METHOD FOR USING 5G EDGE APPLICATION SERVERS FOR LIVE STREAMING OF USER-GENERATED CONTENT

[0001] This application claims priority from U.S. Provisional Application No. 63/187,820, filed on May 12, 2021, in the United States Patent and Trademark Office.

FIELD

[0002] The disclosure generally relates to 5th generation (5G) media streaming (5GMS), and, in particular, to using edge application servers for live streaming of user-generated content.

BACKGROUND

[0003] Network and cloud platforms are used to run various applications. However, there is no existing standard to describe characteristics of a network or cloud platform or its elements

[0004] The 3rd Generation Partnership Project (3GPP) TS23.558 (, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Architecture for enabling Edge Applications (Release 17), V2.0.0) defines the general architecture for enabling edge application, including the discovery of hardware capabilities of an edge element. 3GPP TS 26.501 (3GPP TS 26.501, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects;5G Media Streaming (5GMS); General description and architecture (Release 16), V16.3.1) defines the general architecture for 5G media streaming applications and TS26.512 defines the application programming interface (API) calls for that architecture. 3GPP TR 26.803 (3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Study on 5G Media

Streaming Extensions for Edge Processing (Release 17)V1.5.1) also relates to 5th generation media streaming (5GMS) and edge processing.

[0005] The current 5G edge architecture defined in 3GPP TS23.558 only defines the discovery of the edge application by application clients. The 3GPP TS26.501 only defines the media streaming architecture. However, the discovery, instantiation, and management of the media services on edge resource for the 5GMS application service provider are not defined.

SUMMARY

[0006] According to one or more embodiments, a method for performing live-streaming of user-generated content over a media-streaming network includes transmitting, by a 5th generation media streaming (5GMS)-aware application, a live-streaming request corresponding to the user-generated content to a 5GMS application provider (AP); receiving, by the 5GMS-aware application, from the 5GMS AP, an edge application server (EAS) profile corresponding to the live-streaming request, wherein the EAS profile indicates a service class (SC) from among a plurality of SCs, the SC corresponding to the live-streaming request; selecting, by the 5GMS-aware application, a 5GMS application server (AS) based on the SC; and performing the live-streaming of the user-generated content over the 5GMS network to the 5GMS AS according to the SC.

[0007] According to one or more embodiments, a device for performing live-streaming of user-generated content over a media-streaming network includes at least one memory configured to store program code; and at least one processor configured to read the program code and operate as instructed by the program code, the program code including: transmitting code configured to cause the at least one processor to transmit, by a 5th generation media streaming (5GMS)-aware application, a live-streaming request corresponding to the user-generated content to a 5GMS application provider (AP); receiving code configured to cause

the at least one processor to recieve, by the 5GMS-aware application, from the 5GMS AP, an edge application server (EAS) profile corresponding to the live-streaming request, wherein the EAS profile indicates a service class (SC) from among a plurality of SCs, the SC corresponding to the live-streaming request; selecting code configured to cause the at least one processor to select, by the 5GMS-aware application, a 5GMS application server (AS) based on the SC; and performing code configured to cause the at least one processor to perform the live-streaming of the user-generated content over a 5GMS network to the 5GMS AS according to the SC.

[0008] According to one or more embodiments, a non-transitory computer-readable medium stores instructions including: one or more instructions that, when executed by one or more processors of a device for performing live-streaming of user-generated content over a media-streaming network, cause the one or more processors to: transmit, by a 5th generation media streaming (5GMS)-aware application, a live-streaming request corresponding to the user-generated content to a 5GMS application provider (AP); receive, by the 5GMS-aware application, from the 5GMS AP, an edge application server (EAS) profile corresponding to the live-streaming request, wherein the EAS profile indicates a service class (SC) from among a plurality of SCs, the SC corresponding to the live-streaming request; select, by the 5GMS-aware application, a 5GMS application server (AS) based on the SC; and perform the live-streaming of the user-generated content over a 5GMS network to the 5GMS AS according to the SC.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Further features, the nature, and various advantages of the disclosed subject matter will be more apparent from the following detailed description and the accompanying drawings in which:

[0010] FIG. 1 is a diagram of an environment in which methods, apparatuses, and systems described herein may be implemented, according to embodiments.

[0011] FIG. 2 is a block diagram of example components of one or more devices of FIG. 1.

[0012] FIG. 3 is a block diagram of a media architecture for media streaming, according to embodiments.

[0013] FIG. 4 is a diagram of a 5th generation (5G) edge network architecture, according to embodiments.

[0014] FIG. 5 is a diagram of a 5G media streaming architecture for enabling edge applications, according to embodiments.

[0015] FIG. 6 is a flowchart is an example process for performing live-streaming of user-generated content over a media-streaming network.

[0016] FIG. 7 is a flowchart is an example process for performing live-streaming of user-generated content over a media-streaming network

DETAILED DESCRIPTION

[0017] FIG. 1 is a diagram of an environment 100 in which methods, apparatuses, and systems described herein may be implemented, according to embodiments. As shown in FIG. 1, the environment 100 may include a user device 110, a platform 120, and a network 130. Devices of the environment 100 may interconnect via wired connections, wireless connections, or a combination of wired and wireless connections.

[0018] The user device 110 includes one or more devices capable of receiving, generating, storing, processing, and/or providing information associated with platform 120. For example, the user device 110 may include a computing device (e.g., a desktop computer, a laptop computer, a tablet computer, a handheld computer, a smart speaker, a server, etc.), a

mobile phone (e.g., a smart phone, a radiotelephone, etc.), a wearable device (e.g., a pair of smart glasses or a smart watch), or a similar device. In some implementations, the user device 110 may receive information from and/or transmit information to the platform 120.

[0019] The platform 120 includes one or more devices as described elsewhere herein. In some implementations, the platform 120 may include a cloud server or a group of cloud servers. In some implementations, the platform 120 may be designed to be modular such that software components may be swapped in or out depending on a particular need. As such, the platform 120 may be easily and/or quickly reconfigured for different uses.

