ETSI GS MEC 002 V1.1.1 (2016-03)

Mobile Edge Computing (MEC); Technical Requirements

Disclaimer

The present document has been produced and approved by the Mobile Edge Computing (MEC) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference
DGS/MEC-002TechReq

Keywords
MEC, requirements

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Group Specification (GS) has been produced by ETSI Industry Specification Group (ISG) Mobile Edge Computing (MEC).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

The present document specifies the requirements for Mobile Edge Computing with the aim of promoting interoperability and deployments. It contains normative and informative parts.

The present document also contains an annex describing example use cases and their technical benefits, for the purpose of deriving requirements.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at https://docbox.etsi.org/Reference/.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

Not applicable.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI GS MEC 001: "Mobile Edge Computing (MEC); Terminology".

[i.2] Mobile-Edge Computing - Introductory Technical White Paper. Sept. 2014.

NOTE: Available at https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf.

[i.3] Flinck H. et al. (September 2015): "Mobile Throughput Guidance Inband Signaling Protocol, Internet draft, Internet Engineering Task Force".

NOTE: Available at https://tools.ietf.org/html/draft-flinck-mobile-throughput-guidance-03 (Work in progress).

[i.4] ETSI GS NFV 002: "Network Functions Virtualisation (NFV); Architectural Framework".

[i.5] Byte Caching in Wireless Networks.

NOTE: Available at http://researcher.ibm.com/researcher/files/us-aruni/ByteCachingicdcs2012.pdf.

[i.6] A Protocol-Independent Technique for Eliminating Redundant Network Traffic.

NOTE: Available at https://djw.cs.washington.edu/papers/spring-sigcomm00.pdf.

[i.7] Sprecher N. et al.: "Requirements and reference architecture for Mobile Throughput Guidance Exposure, Internet draft, Internet Engineering Task Force", September 2015.

NOTE: Available at https://tools.ietf.org/html/draft-sprecher-mobile-tg-exposure-req-arch-02 (Work in progress).

[i.8] Small Cells Forum White Paper SCF081: "Enterprise unified communications services with small cells".

NOTE: Available at http://www.scf.io/en/documents/081-Enterprise_unified_communications_services_with_small_cells.php.

[i.9] IEEE 1588™: "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in ETSI GS MEC 001 [i.1] apply.

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in ETSI GS MEC 001 [i.1] and the following apply:

API	Application Programming Interface
BYO	Bring Your Own
DSRC	Digital Short-Range Communications
EAB	Edge Accelerated Browser
GNSS	Global Navigation Satellite System
GPRS	General Packet Radio Service
GPS	Global Positioning System
GTP	GPRS Tunnelling Protocol
HTTP	Hyper Text Transfer Protocol
IM	Instant Messaging
LAN	Local Area Network
NTP	Network Time Protocol
PBX	Private Branch Exchange
PTP	Precision Time Protocol
QCI	Quality Class Indicator
QoE	Quality of Experience
RAT	Radio Access Technology
SLA	Service Level Agreement
SMS	Short Message Service
SPID	Subscriber Profile ID
TCP	Transmission Control Protocol
TEID	Tunnel Endpoint ID
VNF	Virtualised Network Function

4 Generic principles

4.1 Introduction

The following principles are important to understand in the context of Mobile Edge Computing.

4.2 NFV alignment

Mobile Edge Computing uses a virtualisation platform for running applications at the mobile network edge. Network Functions Virtualisation (NFV) provides a virtualisation platform to network functions. The infrastructure that hosts their respective applications or network functions is quite similar.

In order to allow operators to benefit as much as possible from their investment, it would be beneficial to reuse the infrastructure and infrastructure management of NFV to the largest extent possible, by hosting both VNFs (Virtual Network Functions) and mobile edge applications on the same or similar infrastructure. Subject to gap analysis, this might require a number of enhancements (e.g. regarding the sharing of resources with NFV Management and Orchestration, etc.).

4.3 Mobility support

Mobility is an essential functionality of 3GPP networks. Most devices connected to a 3GPP network are moving around within the mobile network. Even fixed devices can "move", especially when located at cell edge, but also when changing RATs, etc., or during exceptional events (e.g. power cut from a base station, etc.).

Some mobile edge applications are state-independent and do not need to keep state information related to the UEs they are serving. For example, an application in the category "network performance and QoE improvements" will only improve the performance of the UE traffic when the traffic goes through that mobile edge host. When the UE moves to a different location covered by another mobile edge host, it will be the application hosted on that mobile edge host that will take care of the UE after a brief transition period. Past interaction is not useful for the application.

Other mobile edge applications, notably in the category "consumer-oriented services", are specifically related to the user activity. Either the whole application is specific to the user, or at least it needs to maintain some application-specific user-related information that needs to be provided to the instance of that application running on another mobile edge host.

As a consequence of UE mobility, the mobile edge system needs to support the following:

- continuity of the service;
- mobility of application (VM); and
- mobility of application-specific user-related information.

4.4 Deployment independence

For reasons of performance, costs, scalability, operator preferred deployments, etc., different deployment scenarios need to be supported:

- deployment at the radio node;
- deployment at an aggregation point;
- deployment at the edge of the Core Network (e.g. in a distributed data centre, at a gateway);
- etc.

In order to fulfil all these deployment options, the framework of the MEC architecture needs to allow all these scenarios and the requirements need to be able to address all these deployment options. Requirements that cannot be fulfilled for all deployment options cannot be made mandatory, but might be conditional or optional.

When a mobile edge platform is deployed on a host located in a cell aggregation site, mobile edge services running on that platform might need to retrieve information from the radio node(s), for instance, to readout the traffic load and resource block usage of a specific cell.

In order to prevent the illegal access from dishonest terminals and mobile edge application developers, authentication and secure tunnel communication are necessary between the radio node(s) and the mobile edge service.

NOTE: The interface between the radio node(s) and the mobile edge service is not specified in Mobile Edge Computing Group Specifications.

4.5 Simple and controllable APIs

In order to enable the development of a strong ecosystem for Mobile Edge Computing, it is very important to develop APIs that are as simple as possible and are directly answering the needs of applications. To the extent this is possible, Mobile Edge Computing specifications need to reuse existing APIs that fulfil the requirements.

In particular circumstances, operators might need to be able to control dynamically the access to certain APIs by a mobile edge application. Examples include the mitigation of high load of a radio node or mobile edge host, or when the information of a specific radio node or cell cannot be provided.

4.6 Smart application location

Mobile edge applications have a number of requirements, in terms of computing, storage and network resources. More importantly, some applications might have requirements in terms of latency (including latency fairness), etc.

For a certain number of mobile edge applications, the conditions might evolve over time and require the mobile edge system to change the location of the application, e.g. as the UEs are moving from cell to cell.

Also, different locations may have different "costs" (in terms of resource availability, etc.), and it might not be always the best choice to run a mobile edge application at the "best" location (to the detriment of other applications).

For these reasons, mobile edge applications need to run "at the right place" at the right moment, and might have to move when the conditions evolve. In order to support this, the mobile edge system needs to provide a system-wide lifecycle management of applications.

4.7 Representation of features

The present document describes requirements towards the framework and architecture of Mobile Edge Computing.

In addition to the definition of requirements applicable to all deployments, this specification introduces the concept of features in order to cater for the different needs of different deployments. A feature is defined as a group of related requirements and is assigned a unique name.

Support for a feature can be mandatory, optional or conditional. Where feature level support is optional or conditional, all other requirements (mandatory or optional) related to that feature are themselves dependent upon support for the feature itself.

The following example illustrates an optional feature with a conditional mandatory and a conditional optional requirement.

EXAMPLE: [Req-1] The Mobile edge system may support a feature called *XYZ*.
[Req-2] When the Mobile edge system supports the feature *XYZ*, the system shall…
[Req-3] When the Mobile edge system supports the feature *XYZ*, the system may…

The architectural framework needs to support mechanisms to identify whether a specific feature is supported. Such information might need to be considered when executing certain tasks, such as the instantiation of an application.

