Category of the type of traffic initiated by an application; a Masque handling indicator instructing the UE to route the packet over a QUIC connection to the PSA UPF; a protocol Description corresponding to the type of media (e.g. h264, audio, metadata); and a range of stream identifiers to route the media of the QUIC connection.

[0284] The second configuration rule may comprise Packet Detection Rule associating uplink traffic of a QUIC connection containing a QoS flow. The packet detection rule may include the stream identifier (or range of stream identifiers).

[0285] The following abbreviations are relevant in the field addressed by this document: 3GPP, 3rd generation partnership project; 5G, fifth generation; 5GS, 5G System; 5QI, 5G QoS Identifier; AF, application function; AMF, access and mobility function; AR, augmented reality; AS, application server; DL, downlink; NAL, network abstraction layer; PCF, policy control function; PDU, packet data unit; PPS, picture parameter set; QoE, quality of experience; QoS, quality of service; RAN, radio access network; RTCP, real-time control protocol; RTP, real-time protocol; SDAP, service data adaptation protocol; SMF, session management function; SRTCP, secure real-time control protocol; SRTP, secure real-time protocol; UE, user equipment; UL, uplink; UPF, user plane function; VCL, video coding layer; VMAF, video multi-method assessment function; VPS, video parameter set; VR , virtual reality; XR, extended reality; XR AS, XR application server; and XRM, XR media.