[0020] In some implementations, as shown, the platform 120 may be hosted in a cloud computing environment 122. Notably, while implementations described herein describe the platform 120 as being hosted in the cloud computing environment 122, in some implementations, the platform 120 may not be cloud-based (i.e., may be implemented outside of a cloud computing environment) or may be partially cloud-based.

[0021] The cloud computing environment 122 includes an environment that hosts the platform 120. The cloud computing environment 122 may provide computation, software, data access, storage, etc. services that do not require end-user (e.g., the user device 110) knowledge of a physical location and configuration of system(s) and/or device(s) that hosts the platform 120. As shown, the cloud computing environment 122 may include a group of computing resources 124 (referred to collectively as “computing resources 124” and individually as “computing resource 124”).

[0022] The computing resource 124 includes one or more personal computers, workstation computers, server devices, or other types of computation and/or communication devices. In some implementations, the computing resource 124 may host the platform 120. The cloud resources may include compute instances executing in the computing resource

124, storage devices provided in the computing resource 124, data transfer devices provided by the computing resource 124, etc. In some implementations, the computing resource 124 may communicate with other computing resources 124 via wired connections, wireless connections, or a combination of wired and wireless connections.

[0023] As further shown in FIG. 1, the computing resource 124 includes a group of cloud resources, such as one or more applications (APPs) 124-1, one or more virtual machines (VMs) 124-2, virtualized storage (VSs) 124-3, one or more hypervisors (HYPs) 124-4, or the like.

[0024] The application 124-1 includes one or more software applications that may be provided to or accessed by the user device 110 and/or the platform 120. The application 124-1 may eliminate a need to install and execute the software applications on the user device 110. For example, the application 124-1 may include software associated with the platform 120 and/or any other software capable of being provided via the cloud computing environment 122. In some implementations, one application 124-1 may send/receive information to/from one or more other applications 124-1, via the virtual machine 124-2.

[0025] The virtual machine 124-2 includes a software implementation of a machine (e.g. a computer) that executes programs like a physical machine. The virtual machine 124-2 may be either a system virtual machine or a process virtual machine, depending upon use and degree of correspondence to any real machine by the virtual machine 124-2. A system virtual machine may provide a complete system platform that supports execution of a complete operating system (OS). A process virtual machine may execute a single program, and may support a single process. In some implementations, the virtual machine 124-2 may execute on behalf of a user (e.g., the user device 110), and may manage infrastructure of the cloud

computing environment 122, such as data management, synchronization, or long-duration data transfers.

[0026] The virtualized storage 124-3 includes one or more storage systems and/or one or more devices that use virtualization techniques within the storage systems or devices of the computing resource 124. In some implementations, within the context of a storage system, types of virtualizations may include block virtualization and file virtualization. Block virtualization may refer to abstraction (or separation) of logical storage from physical storage so that the storage system may be accessed without regard to physical storage or heterogeneous structure. The separation may permit administrators of the storage system flexibility in how the administrators manage storage for end users. File virtualization may eliminate dependencies between data accessed at a file level and a location where files are physically stored. This may enable optimization of storage use, server consolidation, and/or performance of non-disruptive file migrations.

[0027] The hypervisor 124-4 may provide hardware virtualization techniques that allow multiple operating systems (e.g., "guest operating systems") to execute concurrently on a host computer, such as the computing resource 124. The hypervisor 124-4 may present a virtual operating platform to the guest operating systems, and may manage the execution of the guest operating systems. Multiple instances of a variety of operating systems may share virtualized hardware resources.

[0028] The network 130 includes one or more wired and/or wireless networks. For example, the network 130 may include a cellular network (e.g. a fifth generation (5G) network, a long-term evolution (LTE) network, a third generation (3G) network, a code division multiple access (CDMA) network, etc.), a public land mobile network (PLMN), a local area network (LAN), a wide area network (WAN), a metropolitan area network (MAN),

a telephone network (e.g. the Public Switched Telephone Network (PSTN)), a private network, an ad hoc network, an intranet, the Internet, a fiber optic-based network, or the like, and/or a combination of these or other types of networks.

[0029] The number and arrangement of devices and networks shown in FIG. 1 are provided as an example. In practice, there may be additional devices and/or networks, fewer devices and/or networks, different devices and/or networks, or differently arranged devices and/or networks than those shown in FIG. 1. Furthermore, two or more devices shown in FIG. 1 may be implemented within a single device, or a single device shown in FIG. 1 may be implemented as multiple, distributed devices. Additionally, or alternatively, a set of devices (e.g. one or more devices) of the environment 100 may perform one or more functions described as being performed by another set of devices of the environment 100.

[0030] FIG. 2 is a block diagram of example components of one or more devices of FIG. 1. The device 200 may correspond to the user device 110 and/or the platform 120. As shown in FIG. 2, the device 200 may include a bus 210, a processor 220, a memory 230, a storage component 240, an input component 250, an output component 260, and a communication interface 270.

[0031] The bus 210 includes a component that permits communication among the components of the device 200. The processor 220 is implemented in hardware, firmware, or a combination of hardware and software. The processor 220 is a central processing unit (CPU), a graphics processing unit (GPU), an accelerated processing unit (APU), a microprocessor, a microcontroller, a digital signal processor (DSP), a field-programmable gate array (FPGA), an application-specific integrated circuit (ASIC), or another type of processing component. In some implementations, the processor 220 includes one or more processors capable of being programmed to perform a function. The memory 230 includes a

random access memory (RAM), a read only memory (ROM), and/or another type of dynamic or static storage device (e.g. a flash memory, a magnetic memory, and/or an optical memory) that stores information and/or instructions for use by the processor 220.