5 Generic requirements

5.1 Requirements on the framework

[Framework-01] The design of the mobile edge system should attempt to reuse the NFV virtualisation infrastructure and its management functionality, as described in the NFV architecture framework in ETSI GS NFV 002 [i.4], possibly with some enhancements. Concepts that have been developed or studied in NFV Group Specifications and that are needed for Mobile Edge Computing should be reused whenever possible. This might require some enhancements specific to Mobile Edge Computing.

[Framework-02] It shall be possible to enable the deployment of mobile edge applications on the same infrastructure as ETSI NFV-based VNFs.

[Framework-03] It shall be possible to deploy the mobile edge platform on mobile edge hosts in various locations, including radio nodes, aggregation points, gateways, and in a distributed data centre at the edge of the Core Network.

NOTE: Some requirements might not be fulfilled by certain deployment options.

5.2 Application lifecycle

[Lifecycle-01] The mobile edge host shall be available for the hosting of mobile edge applications.

[Lifecycle-02] The mobile edge management shall support the instantiation of an application on a mobile edge host within the mobile edge system.

[Lifecycle-03] The mobile edge management shall support the instantiation of an application on a mobile edge host when required by the operator. This may be in response to a request by an authorized third-party.

[Lifecycle-04] The mobile edge management shall support the termination of a running application when required by the operator. This may be in response to a request by an authorized third-party.

[Lifecycle-05] The mobile edge management shall be able to identify which features and mobile edge services a mobile edge application requires to run, and which additional features and mobile edge services it can use if available.

NOTE 1: This allows the mobile edge system to decide whether and on which mobile edge host to instantiate the application.

[Lifecycle-06] The mobile edge management shall be able to identify which features and mobile edge services are available on a particular mobile edge host.

NOTE 2: This allows the mobile edge management to decide whether a particular application can be instantiated on that host.

5.3 Applications environment

The applications environment describes the security, packaging and run-time environment models for hosting mobile edge applications on the mobile edge host.

[AppEnvironment-01] It shall be possible to deploy mobile edge applications on different mobile edge hosts in a seamless manner, without a specific adaptation to the application.

[AppEnvironment-02] The mobile edge management shall be able to verify the authenticity of a mobile edge application.

[AppEnvironment-03] The mobile edge management shall be able to verify the integrity of a mobile edge application (integrity protection).

5.4 Support of mobility

[Mobility-01] The mobile edge system shall be able to maintain connectivity between a UE and an application instance when the UE performs a handover to another cell associated with the same mobile edge host.

[Mobility-02] The mobile edge system shall be able to maintain connectivity between a UE and an application instance when the UE performs a handover to another cell not associated with the same mobile edge host.

[Mobility-03] The mobile edge platform may use available radio network information to optimize the mobility procedures required to support service continuity.

EXAMPLE: Using UE mobility information to optimize the handling of mobility events by the application (see clause 6.2.2, Connectivity) and of application mobility (see clause 6.3.2, Feature *SmartRelocation*).

6 Services requirements

6.1 General

The mobile edge platform on a mobile edge host provides a framework for delivering mobile edge services and platform essential functionality to mobile edge applications running on the mobile edge host.

A mobile edge service is provided and consumed. Both the mobile edge platform itself and authorized mobile edge applications can provide services. Similarly, both the mobile edge platform itself and authorized mobile edge applications can consume services.

In some cases, and especially in a multi-vendor environment, the service can be provided concurrently by multiple sources. This allows the platform or the applications consuming the service to receive all information required for executing their tasks.

Many of the applications require accurate time information synchronized to the time domain of the operator or application provider. Such applications require exact time of specific events occurrence for analytics information collection and pre-processing, time tagging of the location information, synchronized time intervals for the SLA throughput reports, platform performance monitoring for latency and response times and many others.

Since the platform is located in the synchronized environment required for the mobile network operation, accurate time of day information can be delivered to the platform by the same means as it is provided to the mobile Base Stations. Known techniques include usage of GNSS receivers, running IEEE 1588™ [i.9] PTP protocol, NTP protocol or a combination of the above.

The mobile edge platform will have a means to acquire accurate Time of Day information and make this information available to the hosted applications.

6.2 Platform essential functionality

6.2.1 Mobile edge services

[Services-01] The mobile edge platform shall have the capability to provide mobile edge services that can be consumed by authorized mobile edge applications.

[Services-02] The mobile edge platform shall allow authorized mobile edge applications to provide services that can be consumed by the platform or by authorized mobile edge applications running on the mobile edge host.

NOTE 1: Providing a service by an application to the mobile edge platform includes that the platform can receive information from that application. This information can be used by the mobile edge platform to provide other services.

[Services-03] The mobile edge platform shall provide functionality to allow authorized mobile edge applications to communicate with mobile edge services provided by the platform.

[Services-04] The mobile edge platform shall allow authentication and authorization of providers and consumers of mobile edge services.

[Services-05] When necessary, the mobile edge system shall allow operators to dynamically control the access of running mobile edge applications to certain services.

[Services-06] The mobile edge platform shall provide a secure environment for providing and consuming mobile edge services when necessary.

NOTE 2: Specific services can require end-to-end mechanisms for security.

NOTE 3: A service can be provided concurrently by multiple sources. It depends on the actual service whether the multiple sources are visible to the service consumers or are consolidated and presented as a single source. This will be described as part of the service description.

[Services-07] The mobile edge platform shall allow mobile edge services to announce their availability. The platform shall allow the discovery of available mobile edge services.

[Services-08] The mobile edge platform shall provide functionality that presents the mobile edge service availability and the related interfaces to mobile edge applications.

[Services-09] The access to the information regarding mobile edge service availability and related interfaces shall only be allowed to authenticated and authorized mobile edge applications. Access to information about each mobile edge service shall be separately authorized. Separate authorization shall be possible for registering mobile edge services, and for obtaining information about registered mobile edge services.

6.2.2 Connectivity

[Connectivity-01] The mobile edge platform shall allow authorized mobile edge applications on the same mobile edge host to communicate with each other.

[Connectivity-02] The mobile edge system shall support two instances of a mobile edge application running on different mobile edge hosts to communicate with each other.

NOTE: This allows application-specific procedures to move information from one application instance to another, in order to maintain continuity of the service provided by the application as the UE moves around.

[Connectivity-03] The mobile edge platform shall be able to allow an authorized mobile edge application to communicate with another mobile edge application located on another mobile edge host.

[Connectivity-04] The mobile edge platform shall be able to allow an authorized mobile edge application to communicate with third-party servers located in external networks.

6.2.3 Storage

[Storage-01] The mobile edge platform shall be able to provide access to persistent storage space to an authorized mobile edge application.

6.2.4 Traffic routing

[Routing-01] The mobile edge platform shall provide functionality to allow authorized mobile edge applications to send user plane traffic to UEs.

[Routing-02] The mobile edge platform shall provide functionality to allow authorized mobile edge application to receive user plane traffic from UEs.

[Routing-03] The mobile edge platform shall provide functionality to route selected uplink and/or downlink user plane traffic from the network to authorized mobile edge applications.

[Routing-04] The mobile edge platform shall provide functionality to route selected uplink and/or downlink user plane traffic from authorized mobile edge applications to the network.

[Routing-05] The mobile edge platform shall provide functionality to allow authorized mobile edge applications to inspect selected uplink and/or downlink user plane traffic.

[Routing-06] The mobile edge platform shall provide functionality to allow authorized mobile edge applications to modify selected uplink and/or downlink user plane traffic.

[Routing-07] The mobile edge platform shall provide functionality to allow authorized mobile edge applications to shape selected uplink and/or downlink user plane traffic.

[Routing-08] The mobile edge platform shall provide functionality to route selected uplink and/or downlink user plane traffic from an authorized mobile edge application to another authorized mobile edge application.

[Routing-09] The mobile edge platform shall be able to select one or more applications to which the same traffic will be routed and assign priorities to them. The selection, prioritization and routing of traffic shall be based on traffic rules defined per mobile edge application.