[0032] The storage component 240 stores information and/or software related to the operation and use of the device 200. For example, the storage component 240 may include a hard disk (e.g. a magnetic disk, an optical disk, a magneto-optic disk, and/or a solid state disk), a compact disc (CD), a digital versatile disc (DVD), a floppy disk, a cartridge, a magnetic tape, and/or another type of non-transitory computer-readable medium, along with a corresponding drive.

[0033] The input component 250 includes a component that permits the device 200 to receive information, such as via user input (e.g., a touch screen display, a keyboard, a keypad, a mouse, a button, a switch, and/or a microphone). Additionally, or alternatively, the input component 250 may include a sensor for sensing information (e.g., a global positioning system (GPS) component, an accelerometer, a gyroscope, and/or an actuator). The output component 260 includes a component that provides output information from the device 200 (e.g. a display, a speaker, and/or one or more light-emitting diodes (LEDs)).

[0034] The communication interface 270 includes a transceiver-like component (e.g., a transceiver and/or a separate receiver and transmitter) that enables the device 200 to communicate with other devices, such as via a wired connection, a wireless connection, or a combination of wired and wireless connections. The communication interface 270 may permit the device 200 to receive information from another device and/or provide information to another device. For example, the communication interface 270 may include an Ethernet interface, an optical interface, a coaxial interface, an infrared interface, a radio frequency

(RF) interface, a universal serial bus (USB) interface, a Wi-Fi interface, a cellular network interface, or the like.

[0035] The device 200 may perform one or more processes described herein. The device 200 may perform these processes in response to the processor 220 executing software instructions stored by a non-transitory computer-readable medium, such as the memory 230 and/or the storage component 240. A computer-readable medium is defined herein as a non-transitory memory device. A memory device includes memory space within a single physical storage device or memory space spread across multiple physical storage devices.

[0036] Software instructions may be read into the memory 230 and/or the storage component 240 from another computer-readable medium or from another device via the communication interface 270. When executed, software instructions stored in the memory 230 and/or the storage component 240 may cause the processor 220 to perform one or more processes described herein. Additionally, or alternatively, hardwired circuitry may be used in place of or in combination with software instructions to perform one or more processes described herein. Thus, implementations described herein are not limited to any specific combination of hardware circuitry and software.

[0037] The number and arrangement of components shown in FIG. 2 are provided as an example. In practice, the device 200 may include additional components, fewer components, different components, or differently arranged components than those shown in FIG. 2. Additionally, or alternatively, a set of components (e.g., one or more components) of the device 200 may perform one or more functions described as being performed by another set of components of the device 200.

[0038] A 5G media streaming (5GMS) system may be an assembly of application functions, application servers, and interfaces from the 5G media streaming architecture that

support either downlink media streaming services or uplink media streaming services, or both. A 5GMS Application Provider may include a party that interacts with functions of the 5GMS system and supplies a 5GMS Aware Application that interacts with functions of the 5GMS system. The 5GMS Aware Application may refer to an application in the user equipment (UE), provided by the 5GMS Application Provider, that contains the service logic of the 5GMS application service, and interacts with other 5GMS Client and Network functions via the interfaces and application programming interfaces (APIs) defined in the 5GMS architecture. A 5GMS Client may refer to a UE function that is either a 5GMS downlink (5GMSd) Client or a 5GMS uplink (5GMSu) Client, or both.

[0039] The 5GMSd Client may refer to a UE function that includes at least a 5G media streaming player and a media session handler for downlink streaming and that may be accessed through well-defined interfaces/APIs. The 5GMSu Client may refer to an originator of a 5GMSu service that may be accessed through well-defined interfaces/APIs. A 5GMSu media streamer may refer to a UE function that enables uplink delivery of streaming media content to an Application Server (AS) function of the 5GMS Application Provider, and which interacts with both the 5GMSu Aware Application for media capture and subsequent streaming, and the Media Session Handler for media session control.

[0040] A dynamic policy may refer to a dynamic policy and charging control (PCC) rule for an uplink or downlink application flow during a media session. An egest session may refer to an uplink media streaming session from the 5GMS AS towards the 5GMSu Application Provider. An ingest session may refer to a session to upload the media content to a 5GMSd AS. A policy template may refer to a collection of (semi-static) Policy or Control Function (PCF)/Network Exposure Function (NEF) API parameters which are specific to the 5GMS Application Provider and also the resulting PCC rule. A policy template ID may

identify the desired policy template, which is used by the 5GMSd Application Function (AF) to select the appropriate PCF/NEF API towards the 5G system so that the PCF can compile the desired PCC rule. The Media Player Entry may refer to a document or a pointer to a document that defines a media presentation (e.g., a media presentation description (MPD) for DASH or a uniform resource locator (URL) to a video clip file). A Media Streamer Entry may refer to a pointer (e.g., in the form of a URL) that defines an entry point of an uplink media streaming session. A presentation entry may refer to a document or a pointer to a document that defines an application presentation, such as an HTML5 document.

[0041] A Provisioning Session may refer to a data structure supplied at an interface (M1d) by a 5GMSd Application provider that configures the 5GMSd features relevant to a set of 5GMSd Aware Applications. A 5GMSd Media Player may refer to a UE function that enables playback and rendering of a media presentation based on a media play entry and exposing some basic controls such as play, pause, seek, stop, to the 5GMSd Aware Application. Server Access Information may refer to a set of parameters and addresses (including 5GMSd AF and 5GMSd AS addresses) which are needed to activate the reception of a streaming session. A Service and Content Discovery may refer to functionality and procedures provided by a 5GMSd Application Provider to a 5GMS Aware Application that enables the end user to discover the available streaming service and content offerings and select a specific service or content item for access. A Service Announcement may refer to procedures conducted between the 5GMS Aware Application and the 5GMS Application Provider such that the 5GMS Aware Application is able to obtain 5GMS Service Access Information, either directly or in the form of a reference to that information.

[0042] A third party player may refer to a part of an application that uses APIs to exercise selected 5GMSd functions to play back media content. A third party uplink streamer

may refer to a part of an application that uses APIs to exercise selected 5GMSu functions to capture and stream media content.

[0043] FIG. 3 is a diagram of a media architecture 300 for media streaming according to embodiments. A 5GMS Application Provider 301 may use 5GMS for uplink streaming services or downlink streaming services. 5GMS Application provider 301 may provide a 5GMS Aware Application 302 on the UE 303 to make use of 5GMS Client 304 and network functions using interfaces and defined in 5GMS. The 5GMS Aware Application 302 may contain the service logic of the 5GMS application service, and may interact with other 5GMS client and network functions via the interfaces and APIs defined in the 5GMS architecture. 5GMS AS 305 may be an AS dedicated to 5G Media Uplink Streaming. 5GMS Client 304 may be a UE 303 internal function dedicated to 5G Media Uplink Streaming.

[0044] 5GMS AF 306 and 5GMS AS 305 may be Data Network (DN) 307 functions. Functions in trusted DNs may be trusted by the operator's network. Therefore, AFs in trusted DNs may directly communicate with all 5G Core functions. Functions in external DNs may only communicate with 5G Core functions via the NEF 308 using link 320.

[0045] The media architecture 300 may connect UE 303 internal functions and related network functions for 5G Media Uplink Streaming. Accordingly, media architecture 300 may include a number of functions. For example, 5GMS Client 304 on UE 303 may be an originator of 5GMS service that may be accessed through interfaces/APIs. 5GMS Client 304 may include two sub-functions, media session handler 309 and media streamer 310. Media session handler 309 may communicate with the 5GMS AF 306 in order to establish, control and support the delivery of a media session. The Media Session Handler 309 may expose APIs that can be used by the 5GMS Aware Application 302. Media Streamer 310 may communicate with 5GMS AS 305 in order to stream the media content and provide a service to the 5GMS Aware Application

302 for media capturing and streaming, and the Media Session Handler 309 for media session control. 5GMS Aware Application 302 may control 5GMS Client 304 by implementing external application or content service provider specific logic and enabling the establishment of a media session. 5GMS AS 305 may host 5G media functions and may be implemented as a content delivery network (CDN), for example. 5GMS Application Provider 301 may be an external application or content specific media functionality, e.g., media storage, consumption, transcoding and redistribution that uses 5GMS to stream media from 5GMS Aware Application 302. 5GMS AF 306 may provide various control functions to the Media Session Handler 309 on the UE 303 and/or to 5GMS Application Provider 301. 5GMS AF 306 may relay or initiate a request for different PCF 311 treatment or interact with other network functions. The 5GMS AF 306 may be connected to the PCF 311 by an N5 interface 319.

[0046] Media architecture 300 may include a number of different interfaces. For example, link 321 may relate to M1u, which may be a 5GMS Provisioning API exposed by 5GMS AF 306 to provision usage of media architecture 300 and to obtain feedback. Link 322 may relate to M2u, which may be a 5GMS Publish API exposed by 5GMS AS 305 and used when 5GMS AS 305 in trusted DN, such as DN 307, is selected to receive content for streaming service. Link 323 may relate to M3u, which may be an internal API used to exchange information for content hosting on 5GMS AS 305 within a trusted DN such as DN 307. Link 324 may relate to M4u, which may be a Media Uplink Streaming API exposed by 5GMS AS 323 to Media Streamer 310 to stream media content. Link 325 may relate to M5u, which may be a Media Session Handling API exposed by 5GMS AF 305 to Media Session Handler for media session handling, control and assistance that also include appropriate security mechanisms e.g. authorization and authentication. Link 326 may relate to M6u, which may be a UE 303 Media Session Handling API exposed by Media Session Handler 309 to 5GMS Aware Application

302 to make use of 5GMS functions. Link 327 may relate to M7u, which may be a UE Media Streamer API exposed by Media Streamer 310 to 5GMS Aware Application 302 and Media Session Handler 309 to make use of Media Streamer 310. Link 328 may relate to M8u, which may be an Application API which is used for information exchange between 5GMS Aware Application 302 and 5GMS Application Provider 301, for example to provide service access information to the 5GMS Aware Application 302. The UE 303 may also be implemented in a self-contained manner such that interfaces M6u 326 and M7u 327 are not exposed.

[0047] FIG. 4 is a diagram of a 5G edge network architecture 400, according to embodiments. Edge Data Network (EDN) 401 is a local Data Network. Edge Application Server (EAS) 402 and Edge Enabler Server (EES) 403 are contained within the EDN 401. Edge Configuration Server (ECS) 404 provides configurations related to EES 403, including details of EDN 401 hosting EES 403. User Equipment (UE) 405 contains Application Client (AC) 406 and Edge Enabler Client (EEC) 407. EAS 402, EES 403 and ECS 404 may interact with the 3GPP Core Network 408.

[0048] EES 403 provides supporting functions needed for EAS 402 and EEC 407. Functionalities of EES 403 may include: provisioning of configuration information to EEC 407, enabling exchange of application data traffic with EAS; supporting the functionalities of API invoker and API exposing function, for example as specified in 3GPP TS 23.222; interacting with 3GPP Core Network 408 for accessing the capabilities of network functions either directly (e.g. via PCF) or indirectly (e.g. via Service Capability Exposure Function (SCEF)/NEF/SCEF+NEF); supporting the functionalities of application context transfer; supporting external exposure of 3GPP network and service capabilities to EASs 402 over link EDGE-3; supporting the functionalities of registration (i.e., registration, update, and de-

registration) for EEC 407 and EAS; and supporting the functionalities of triggering EAS 402 instantiation on demand.

[0049] EEC 407 provides supporting functions needed for AC. Functionalities of EEC 407 may include: retrieval and provisioning of configuration information to enable the exchange of Application Data Traffic with EAS 402; and discovery of EASs 402 available in the EDN 401.