NOTE 1: The prioritization is used to determine the routing order between the mobile edge applications.

[Routing-10] The mobile edge management shall allow the configuration of the traffic rules.

[Routing-11] The traffic rules shall allow setting packet filters based on network address and/or IP protocol.

[Routing-12] The traffic rules may allow setting packet filters based on the Tunnel Endpoint ID (TEID) and/or the Subscriber Profile ID (SPID) and/or the Quality Class Indicator (QCI) value(s).

[Routing-13] When the user plane traffic is encapsulated, then:

- the mobile edge host shall support the de-capsulation of the encapsulated (uplink) IP traffic and its routing to the authorized mobile edge applications;
- the mobile edge host shall support the encapsulation of (downlink) IP traffic received from authorized mobile edge applications before routing it to the network.

NOTE 2: IP traffic for example can be encapsulated with GTP header.

[Routing-14] The mobile edge host shall support routing user plane traffic to/from authorized mobile edge applications according to configurable parameters received from the mobile edge platform.

6.2.5 DNS support

[DNS-01] The mobile edge platform shall provide functionality that supports routing all DNS traffic received from any UE to a local DNS server/proxy.

[DNS-02] The mobile edge platform shall support configuring the local DNS server/proxy with the association of specific FQDN with IP addresses allocated to mobile edge application instances.

6.2.6 Timing

[Timing-01] The mobile edge platform shall provide a capability of supplying UTC time of day information to the authorized mobile edge applications. The information regarding time of day accuracy provided by the platform should be available to the applications.

[Timing-02] The mobile edge platform may provide authorized mobile edge applications with accurate time of specific user packets received or transmitted by the platform.

6.3 Features

6.3.1 Feature *UserApps*

[UserApps-01] The mobile edge system may support the feature called *UserApps*.

[UserApps-02] When the mobile edge system supports the feature *UserApps*, the mobile edge management shall support the instantiation of a mobile edge application on multiple mobile edge hosts following a single instantiation request.

[UserApps-03] When the mobile edge system supports the feature *UserApps*, the mobile edge management shall support the instantiation of a mobile edge application on a mobile edge host when required by the operator in response to a request by the user. The application instance needs to fulfil a number of potential constraints predefined for the application. Once instantiated, connectivity shall be established between the UE and the application instance.

NOTE 1: Potential constraints can include latency, location, compute resources, storage resources, network capability, security conditions.

[UserApps-04] When the mobile edge system supports the feature *UserApps*, the system shall, in response to a request by the user, support the establishment of connectivity between the UE and an instance of a specific mobile edge application fulfilling the requirements of the application regarding this UE. If no instance of the application fulfilling these requirements is currently running, the mobile edge system management shall create a new instance of the application on a mobile edge host that fulfils the requirements of the application. Once instantiated, connectivity shall be established between the UE and the new application instance.

NOTE 2: Requirements of the application can include latency, location, compute resources, storage resources, network capability, security conditions.

[UserApps-05] When the mobile edge system supports the feature *UserApps*, the system shall support the on-boarding of a mobile edge application during the execution of an instantiation request.

[UserApps-06] When the mobile edge system supports the feature *UserApps*, the system shall allow the establishment of connectivity between a UE and a specific instance of a mobile edge application.

[UserApps-07] When the mobile edge system supports the feature *UserApps*, the mobile edge management shall support the capability to terminate a mobile edge application instance when no UE is connected to it anymore.

[UserApps-08] When the mobile edge system supports the feature *UserApps*, the mobile edge management shall support the termination of a mobile edge application running on multiple mobile edge hosts following a single termination request.

6.3.2 Feature *SmartRelocation*

[SmartReloc-01] The mobile edge system may support the feature called *SmartRelocation*.

[SmartReloc-02] When the mobile edge system supports the feature *SmartRelocation*, the system shall support the feature *UserApps*.

[SmartReloc-03] When the mobile edge system supports the feature *SmartRelocation*, the mobile edge management shall support the relocation of a mobile edge application instance from one mobile edge host to a different host within the system.

[SmartReloc-04] When the mobile edge system supports the feature *SmartRelocation* , a mobile edge host may support the relocation of a mobile edge application instance from a different host (within the system) to this particular host, and from this particular host to a different host (within the system).

NOTE 1: Both hosts (source and target) need to support the feature *SmartRelocation* for relocation to be executed. If the ability to perform relocation is required for a mobile edge application, the mobile edge management will select a host that supports the feature *SmartRelocation* when instantiating the mobile edge application.

[SmartReloc-05] When the mobile edge system supports the feature *SmartRelocation*, the system shall be able to move mobile edge application instances between mobile edge hosts in order to continue to satisfy the requirements of the mobile edge application.

NOTE 2: Requirements of the application can include latency, compute resources, storage resources, etc.

[SmartReloc-06] When the mobile edge system supports the feature *SmartRelocation*, and based on a request from the UE, the system shall be able to relocate a mobile edge application running in a cloud environment to a mobile edge host fulfilling the requirements of the mobile edge application, and relocate a mobile edge application from a mobile edge host to a cloud environment outside the mobile edge system.

6.3.3 Feature *RadioNetworkInformation*

[RNI-01] The mobile edge system may support the feature called *RadioNetworkInformation*.

[RNI-02] When the mobile edge system supports the feature *RadioNetworkInformation*, there shall be a mobile edge service that exposes up-to-date radio network information regarding the current radio network conditions.

[RNI-03] When the mobile edge system supports the feature *RadioNetworkInformation*, there shall be a mobile edge service that provides appropriate up-to-date radio network information.

NOTE: The radio network information can be based on information received from external sources and/or generated locally.

[RNI-04] When the mobile edge system supports the feature *RadioNetworkInformation*, the radio network information shall be provided at the relevant granularity (e.g. per User Equipment (UE) or per cell, per period of time).

[RNI-05] When the mobile edge system supports the feature *RadioNetworkInformation*, the provided radio network information shall include measurement and statistics information related to the user plane. This information shall be based on information defined by 3GPP specifications.

[RNI-06] When the mobile edge system supports the feature *RadioNetworkInformation*, the provided radio network information shall include information related to UEs connected to the radio node(s) associated with the mobile edge host, their UE context and the related radio access bearers.

[RNI-07] When the mobile edge system supports the feature *RadioNetworkInformation*, the provided radio network information shall include information on changes related to UEs connected to the radio node(s) associated with the mobile edge host, their UE context and the related radio access bearers.

6.3.4 Feature *LocationService*

[Location-01] The mobile edge system may support the feature called *LocationService*.

[Location-02] When the mobile edge system supports the feature *LocationService*, there shall be a mobile edge service that provides information about the location of specific UEs currently served by the radio node(s) associated with the mobile edge host.

[Location-03] When the mobile edge system supports the feature *LocationService*, there shall be a mobile edge service that provides information about the location of all UEs currently served by the radio node(s) associated with the mobile edge host.

[Location-04] When the mobile edge system supports the feature *LocationService*, there may be a mobile edge service that provides information about the location of a certain category of UEs currently served by the radio node(s) associated with the mobile edge host.

[Location-05] When the mobile edge system supports the feature *LocationService*, there shall be a mobile edge service that presents a list of UEs in a particular location.

NOTE: Location can be geolocation, Cell ID, etc.

[Location-06] When the mobile edge system supports the feature *LocationService*, there shall be a mobile edge service that provides information about the location of all radio nodes currently associated with the mobile edge host.

6.3.5 Feature *BandwidthManager*

[Bandwidth-01] The mobile edge system may support the feature called *BandwidthManager*.

[Bandwidth-02] When the mobile edge system supports the feature *BandwidthManager* , the mobile edge platform or a dedicated mobile edge application shall enable an authorized mobile edge application to register statically and/or dynamically its bandwidth requirements and/or priority.

[Bandwidth-03] When the mobile edge system supports the feature *BandwidthManager*, the mobile edge platform or a dedicated mobile edge application may allocate bandwidth and/or assign priority to any session or to any application.