[0050] ECS 404 provides supporting functions needed for the EEC 407 to connect with an EES 403. Functionalities of ECS 404 are: provisioning of Edge configuration information to the EEC 407, for example the information for the EEC 407 to connect to the EES 403 (e.g. service area information applicable to LADN); and the information for establishing a connection with EESs 403 (such as URI); supporting the functionalities of registration (i.e., registration, update, and de-registration) for the EES 403; supporting the functionalities of API invoker and API exposing function as specified in 3GPP TS 23.222; and interacting with 3GPP Core Network 408 for accessing the capabilities of network functions either directly (e.g. PCF) or indirectly (e.g. via SCEF/NEF/SCEF+NEF).

[0051] AC 406 is the application resident in the UE 405 performing the client function.

[0052] EAS 402 is the application server resident in the EDN 401, performing the server functions. The AC 406 connects to EAS 402 in order to avail the services of the application with the benefits of Edge Computing. It is possible that the server functions of an application are available only as an EAS 402. However, it is also possible that certain server functions are available both at the edge and in the cloud, as an EAS 402 and an Application Server resident in the cloud respectively. The server functions offered by an EAS 402 and its cloud Application Server counterpart may be the same or may differ; if they differ, the Application Data Traffic exchanged with the AC may also be different. EAS 402 may consume the 3GPP

Core Network 408 capabilities in different ways, such as: it may invoke 3GPP Core Network 408 function APIs directly, if it is an entity trusted by the 3GPP Core Network 408; it may invoke 3GPP Core Network 408 capabilities through EES 403; and it may invoke the 3GPP Core Network 408 capability through the capability exposure functions e.g., SCEF or NEF.

[0053] Architecture 400 may include a number of different interfaces for enabling edge applications, which may be referred to as reference points. For example, link EDGE-1 may be a reference point which enables interactions between the EES 403 and the EEC 407. It supports: registration and de-registration of EEC 407 to EES 403; retrieval and provisioning of EAS 402 configuration information; and discovery of EASs 402 available in the EDN 401.

[0054] Link EDGE-2 may be a reference point which enables interactions between EES 403 and the 3GPP Core Network 408. It supports: access to 3GPP Core Network 408 functions and APIs for retrieval of network capability information, e.g. via SCEF and NEF APIs as defined in 3GPP TS 23.501, 3GPP TS 23.502, 3GPP TS 29.522, 3GPP TS 23.682, 3GPP TS 29.122; or with EES 403 deployed within the MNO trust domain (see 3GPP TS 23.501 clause 5.13, 3GPP TS 23.503, 3GPP TS 23.682). Link EDGE-2 may reuse 3GPP reference points or interfaces of EPS or 5GS considering different deployment models.

[0055] Link EDGE-3 may be a reference point which enables interactions between EES 403 and EASs 402. It supports: registration of EASs 402 with availability information (e.g. time constraints, location constraints); de-registration of EASs 402 from EES 403; discovery of target EAS 402 information to support application context transfer; providing access to network capability information (e.g. location information, Quality of Service (QoS) related information); and requesting the setup of a data session between AC and EAS 402 with a specific QoS.

[0056] Link EDGE-4 may be a reference point which enables interactions between ECS 404 and EEC 407. It supports: provisioning of Edge configuration information to the EEC 407.

[0057] Link EDGE-5 may be a reference point which enables interactions between AC and EEC 407.

[0058] Link EDGE-6 may be a reference point which enables interactions between ECS 404 and EES 403. It supports: registration of EES 403 information to ECS 404.

[0059] Link EDGE-7 may be a reference point which enables interactions between EAS 402 and the 3GPP Core Network 408. It supports: access to 3GPP Core Network 408 functions and APIs for retrieval of network capability information, e.g. via SCEF and NEF APIs as defined in 3GPP TS 23.501, 3GPP TS 23.502, 3GPP TS 29.522, 3GPP TS 23.682, 3GPP TS 29.122; or with EAS 402 deployed within the MNO trust domain (see 3GPP TS 23.501 clause 5.13, 3GPP TS 23.682). Link EDGE-7 may reuse 3GPP reference points or interfaces of EPS or 5GS considering different deployment models.

[0060] Link EDGE-8 may be a reference point which enables interactions between the ECS 404 and the 3GPP Core Network 408. It supports: a) access to 3GPP Core Network 408 functions and APIs for retrieval of network capability information, e.g. via SCEF and NEF APIs as defined in 3GPP TS 23.501, 3GPP TS 23.502, 3GPP TS 29.522, 3GPP TS 23.682, 3GPP TS 29.122; and with the ECS 404 deployed within the MNO trust domain (see 3GPP TS 23.501 clause 5.13, 3GPP TS 23.682). Link EDGE-8 may reuse 3GPP reference points or interfaces of EPS or 5GS considering different deployment models.

[0061] The AC 406 may send an inquiry to the EES 403 through the EEC 407, to discover the suitable EASs. In this inquiry, the AC 406 includes EAS discovery filters that define the desired characteristics of the suitable EAS. In the response, the EEC 407 provides the AC

406 the list of matching EASs and some of their characteristics. The AC 406 then selects the best EAS from the list.

[0062] FIG. 5 is a diagram of a 5G media streaming architecture 500 for enabling edge applications, according to embodiments. The 5GMS Application Provider 501, 5GMS Aware Application 502, UE 503, 5GMS client 504, 5GMS AS 505, 5GMS AF 506, DN 507, NEF 508, media session handler 509, media streamer 510, and PCF 511, along with interfaces M1 521, M2 522, M3 523, M4 524, M5 525, M6 526, M7 527, M8 528, N5 519 and N33 520 are similar to their counterparts in FIG. 3 and therefore the detailed description of these components will be omitted.

[0063] The architecture 500 includes an EES 550 as part of the 5GMS AF 506, an EAS 522 on the DN 507, an EEC 554 as part of the media session handler 509, and an ECS 556. The EEC 554 is connected to the EES 550 by the Edge-1 interface 570. The EAS 552 is connected to the EES 550 by the Edge-3 interface 572. The EEC 554 is connected to the ECS 556 by the Edge-4 interface 274. The EEC 554 is connected to the 5GMS Aware Application 502 by the Edge-5 interface 576. The ECS 556 is connected to the EES 550 by the Edge-6 interface 578. Lastly, the Edge-9 interface 580 is connected to the EES 550.

[0064] In FIG. 5, the 5GMS Application Provider (AP) requests to 5GMS AF the provisioning of the session using the M1 interface. Then, through the same interface, AP requests provisioning of different session features including server certificate, content preparation, content hosting configuration, reporting, consumption reporting, policy, and others.

[0065] In FIG. 5, the 5GMS Application Provider (AP) requests to 5GMS AF the provisioning of the session using the M1 interface. Then, through the same interface, AP requests provisioning of different session features including server certificate, content preparation, content hosting configuration, reporting, consumption reporting, policy, and others.

[0066] As discussed above, the current 5G Edge architecture defined in 3GPP TS23.558 only defines the discovery of the edge application by Application clients. The 3GPP TS26.501 only defines the media streaming architecture. The 3GPP SWG4 recently developed an architecture for running 5G media applications on 5G edge servers by combining the two architectures.

[0067] Accordingly, embodiments may provide a method of using the above architecture for running live streaming services for user-generated content.

[0068] In embodiments, an example use case may correspond to user-generated live streaming. The use-case can be summarized as the following. A social influencer starts a live captured media session and publishes the content through 5G Media Uplink Streaming. The content is then distributed live to several or many viewers through 5G Media Downlink Streaming.

[0069] Embodiments may provide several usage scenarios with various configuration options which may change during the live session. For example:

1) The social influencer may:
    a) be static, occasionally moving, or highly mobile, for example in a vehicle, on skis, or on a bicycle,
    b) produce different quality of content, depending on lighting conditions, speed, as well as based on the quality of the camera and the available uplink bandwidth,
    c) produce highly valuable content that requires extra content protection,
    d) want to capitalize on the stream by allowing ad-insertion in the content (targeted pre-roll and/or mid-roll).

2) The viewers may be quite diverse and changing because:
   a) they may be dynamically joining or leaving the live stream,
   b) their number might be just a few or they may quite many, in the range of tens of thousands or more, for example for a popular influencer,
   c) they may be geographically spread with different densities in various areas, and their densities may change during the session,
   d) they may consume the service on different devices, for example on 4K TV sets, mobile phones, in-car receivers, or tablets, with different operating systems, DRM capabilities as well as different codec hardware capabilities,
   e) some of them may be mobile, i.e on a car or public transport,
   f) they may react (smilies, comments, likes, audio dubs, images, avatars, and animations) to the content or previous reactions by the viewers who watched the content earlier.
3) The service requirements may be quite different. The content may need to be:
   a) available for live and/or on-demand consumption,
   b) only available for consumption after the end of the live uplink session, i.e. it is uploaded entirely before being made available to followers,
   c) available with a required target latency with ranges in between capture and display of as low as 1 second up to several tens of seconds,
   d) dubbed into the same or different language,
   e) provided with automatic extraction and addition of captions/subtitles from the audio,
   f) post-processed to improve the audio and/or visual quality,
   g) processed by adding overlays and content tags and other augmented material,
   h) indexed, including the addition of thumbnail navigation in real time,
   i) provided to regional proxies with specific metadata such as black-out information, ad insertion opportunities, language settings or other service metadata,
   j) profanity checked and appropriately altered before being distributed,
   k) available for viewing for some time (from a few minutes to forever) after the end of the live session.

[0070] According to embodiments, in a simple reference scenario of 5G Media Uplink and Downlink Streaming, one or more of the following aspects may be supported:

1) An application manages the service. The server application may be run by external

application providers, by MNO, or by a collaboration of the application provider and MNO.

2) Uplink streaming is provided through 5G media uplink streaming to the application.
3) The application may perform one or more of the following processes:
   a) It decodes the received content
   b) It applied the various processes the content such as:
      i) Upscaling
      ii) Light correction
      iii) Stabilization
      iv) Dubbing
      v) Captioning
      vi) Overlaying and tagging
      vii) Indexing
      viii) Navigation improvements.
   c) It encodes the uploaded content to 5GMS downlink streaming formats.
   d) It packages the content and adds appropriate ad metadata.
   e) It applies content protection and DRM.
4) As soon as the content becomes available, the server application uses downlink streaming for distribution.
   a) It provisions a 5GMS downlink streaming service.
   b) It ingests the content into a 5GMS streaming service.
   c) It gets feedback from consumption reporting on what content is consumed and may change the encoded streaming formats.

[0071] In embodiments, one or more of the following aspects may be important for the service:

1) Can the service be provided throughout the session without any interruption considering the dynamic aspect of the service, for example when a highly mobile edge media streaming client makes a transition between two edge processing environments?
2) Can the desired end-to-end latency be met with the reference scenario and if not, what are possible ways to realize this?
3) Is the content generation flexible and fast enough to address different user populations, yet highly utilized, i.e. the variations of the content is consumed by one or more viewers and is shared as much as possible by many viewers?
4) Is there a benefit to push certain processing closer to the influencer, certain viewers, or in-

between (in terms of bandwidth, latency, and processing requirements)?

Service Class

[0072] Accordingly, embodiments may make use of service classes that may define the set of services offered to a user. For example, user-generated live streaming use case covers scenarios where various processing may occur depending on a requested class of service. In many sessions, the service may occur in the cloud or at an edge close to the capturing device. In this use case, an Application Provider, for example 5GMS Application Provider 501, may define various classes of service.