6.3.6 Feature *UEIdentity*

[UEIdentity-01] The mobile edge system may support a feature called *UEIdentity*.

[UEIdentity-02] When the mobile edge system supports the feature *UEIdentity*, the mobile edge platform shall provide functionality for a mobile edge application to register a token (representing a UE) or a list of tokens.

NOTE 1: Whether and how the mapping of IP addresses to tokens is exposed to the application is described as part of the relevant API definition.

[UEIdentity-03] When the mobile edge system supports the feature *UEIdentity*, the mobile edge platform shall allow setting packet filters for routing traffic based on a token representing the UE.

NOTE 2: The mobile edge application can obtain tokens through mechanisms that are not defined within the system.

[UEIdentity-04] When the mobile edge system supports the feature *UEIdentity*, the mobile edge platform shall allow the routing of user-plane traffic of authorized UEs to a local network (e.g. enterprise network) connected to the mobile edge host without having to route the traffic via a mobile edge application.

[UEIdentity-05] When the mobile edge system supports the feature *UEIdentity*, the mobile edge host shall support the connectivity between authorized mobile edge applications and local networks (e.g. enterprise network) connected to the host.

7 Operation and management requirements

[OAM-01] It shall be possible to control the access of a mobile edge application to mobile edge services.

[OAM-02] The mobile edge platform management shall be able to collect and expose performance data regarding the virtualisation environment of the mobile edge host related to a specific mobile edge application.

8 Security, regulation, charging requirements

[Security-01] The mobile edge system shall provide a secure environment for running services for the following actors: the user, the network operator, the third-party application provider, the application developer, the content provider, and the platform vendor.

[Security-02] The mobile edge platform shall only provide a mobile edge application with the information for which the application is authorized.

[Lawful-01] The mobile edge system shall comply with regulatory requirements for lawful interception.

[Charging-01] The mobile edge system shall allow the collection of charging-related information, log it in a secure way and make it available for further processing.

NOTE: Charging-related information can include traffic usage, application instantiation, access, usage duration, resource usage, etc.

Annex A (informative): Use cases

A.1 Use case categorization

The presence of Mobile Edge Computing at the edge of operators' networks enables a large number of new features to be developed in the mobile industry.

In order to develop the proper architecture and APIs for Mobile Edge Computing, a number of use cases are described in order to derive a consistent set of requirements, listing the capabilities that the mobile edge system needs to support to enable the features mentioned above.

Three main categories have been identified for use cases. Requirements on the architecture are generally quite similar for use cases within a category, and quite different between the categories. However, all these categories need to be supported to allow Mobile Edge Computing to enable a new era of services within the operators' networks.

The three categories are:

- **Consumer-oriented services**: these are innovative services that generally benefit directly the end-user, i.e. the user using the UE. This can include:
  - gaming;
  - remote desktop applications;
  - augmented and assisted reality;
  - cognitive assistance;
  - etc.
- **Operator and third party services**: these are innovative services that take advantage of computing and storage facilities close to the edge of the operator's network. These services are usually not directly benefiting the end-user, but can be operated in conjunction with third-party service companies:
  - active device location tracking;
  - big data;
  - security, safety;
  - enterprise services;
  - etc.
- **Network performance and QoE improvements**: these services are generally aimed at improving performance of the network, either via application-specific or generic improvements. The user experience is generally improved, but these are not new services provided to the end-user:
  - content/DNS caching;
  - performance optimization;
  - video optimization;
  - etc.

The purpose of describing these use cases is to derive useful requirements. However, some requirements are defined by design constraints and do not originate from use cases.

A.2 Mobile video delivery optimization using throughput guidance for TCP

A.2.1 Description

Category: Network performance and QoE improvements

Media delivery is nowadays usually done via HTTP streaming which in turn is based on the Transmission Control Protocol (TCP). The behaviour of TCP, which assumes that network congestion is the primary cause for packet loss and high delay, can lead to the inefficient use of a cellular network's resources and degrade application performance and user experience. The root cause for this inefficiency lies in the fact that TCP has difficulty adapting to rapidly varying network conditions. In cellular networks, the bandwidth available for a TCP flow can vary by an order of magnitude within a few seconds due to changes in the underlying radio channel conditions, caused by the movement of devices, as well as changes in system load when other devices enter and leave the network.

In this use case, a radio analytics mobile edge application, which uses services of Mobile Edge Computing, provides a suitably equipped backend video server with a near real-time indication on the throughput estimated to be available at the radio downlink interface in the next time instant. The video server can use this information to assist TCP congestion control decisions. With this additional information, TCP does not need to overload the network when probing for available resources, nor does it need to rely on heuristics to reduce its sending rate after a congestion episode.

The throughput guidance is an application-specific figure which gives the video server a hint about the bitrate that can be expected to be available for its use during the upcoming time interval. Sprecher N et al. [i.7] and Flinck H. et al. [i.3] describe the concept and its integration with TCP in more detail. Note that [i.7] and [i.3] are individual submissions which are not endorsed by IETF. It is not foreseen to standardize or support standardization of throughput guidance as part of ETSI ISG MEC.

A.2.2 Use of Mobile Edge Computing

The Throughput Guidance radio analytics application computes throughput guidance based on the required radio network information it obtains from a mobile edge service running on the mobile edge host, and uses functionality of the platform to communicate this information to the video server.

When the Throughput Guidance radio analytics application has been started, it uses a service registry functionality to find the services that provide radio network information. Based on the relevant radio network information, the application computes throughput guidance values, using an application-specific algorithm.

These throughput guidance values are then transmitted in-band to the video server, by embedding the information into the uplink data packets that are sent to the video server by the video client on the user's device. These data packets are routed through the Throughput Guidance radio analytics application based on the traffic rules related to the application.

A.2.3 Related requirements

- [Lifecycle-01]
- [Services-03], [Services-06], [Services-07], [Services-08]
- [Connectivity-01]
- [Routing-01], [Routing-02], [Routing-03], [Routing-04], [Routing-05], [Routing-06], [Routing-07], [Routing-08], [Routing-09], [Routing-10], [Routing-11]
- [RNI-01], [RNI-02], [RNI-03], [RNI-04], [RNI-05], [RNI-06], [RNI-07]

A.3 Local content caching at the mobile edge

A.3.1 Description

The fast paced development of smart phones, tablets and other handheld devices along with the success of global web based services has resulted in a significant increase in the use of mobile broadband services. The display and graphic processing technologies have evolved dramatically and high resolution video can be played on the go with handheld devices. Also wide adoption of social media enables quick and efficient sharing of the topics, which start spreading in viral fashion.

Because of viral spreading, the content is many times consumed at about the same time in the same geographical area. This creates increased pressure to ensure sufficient bandwidth, and usually the capacity in the mobile broadband network becomes a bottleneck.

This problem can be alleviated with caching the content locally, which can provide savings both in the backhaul and in the transport and at the same time improve the QoE of the consumer. Content caching has the potential to reduce the backhaul capacity requirements up to 35 % [i.2].

A mobile edge application can store locally the most popular content that is consumed in the geographical area and once requested then provide the content from the local cache. In that case there is no need to transfer the content over core network and therefore significant savings in the backhaul capacity can be achieved. In addition to capacity savings, the download times to receive the content can be greatly reduced.

A.3.2 Use of Mobile Edge Computing

Local content caching at the mobile edge host can be realized with an authorized application. The content cache application can store the content that has been identified frequently used or otherwise beneficial from the service point of view. As any application, the content cache application needs to be authorized by the platform. Content cache application can use information obtained from other applications to identify the content that could be cached. Also other criteria to decide the content to be cached can be used.

Once the content cache application receives a request for content that is stored in its local cache, the application starts directing the requested content to the user equipment, which requested the content. This results in savings in the backhaul capacity as well as improvement in QoE as content can be transferred without the additional delays caused by the core network and public internet.

A.3.3 Related requirements

- [Lifecycle-01]
- [Services-03], [Services-06], [Services-07], [Services-08]
- [Routing-01], [Routing-02], [Routing-05], [Routing-06], [Routing-07], [Routing-09], [Routing-10], [Routing-11]
- [RNI-03]

A.4 Security, safety, data analytics

A.4.1 Description

Category: operator and third-party services

This use case groups a number of innovative services for the operator or third-party vendors based on the gathering of huge amounts of data (video, sensor information, etc.) from devices analysed through a certain amount of processing to extract meaningful information before being sent towards central servers.

Applications might run in a single location (i.e. on a single host), or be spread over a given area (e.g. campus coverage) or even in the whole network. In order to support the constraints of the operator or the third party requesting the service, the applications might have to be run on all requested locations (mobile edge hosts).

This use case describes an application running on a mobile edge host close to the radio network, that receives a very large amount of information from devices and sensors connected to the radio node(s)) associated with the mobile edge host. The application then processes the information and extracts the valuable metadata, which it sends to a central server. A subset of the data might be stored locally for a certain period for later cross-check verification.

A number of service scenarios can be enabled via this use case:

- Security, safety: monitoring of an area for specific events, such as abandoned luggage, authorized access (e.g. with face recognition), car park car monitoring, etc.
- Big data: massive sensor data pre-processing, smart city, etc.

Information can be completed for example with device location tracking (see use case "Active device location tracking" in clause A.7).

A.4.2 Use of Mobile Edge Computing

A mobile edge application can either be running permanently on the mobile edge host, or based on demand from the operator, possibly in response to a request by a third-party. The application can be instantiated on a number of different hosts.

Once running, the application connects to a number of specific UEs (devices, sensors) connected to the radio node(s) associated with the mobile edge host. It then interacts with the UEs for collecting information.

The application might need to store a large amount of data locally. The data might need to survive the application instance termination.

The application performs the required (application-specific) analysis and provides the analysis results to an external entity. In order to do this, the application needs to be able to connect to external applications.

The application might need to get location information regarding UEs.

A.4.3 Related requirements

- [Lifecycle-03], [Lifecycle-04]
- [Connectivity-04], [Storage-01]
- [UserApps-01], [UserApps-02]
- [Location-01], [Location-02], [Location-03], [Location-04]

A.5 Augmented reality, assisted reality, virtual reality, cognitive assistance

A.5.1 Description

Category: consumer-oriented services

Augmented reality allows users to have additional information from their environment by performing an analysis of their surroundings, deriving the semantics of the scene, augment it with additional knowledge provided by databases, and feed it back to the user within a very short time. The device can be for example a smartphone or any wearable device with a camera and other sensors.

Assisted reality is similar to augmented reality, but its purpose is to actively inform the user of matter of interest to him (danger warning, ongoing conversations, etc.). This might be used for example to support people with disabilities (blind, deaf, of old age, etc.) to improve their interactions with their surroundings.

Virtual reality is similar to augmented reality, but its purpose is to render the entire field of view with a virtual environment either generated or based on recorded/transmitted environments. This might be used for example to support gaming implementations or remote viewing while using the most natural input device available.

Cognitive assistance takes the concept of augmented reality one step further, by providing personalized feedback to the user on activities the user might be performing (e.g. cooking, recreational activities, furniture assembling, etc.). The analysis of the scene and the advices to the user need to be provided within a very short time.

Low latency applications, such as games, AR, or VR applications, can choose to implement the rendering pipeline either in a mobile edge application on the mobile edge host or directly on the client device (such as UE).

These applications can choose to offload part of the device computational load to a mobile edge application running on a mobile edge host. This can include for example simulation of physics, artificial intelligence and other components.

For all these cases, the interaction between the user and the application needs to be personalized, and continuity of the service needs to be maintained as the user moves around.

Innovative applications are developed at a rapid pace and will evolve and be replaced in a very dynamic environment. In order to support and stimulate rapid innovation, it is necessary that new applications and new versions of applications can be provisioned dynamically, up to the point where the user requests the application to be run. If a specific application has not yet been on-boarded and the mobile edge system is able to fetch the application in a defined location, it needs to be able to do so.

Users are not necessarily going to be permanently using the mobile network environment for running their augmented reality, assisted reality and cognitive assistance applications. In some cases (e.g. in their home or at work), they might access their applications located in a cloud environment over other radio accesses, such as local Wi-Fi. However, when moving away from their static environment or going back to it, they might want to continue using the application over the mobile network environment. Applications would then need to be relocated between the external cloud environment and the mobile edge system dynamically.

A.5.2 Use of Mobile Edge Computing

In response to a request from the user, the UE needs to be connected to an instance of a specific application, running on an appropriate mobile edge host fulfilling the latency requirements of the application. A new instance of the application needs to be started if it is not yet running.

The application might have a set of requirements (e.g. latency, compute resources, storage resources, etc.) that needs to be fulfilled by the host. The mobile edge system needs to select a mobile edge host that fulfils all the requirements.

When the UE moves to an area which is not associated with the host on which the application is running, in order to fulfil the application requirements and depending on the application, either;

- the instance of the application might need to move to another appropriate host; or
- the application instance might need to transfer state information regarding the UE to another instance of the same application running on an appropriate host. To support this, the mobile edge system has to provide connectivity between these instances.

When all the users connected to a specific instance of an application have disconnected, the application instance can be terminated.

When a UE requests the mobile edge system to instantiate an application that is not already on-boarded in the system and when this is possible, the system needs to on-board the application dynamically.

The mobile edge system needs to be able to relocate applications from an external cloud environment to a mobile edge host fulfilling the requirements of the applications, and from a host to an external cloud environment, based on a request from the UE.

To support application-based distributed computation on UEs, an application can identify devices capable of supporting computation assistance and their connectivity capabilities. The application is able to determine, based on these devices' location and additional connectivity specific information, their ability to support distributed computation requests. Upon change of conditions, the application will be able to recover the connection.

A.5.3 Related requirements

- [Mobility-01], [Mobility-02], [Mobility-03]
- [Connectivity-02]
- [UserApps-01], [UserApps-03], [UserApps-04], [UserApps-05], [UserApps-06], [UserApps-07]
- [SmartReloc-01], [SmartReloc-05], [SmartReloc-06]

A.6 Gaming and low latency cloud applications

A.6.1 Description

Category: Consumer-oriented services

Games are very popular application on computers, tablets and smartphones. However, many games played on computers connected via LAN and/or broadband Internet connection require low latency values typically not available today for UEs, as the servers are usually reachable via the Internet, located beyond the RAN and Core Network.

By locating game server applications closer to the radio, a new kind of low latency-based games will become available to UE users. Of course, the use is not restricted to games, and can benefit any kind of applications requiring low latency, for example, using a "remote desktop" protocol to access cloud virtual machines, e.g. in the case the compute capacity is too high to be run efficiently on a tablet. In this case, for the user experience to be satisfactory, the latency between an action done by the user to the feedback received by the device needs to be very short.

Both general applications and games might be for either single or multiple users (e.g. equivalent to LAN gaming or shared whiteboard).

The user would start an application locally on the device that requests the connection to an application (remote desktop application, game server, etc.). For multi-user applications, the other users, via their local application, will request to access to the same instance of the game or application server. Latency constraints and other applications requirements (e.g. computing capacity, storage, etc.) will be defined for the applications. Latency constraints might include for example maximum acceptable latency, or maximum latency difference between users (e.g. in games, all users need to have a relatively similar latency to avoid unfair behaviour).

While the application / game is running, one or more users might move around, and be connected to a different radio node (handover). As this is happening, the service still needs to be provided (the connectivity between the UE and the application needs to be maintained).

As the user moves away from the original location, the latency between the UE and the application is likely to lengthen. In order to maintain the latency requirements (e.g. maximum value, fairness), the application might have to be relocated to another server.

Low latency applications, such as games or interactive applications, can choose to implement the rendering pipeline either in a mobile edge application running on the mobile edge host or directly on the client device (such as UE).