[0073] Each class of service, which may be referred to as a Service Class, may provide a content preparation process that supports one or more specific features from the list below:

1) Quality improvement such as
    a) Upscaling
    b) Light correction
    c) Stabilization
    d) Audio quality improvement
2) Enriching and adding interactivity features such as
    a) Dubbing and captioning
    b) Overlaying and tagging
    c) Indexing and key feature detection
    d) Navigation improvements
3) Multirate encoding of content
    a) Content-aware encoding
    b) Low latency
    c) Just in time encoding and transcoding
    d) Multi-path encoding for on-demand content
4) Splicing and ad insertion
    a) Splicing content and manifest signaling
    b) Ad insertion queues and information
    c) Language setting and other service metadata

5) Applying content protection and DRM

Service Class Requirements

[0074] For each SC, a set of Service Class Requirements (SCR) may be defined for a 5GMS AS running the service, which may include one or more of the following aspects:

1) EAS type
2) Hardware resources:
    a) Compute
    b) Graphical compute
    c) Memory
    d) Storage
3) Connectivity
    a) Bandwidth
    b) Latency
    c) Maximum request rate
    d) Maximum response time
4) Availability
5) Functional support
    a) Accelerated encoders/transcoders
    b) Quality improvement functions
    c) Content enriching and adding interactivity
    d) Splicing and ad-insertion tools
    e) Stabilization

[0075] In order to set up the service, the Application Provider may maintain three objects:

1. Service Class (SC)
2. Service Class Requirements (SCR)
3. The audience Scale and Geographical Distribution (SGD)

[0076] FIG. 6 shows an example of a high-level call flow for the use case discussed above.

[0077] First, at operation 602, the 5GMS Application Provider (AP) 501 requests provisioning a session for user-generated live stream service.

[0078] At operation 604, the 5GMS-Aware Application 502 requests the start of service from the 5GMS AP 501.

[0079] At operation 606, the 5GMS AP 501, using the user's SC, SCR, and SGD in its database and the current user location, may transmits the 5GMS-Aware Application an EAS Profile (for example as defined in TS23.558 section 8.2.4).

[0080] At operation 608, the 5GMS-Aware Application 502 requests locating a server that can provide the class of service which the user is signed up for, and MSH 509 and EEC 554 may assist in locating such a server.

[0081] At operation 610, a new 5GMS AS may be provisioned if needed. Operation 610 may be performed by one or more of 5GMS AF 506, EES 550, and ECS 556.

[0082] At operation 612, the 5GMS AF may provide the list of suitable 5GMS AS and their information to one or more of MSH 509, EEC 554, and 5GMS-Aware Application 502.

[0083] At operation 614, one or more of MSH 509, EEC 554, and 5GMS-Aware Application 502 may select a preferred 5GMS AS instance based for example on the user class of service or SC and the user's location.

[0084] At operation 616, the uplink streaming may be performed. The media may be uplink streamed to the 5GMS AS, and the content may be prepared according to the SCR and is delivered to distribution (CDN).

[0085] In embodiments corresponding to this use case, the relocation of 5GMS AS may occur due to various reasons, for example:

1. The relocation of the user.
2. Audience diversity/geographical change may require increase transcoders.
3. Change of service class which requires more computational resources

[0086] In any of the above cases, the service may need to be transferred to a new 5GMS AS. The relocation may be performed according to TS23.558 using the ACR detection entity in Table 1.

Table 1: ACR detection entities for each relocation reason

Relocation reason	ACR detection entity
User relocation	MSH/EEC
Audience diversity increase	5GMS AS/EAS
Change of service class	The Application/MSH/EEC

[0087] Accordingly, embodiments may relate to a method for live streaming of user-generated content on 5G networks, wherein the content preparation is performed on the 5G edge server, wherein various services classes are offered, wherein in each class of service, a set of content preparation processing is defined and for that class of service, and a set of service class requirements is defined, wherein the application on the user device uses these requirements to locate and discover the appropriate 5G edge servers to run the service, and chooses the best one based on additional information provided by the 5G edge servers, such as their location, efficiency, cost, and other parameters, wherein the service can be relocated to a new edge server based on various reasons.

[0088] FIG. 7 is a flowchart is an example process 700 for enabling edge applications, according to embodiments. In some implementations, one or more process blocks of FIG. 7 may be performed by any of the elements discussed above with respect to FIGs. 1-6.

[0089] As shown in FIG. 7, process 700 may include transmitting, by a 5th generation media streaming (5GMS)-aware application, a live-streaming request corresponding to the user-generated content to a 5GMS application provider (AP) (block 702).

[0090] As further shown in FIG. 7, process 700 may include receiving, by the 5GMS-aware application, an edge application server (EAS) profile corresponding to the live-streaming request from the 5GMS AP, wherein the EAS profile indicates a service class (SC) corresponding to the live-streaming request from among a plurality of SCs (block 704).

[0091] As further shown in FIG. 7, process 700 may include selecting, by the 5GMS-aware application, a 5GMS application server (AS) based on the SC (block 706).

[0092] As further shown in FIG. 7, process 700 may include performing the live-streaming of the user-generated content over the media streaming network, for example a 5GMS network, to the 5GMS AS according to the SC (block 708).

[0093] In embodiments, the selecting may further include transmitting, by the 5GMS-aware application to a 5GMS application function (AF), a request to locate at least one AS corresponding to the SC; receiving, by the 5GMS-aware application from the 5GMS AF, a list of the at least one 5GMS AS corresponding to the SC; and selecting the 5GMS AS from the list based on the SC.

[0094] In embodiments, the SC may specify a content preparation process supporting at least one from among upscaling, light correction, stabilization, audio quality improvement, dubbing, captioning, overlaying, tagging, indexing, key feature detection, navigation improvement, multirate encoding, content-aware encoding, low latency encoding, just-in-time encoding, transcoding multi-path encoding for on-demand content, splicing content, manifest signaling, ad insertion queues, ad-insertion information, language setting, service metadata, content protection, and digital rights management.

[0095] In embodiments, the EAS profile may further indicate a service class requirement (SCR) corresponding to the SC, and the 5GMS AS may be selected based on the SCR.

[0096] In embodiments, the SC and the SCR may be stored in database corresponding to the 5GMS AP.