These applications can choose to offload part of the device computational load to a mobile edge application running on a mobile edge host. This can include for example simulation of physics, artificial intelligence and other components.

When all the users disconnect from the application, the application might be terminated, or frozen.

Innovative applications are developed at a rapid pace and will evolve and be replaced in a very dynamic environment. In order to support and stimulate rapid innovation, it is necessary that new applications and new versions of applications can be provisioned dynamically, up to the point where the user requests the application to be run. If a specific application has not yet been on-boarded and the mobile edge system is able to fetch the application in a defined location, it needs to be able to do so.

Users are not necessarily going to be permanently using the mobile network environment for running their gaming or low latency cloud applications. In some cases (e.g. in their home or at work), they might access their applications located in a cloud environment over other radio accesses, such as local Wi-Fi. However, when moving away from their static environment or going back to it, they might want to continue using the application over the mobile network environment. Applications would then need to be relocated between the external cloud environment and the mobile edge system dynamically.

A.6.2 Use of Mobile Edge Computing

In response to a request from the user, a new instance of a specific application needs to be started on an appropriate mobile edge host fulfilling the latency and resources requirements of the application.

In response to requests from other users, connectivity needs to be established between their UEs and a specific instance of an already running application.

The application might have a set of requirements (e.g. latency, compute resources, storage resources, etc.) that needs to be fulfilled by the mobile edge host. The mobile edge system needs to select a mobile edge host that fulfils all the requirements.

When a UE moves to another radio node associated with the same host, connectivity needs to be maintained between the UE and the application.

When a UE is connected to a radio node not associated with the same mobile edge host where the application is running (e.g. after a UE moves between radio nodes), connectivity needs to be maintained between the UE and the application. In order to maintain the latency requirements, the mobile edge management might need to relocate the application to another host, while connectivity between the application and the UE(s) is maintained.

NOTE: This functionality is required for applications such as the ones described by this use case. This functionality is not required for some of the use cases.

When all the users connected to a specific instance of an application have disconnected, the application instance can be terminated.

When a UE requests the mobile edge system to instantiate an application that is not already on-boarded in the system and when this is possible, the system needs to on-board the application dynamically.

The mobile edge system needs to be able to relocate applications from an external cloud environment to a mobile edge host fulfilling the requirements of the applications, and from a host to an external cloud environment, based on a request from the UE.

To support application-based distributed computation on UEs, an application can identify devices capable of supporting computation assistance and their connectivity capabilities. The application is able to determine, based on these devices' location and additional connectivity specific information, their ability to support distributed computation requests. Upon change of conditions, the application will be able to recover the connection.

A.6.3 Related requirements

- [Mobility-01], [Mobility-02], [Mobility-03]
- [UserApps-01], [UserApps-03], [UserApps-05], [UserApps-06], [UserApps-07]
- [SmartReloc-01], [SmartReloc-05], [SmartReloc-06]

A.7 Active device location tracking

A.7.1 Description

Category: operator and third-party services

This use case enables real-time, network measurement based tracking of active (GPS independent and network determined) terminal equipment using 'best-in-class' geo-location algorithms.

This provides an efficient and scalable solution with local measurement processing and event based triggers. It enables location based services for enterprises and consumers (e.g. on opt-in basis), for example in venues, retail locations and traditional coverage areas where GPS coverage is not available.

Services can include mobile advertising, 'Smart City', footfall analysis, campus management, crowd management, personnel management, etc.

A.7.2 Use of Mobile Edge Computing

The application can either be running permanently on the mobile edge host, or based on demand from the operator, possibly in response to a request from a third-party.

Once running, the application collects location-related information from the UEs connected to the radio node(s) which the mobile edge host is associated with. Depending on the application, specific UEs, specific categories of UEs, or all UEs need to be tracked, possibly anonymously (based on authorization).

The application performs the required (application-specific) analysis and provides the analysis results to an external entity. In order to do this, the mobile edge application needs to be able to connect to external applications.

A.7.3 Related requirements

- [Lifecycle-03], [Lifecycle-04]
- [Connectivity-04]
- [UserApps-01], [UserApps-02]
- [Location-01], [Location-02], [Location-03], [Location-04]

A.8 Application portability

A.8.1 Description

An application provider develops an application. This application can be instantiated on mobile edge hosts from different vendors without any modification.

A.8.2 Use of Mobile Edge Computing

One of the main targets of ETSI ISG MEC is to ensure the portability of mobile edge applications across mobile edge systems from different vendors. Portability in this context means that the application needs no platform specific modifications, and can be installed on every platform.

In addition to this, the mobile edge system has to verify the authenticity and integrity of the application.

The mobile edge system needs to be able to control the access of applications to mobile edge services.

A.8.3 Related requirements

- [AppEnvironment-01], [AppEnvironment-02], [AppEnvironment-03]
- [OAM-01]

A.9 SLA management

A.9.1 Description

An application provider develops a mobile edge application. This application is instantiated on a mobile edge host. The application has certain performance requirements regarding the virtualisation environment of the host and the allocated virtual resources. These requirements are typically agreed and specified in Service Level Agreements (SLAs).

In order to verify how well the SLAs are met, performance data regarding the virtualisation environment of the mobile edge host has to be collected and made available for further processing.

A.9.2 Related requirements

- [OAM-02]

A.10 MEC edge video orchestration

A.10.1 Description

Mobile Edge Computing is expected to provide excellent performance and quality while at the same time providing savings in the backhaul capacity by being able to provide content as close to end user as possible. The biggest gains can be likely achieved in a scenario with a dense population of consumers in small geographical area.

The most popular content over mobile broadband is video, which already generates over 55 % of the total traffic volume. This is greatly due to fast paced enhancements on processing and graphics capabilities of handheld devices together with the top notch services offered by the service and content providers.

Proposed use case of edge video orchestration suggests a scenario where visual content can be produced and consumed at the same location close to consumers in a densely populated and clearly limited area. Such a case could be a sports event or concert where a remarkable number of consumers are using their handheld devices to access user select tailored content. The overall video experience is combined from multiple sources including locally produced video and additional information as well as master video from central production server. The user is given an opportunity to select tailored views from set of local video sources.

The mobile edge system is ideally suited for proposed edge video orchestration due to specific characteristics of the use case. The service production and consumption take both place and strictly limited area, which also gives the best chances to control the service quality and performance.

A.10.2 Use of Mobile Edge Computing

The edge video orchestration application runs on top of a mobile edge host and uses the mobile edge services. The platform needs to provide mechanisms to connect the UEs used for local production devices (e.g. video cameras and sensors) to the video orchestration mobile edge application as well as the UEs of the consumers that use the video orchestration service.

The mobile edge platform needs to provide mechanisms to route the data traffic from the local video cameras and sensors to the video orchestration application. When UEs used for local data sources (e.g. cameras and sensors) send their data, the platform provides its services to route the data to the edge video orchestration application.

Requests from UEs to receive video are directed to the edge video orchestration application.

When a consumer selects the video stream, the edge video orchestration application sends the selected content to the user. The mobile edge platform is responsible of routing the data to the UE of the user, according to configurable rules.

A.10.3 Related requirements

- [Routing-01], [Routing-02], [Routing-03], [Routing-04], [Routing-05], [Routing-06], [Routing-10], [Routing-11]

A.11 Mobile backhaul optimization

A.11.1 Description

Category: Network performance and QoE improvements

Today there is no real coordination between the radio network and the backhaul network. When there is capacity degradation in the backhaul, the radio network is not informed about it, and vice versa, when radio network need less capacity the backhaul is not aware of it too.

It is intended in this use case to combine information from the radio network together with information from the backhaul network to optimize the resources in the backhaul (in the future optimizing resources in the radio network can also be considered).

The analytic application uses services of Mobile Edge Computing (like traffic monitoring, performance monitoring) to provide real time information about the traffic requirements of the radio network (taking into account the radio access scheduling, the application and backhaul condition).

The analytics application gets real time information from a monitoring application within the backhaul, and sends the traffic requirement to an optimization application within the backhaul network.

The optimization application can optimize the backhaul is several ways:

- shape the traffic per application at a remote aggregation point;
- reroute some of the traffic;
- increase/reduce power of microwave link based on actual capacity need.

A.11.2 Use of Mobile Edge Computing

The traffic analytics application computes throughput based on the required radio network information it obtains from mobile edge service available via the mobile edge platform and the backhaul information it obtains from the monitoring application. The traffic analytics can use the traffic monitoring service to get the user plane traffic and identify the applications that the user uses. The traffic analytics application communicates this information to the optimization application within the backhaul using an interface which is out of scope of ETSI ISG MEC.

A.11.3 Related requirements

- [Connectivity-04]
- [Routing-05]
- [RNI-04], [RNI-05], [RNI-06], [RNI-07]

A.12 Direct interaction with Mobile edge application

A.12.1 Description

Mobile edge applications might require a flow/session that is originally intended to run between the UE and some application running on the Internet to be actually setup directly between the UE and that same application now running on the mobile edge host. This requires routing traffic between the UEs to/from each of these applications in the mobile edge host.

Traffic routing as described above requires two distinct functionalities:

- At application startup: IP connection setup (i.e. IP socket setup) with the mobile edge application instead of the application server in the cloud.
- While the IP session is active: termination and decapsulation of the GTP tunnel layer providing the application with "standard" IP traffic.

In practice, satisfying the above two requirements means that at least the following functionality needs to be supported:

- DNS re-direction: for many common applications, DNS is the method by which the application server IP address is discovered and therefore DNS re-direction is the way to satisfy the required re-direction at application startup.
- GTP tunnel termination: since GTP is the protocol used by 3GPP to encapsulate bearers, satisfying the required operation while the IP session is active requires the support of GTP tunnel termination, including encapsulation and decapsulation of traffic to/from the UE.

This can include:

- A DNS cache + server functionality which can be preconfigured with DNS resolutions including the IP addresses of applications that are running on the mobile edge host.
- A monitoring capability to analyse GTP tunnelled packets and identify the destination IP address inside the IP packets.
- Capability to strip the GTP layer from the IP packets, and route these packets as IP packets to the requested application.
- Capability to receive IP packets from other applications, re-encapsulation these IP packets into the GTP tunnel and send them out towards the UE.

A.12.2 Use of Mobile Edge Computing

The UE application wants to interact with an application running on a mobile edge host (the UE application itself might be unaware that it is interacting with an application running in a Mobile Edge Computing environment).

In order to do this, it first sends a DNS request for an FQDN to get the IP address of the application.

As part of the traffic handling functionality, the mobile edge host retrieves the IP packet from the user plane traffic, possibly after decapsulating the GTP bearer if necessary. The DNS traffic (i.e. port 53) is routed to the DNS server/proxy running on the host.

If the DNS server/proxy can resolve the FQDN locally, either because the FQDN was preconfigured or because it is cached from a previous request and is still valid, then it replies to the request. Otherwise it might receive the response from the core network to which it could have forwarded the query. It then sends back the DNS response it receives from its own request or from the DNS server in the network.

The traffic handling functionality then sends the IP packet containing the answer back to the UE, possibly after encapsulating it in the GTP bearer if necessary.

The UE then interacts with the mobile edge application running on the mobile edge host using the IP address provided by the DNS server/proxy.

The traffic handling functionality retrieves IP packets with the destination IP address of the mobile edge application, possibly after decapsulating the GTP bearers if necessary.

The traffic from the application is then sent back to the UE, possibly after encapsulating it in the GTP bearer if necessary.

A.12.3 Related requirements

- [DNS-01], [DNS-02]
- [Routing-13], [Routing-14]

A.13 Traffic deduplication

A.13.1 Description

Category: Network performance and QoE improvements

The traffic in the network today contain a lot of content that arise from content providers like video content providers, etc. Studies show that a lot of the traffic repeat itself since many users consume same traffic, for example many users are upgrading apps on their smart phone, or looking at the same video clip.

Traffic deduplication is a technique is made of two functions, compressor and decompressor [i.5]. The compressor and decompressor identify repeating patterns in the traffic, which are stored close to the user at the decompressor. The compressor identify the pattern as well and instead of sending them, it sends just indexes that identify the specific pattern stored at the decompressor [i.6]. This technique reduces significantly the amount of traffic send for a repeating pattern, and overall studies show reduction of more than 30 % of the overall traffic in the segment between the compressor and decompressor. In this use case the decompressor application uses services of Mobile Edge Computing like traffic routing to perform in real time the detection of the repeating pattern, storing them and reconstruct the original content. It is important to note that if there are other applications on the platform that use the same traffic flows, this application needs to be performed first to allow the other applications to see the original traffic (and not the indexes).

A.13.2 Use of Mobile Edge Computing

When De-Compressor application starts, it interacts with the mobile edge platform to enable traffic routing.

The De-Compressor application starts a communication channel with the Compressor application and coordinates with it the detection of repeated pattern and their storage.

The compression application is responsible for detecting repeated patterns and sends just indexes of those pattern to the de-compressor. The De-Compressor application is responsible to regenerate the original traffic from the received indexes.

A.13.3 Related requirements

- [Connectivity-04]
- [Routing-03], [Routing-04], [Routing-09]

A.14 Vehicle-to-infrastructure communication

A.14.1 Description

Category: Operator and third party services

Communication of vehicles and roadside-sensors with a roadside unit is intended to increase the safety, efficiency, and convenience of the transportation system, by the exchange of critical safety and operational data.

The roadside application incorporates algorithms that use data received from vehicles and roadside sensors to recognize high-risk situations in advance, and send alerts and warnings to the vehicles in the area. The drivers of the vehicles can immediately react, for example by avoiding the lane hazard, slowing down or changing the route.

LTE™ can significantly accelerate the deployment of car-to-roadside communications. LTE™ cells can provide "beyond the horizon" visibility in the critical 2 km range. Cars can leverage their increasingly inbuilt LTE™ connectivity.

Vehicle-to-Infrastructure communication has tight latency requirements, below 10ms in some use cases. Messages, such as hazard warnings (e.g. accident, danger on lane, etc.), could be distributed in real time over LTE™, eliminating the need to build a countrywide, Digital Short-Range Communications (DSRC) network. In deployments where DSRC exists, the LTE™ would be able complement it.

Mobile Edge Computing can be used to extend the connected car cloud into the highly distributed mobile network environment. Applications can be deployed on mobile edge hosts to provide the roadside functionality. The roadside applications can receive local messages directly from the applications in the vehicles and the roadside sensors, analyse them and then propagate (with extremely low latency) hazard warnings and other latency-sensitive messages to other cars in the area. This enables a nearby car to receive data in a matter of milliseconds, allowing the driver to immediately switch lanes, slow down or change his route.

The roadside application can inform other applicable mobile edge applications running on other mobile edge hosts about the event(s), and in so doing, enable the propagation of hazard warnings to cars that are close to the affected area.

The roadside application can send local information to the applications at the connected car cloud for further processing.

A.14.2 Use of Mobile Edge Computing

The roadside Mobile edge application uses functionality of the mobile edge platform to receive data from applications in the vehicles or the roadside sensor that can be used to recognize high-risk situations in advance. It uses the computing resources to analyse the data and then utilizes the functionality of the platform to propagate (with extremely low latency) hazard warnings and other latency-sensitive messages to other cars.

When the roadside application starts, it uses the mobile edge platform to enable traffic routing. Based on the data received from the vehicles, the roadside application uses application-specific algorithms to recognize high-risk situations in advance. The application generates warnings messages and sends them to nearby cars with extremely low latency.

The roadside application performs the required (application-specific) analysis and provides the analysis results to an external entity (either on an adjacent mobile edge host or at a connected car cloud). In order to do this, the application needs to be able to connect to external applications.