[0097] In embodiments, the SCR may specify at least one from among a hardware resource corresponding to the SC, a connectivity parameter corresponding to the SC, an availability corresponding to the SC, and a function corresponding to the SC.

[0098] In embodiments, the EAS profile may further indicate a scale and geographical distribution (SGD) corresponding to the SC, wherein the SGD indicates at least one of an audience size and an audience location, and wherein the 5GMS AS is selected based on the SGD and a location of a user corresponding to the live-streaming request.

[0099] In embodiments, based on the 5GMS-aware application determining that the SC changed to a new SC, process 700 may further include selecting a new 5GMS AS corresponding to the new SC, and performing the live-streaming of the user-generated content to the new 5GMS AS according to the new SC.

[0100] Although FIG. 6 shows example blocks of process 600, in some implementations, process 600 may include additional blocks, fewer blocks, different blocks, or differently arranged blocks than those depicted in FIG. 6. Additionally, or alternatively, two or more of the blocks of process 600 may be performed in parallel.

[0101] Further, the proposed methods may be implemented by processing circuitry (e.g., one or more processors or one or more integrated circuits). In one example, the one or more processors execute a program that is stored in a non-transitory computer-readable medium to perform one or more of the proposed methods.

[0102] The techniques described above can be implemented as computer software using computer-readable instructions and physically stored in one or more computer-readable media.

[0103] Embodiments of the present disclosure may be used separately or combined in any order. Further, each of the embodiments (and methods thereof) may be implemented by processing circuitry (e.g., one or more processors or one or more integrated circuits). In one example, the one or more processors execute a program that is stored in a non-transitory computer-readable medium.

[0104] The foregoing disclosure provides illustration and description, but is not intended to be exhaustive or to limit the implementations to the precise form disclosed. Modifications and variations are possible in light of the above disclosure or may be acquired from practice of the implementations.

[0105] As used herein, the term component is intended to be broadly construed as hardware, firmware, or a combination of hardware and software.

[0106] Even though combinations of features are recited in the claims and/or disclosed in the specification, these combinations are not intended to limit the disclosure of possible implementations. In fact, many of these features may be combined in ways not specifically recited in the claims and/or disclosed in the specification. Although each dependent claim listed below may directly depend on only one claim, the disclosure of possible implementations includes each dependent claim in combination with every other claim in the claim set.

[0107] No element, act, or instruction used herein should be construed as critical or essential unless explicitly described as such. Also, as used herein, the articles “a” and “an” are intended to include one or more items, and may be used interchangeably with “one or more.” Furthermore, as used herein, the term “set” is intended to include one or more items (*e.g.*, related items, unrelated items, a combination of related and unrelated items, etc.), and may be used interchangeably with “one or more.” Where only one item is intended, the term “one” or

similar language is used. Also, as used herein, the terms “has,” “have,” “having,” or the like are intended to be open-ended terms. Further, the phrase “based on” is intended to mean “based, at least in part, on” unless explicitly stated otherwise.

CLAIMS

1. A method for performing live-streaming of user-generated content over a media-streaming network, the method comprising:

transmitting, by a 5th generation media streaming (5GMS)-aware application, a live-streaming request corresponding to the user-generated content to a 5GMS application provider (AP);

receiving, by the 5GMS-aware application, from the 5GMS AP, an edge application server (EAS) profile corresponding to the live-streaming request, wherein the EAS profile indicates a service class (SC) from among a plurality of SCs, the SC corresponding to the live-streaming request;

selecting, by the 5GMS-aware application, a 5GMS application server (AS) based on the SC; and

performing the live-streaming of the user-generated content over a 5GMS network to the 5GMS AS according to the SC.

2. The method of claim 1, selecting further comprises:

transmitting, by the 5GMS-aware application to a 5GMS application function (AF), a request to locate at least one AS corresponding to the SC;

receiving, by the 5GMS-aware application from the 5GMS AF, a list of the at least one 5GMS AS corresponding to the SC; and

selecting the 5GMS AS from the list based on the SC.

3. The method of claim 1, wherein the SC specifies a content preparation process supporting at least one from among upscaling, light correction, stabilization, audio quality improvement, dubbing, captioning, overlaying, tagging, indexing, key feature detection, navigation improvement, multirate encoding, content-aware encoding, low latency encoding, just-in-time encoding, transcoding multi-path encoding for on-demand content, splicing content, manifest signaling, ad insertion queues, ad-insertion information, language setting, service metadata, content protection, and digital rights management.

4. The method of claim 1, wherein the EAS profile further indicates a service class requirement (SCR) corresponding to the SC, and

wherein the 5GMS AS is selected based on the SCR.

5. The method of claim 4, wherein the SC and the SCR are stored in database corresponding to the 5GMS AP.

6. The method of claim 4, wherein the SCR specifies at least one from among a hardware resource corresponding to the SC, a connectivity parameter corresponding to the SC, an availability corresponding to the SC, and a function corresponding to the SC.

7. The method of claim 1, wherein the EAS profile further indicates a scale and geographical distribution (SGD) corresponding to the SC,

wherein the SGD indicates at least one of an audience size and an audience location, and

wherein the 5GMS AS is selected based on the SGD and a location of a user corresponding to the live-streaming request.

8. The method of claim 1, wherein based on the 5GMS-aware application determining that the SC changed to a new SC, the method further comprises selecting a new 5GMS AS corresponding to the new SC, and performing the live-streaming of the user-generated content to the new 5GMS AS according to the new SC.

9. A device for performing live-streaming of user-generated content over a media-streaming network, the device comprising:

at least one memory configured to store program code; and

at least one processor configured to read the program code and operate as instructed by the program code, the program code including performing the method according to any one of claims 1 to 8.

10. A non-transitory computer-readable medium storing instructions, the instructions comprising: one or more instructions that, when executed by one or more processors of a device for performing live-streaming of user-generated content over a media-streaming network, cause the one or more processors to perform the method according to any one of claims 1 to 8.