A.14.3 Related requirements

- [Services-03], [Services-06], [Services-07], [Services-08]
- [Connectivity-03], [Connectivty-04]
- [Routing-01], [Routing-02], [Routing-03], [Routing-04], [Routing-05], [Routing-06], [Routing-10], [Routing-11], [Routing-13]

A.15 Location-based service recommendation

A.15.1 Description

Category: User-oriented services

Users' activities usually largely depend on where users are. For instance, users at shopping mall probably buy some goods and users at a museum would be interested in a specific art work. In that sense, if some user services related to users' context are to be recommended, users' location can be of great significance. Location information exactly tells that users are in certain place at certain moment.

Location based service recommendation allows a variety of services which are tightly coupled with a specific place such as shopping mall, museum, etc. to be recommended to users at the right time. In order to improve QoE with respect to usefulness of recommended user services, users' behaviour log and preference can be collected in users' mobile terminal, and then be delivered to the location based service recommendation application, and also data from sensors around users can be delivered to service recommendation application.

Location can be detected as users are connected to a specific cell. Attachment information of user terminals to a specific radio node can be used to get location information of users especially when users are connected to an indoor small cell. In that sense, location information would represent the proximity to a certain place.

The user service recommendation application can make use of machine learning and/or inference engine to determine proper services for users at the moment. The service recommendation can be done in cooperation with big data analysis which is fulfilled by backend server in internet. Services can be represented in a variety of form. For instance, at a department store, coupons can be recommended or some of goods can be recommended. On the other hand, at the museum, video clips can be recommended for a specific art piece.

A.15.2 Use of Mobile Edge Computing

The application can run on an appropriate mobile edge host and use some of the mobile edge services provided via the mobile edge platform that fulfils the requirements of the application in order to respond the user's request. These can include requirements on latency, computing resources, response time. The mobile edge system needs to select a host fulfilling all the requirements. A new instance of the application might be created by the first request of the user, or can be created prior to any request from users.

The user's request can be routed to an appropriate mobile edge application, and also the mobile edge application might be able to send request to a specific user through the mobile edge host. The application might be connected to the external servers through the mobile edge host.

A.15.3 Related requirements

- [Location-05]

A.16 Bandwidth allocation manager for applications

A.16.1 Description

Different mobile edge application running in parallel on the same mobile edge host might require specific static/dynamic up/down bandwidth resources. In some cases different sessions running in parallel in the same application can each have specific bandwidth requirements. As all these applications and application sessions are competing over the same shared bandwidth resources, it is suggested that a central bandwidth resource allocator exists on the mobile edge host, preferably on the mobile edge platform. The proposed function can include the following:

- an API enabling all registered application to statically and/or dynamically register for specific bandwidth allocation;
- an interface with the radio network information service to receive network conditions and available bandwidth;
- the capability to calculate optimal bandwidth allocation per session/application according to available and required bandwidths;
- the capability to manage the allocated bandwidth to each of the sessions/applications according to the calculations.

A.16.2 Use of Mobile Edge Computing

The mobile edge host can support applications running in parallel. These applications will be running over the same hardware and network resources, with each having its own requirements.

In order to assure that all the applications/sessions are receiving the bandwidth resources they require in an optimal way, the mobile edge platform can include functionality that will collect required bandwidth resources and available bandwidth resources and allocate bandwidth to each session/application according to static/dynamic requirements.

A.16.3 Related requirements

- [Bandwidth-01], [Bandwidth-02], [Bandwidth-03]

A.17 Mobile edge platform consuming information from operator trusted mobile edge application

A.17.1 Description

Category: operator and third-party services

This use case allows an application to target a specific subscriber or a group of subscribers. For example:

- allowing an anonymous group of flat rate billing subscribers access content locally from the mobile edge host;
- targeted advertising for a certain group of users within the mobile network;
- providing content to a specific group of users that might be in the same club, association, public service group, etc.;
- providing enterprise services to company employees.

The ability to target a particular subscriber or a group of subscribers from the mobile edge platform can add significant value to the Mobile Edge Computing offering. In order to do this within the network, the mobile edge host needs to be capable of routing traffic to the mobile edge application based on UE IP address rather than destination IP address. The mapping of UE IP address to subscriber is not available on the mobile edge platform. This information needs to be provided by an external source (e.g. in the core network). In this use case an operator trusted mobile edge application (see annex B) receives this information from an external source and provides this information to the platform.

Figure A.17.1 illustrates an example use case for this concept.

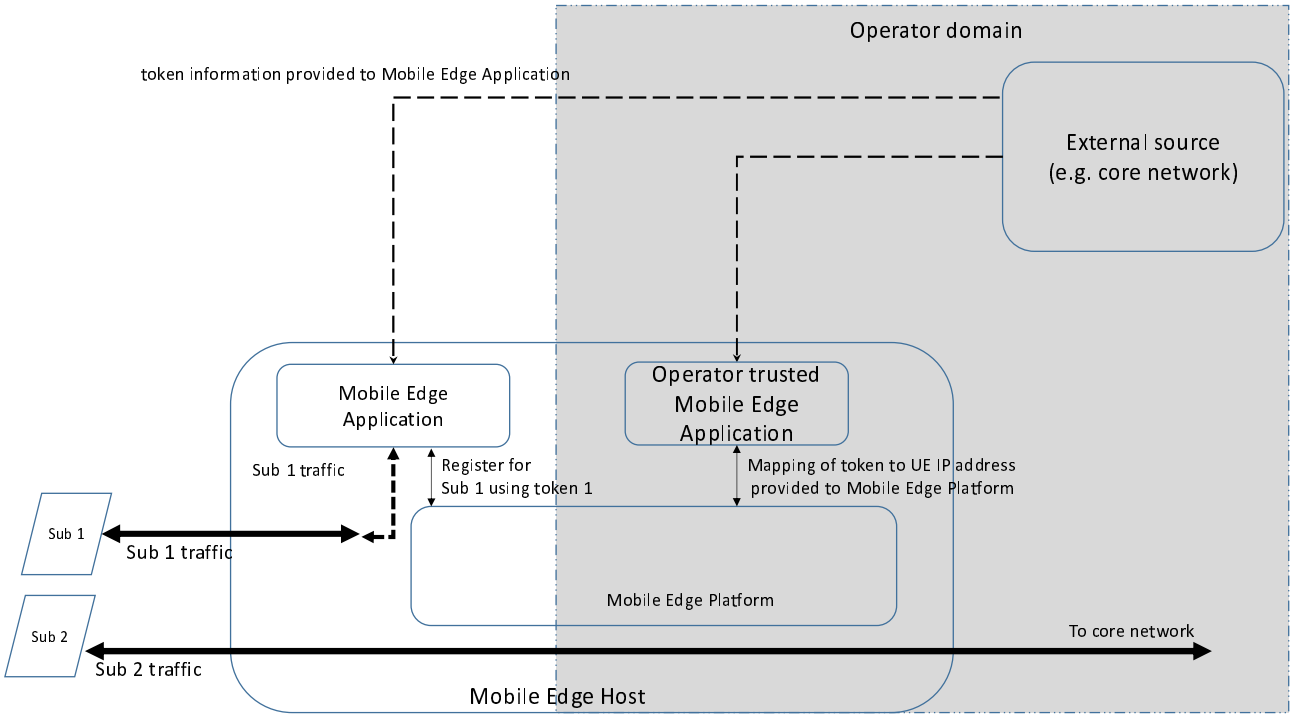

Figure A.17.1: Subscriber based routing

After being provided the subscriber token information, the application registers to receive traffic associated with a subscriber, in this example Subscriber 1. The mobile edge platform registers this information. On receiving information associated with Subscriber 1, the platform triggers the routing of all application traffic to the application as illustrated above. The traffic reference (UE IP address) can change periodically, which will be updated via the operator trusted application, but the reference provided to the application will remain the same.

A.17.2 Use of Mobile Edge Computing

The mobile edge platform will support receiving information from an authorized application and using this information to provide a service to other applications.

A.17.3 Related requirements

- [Services-02]

A.18 Video caching, compression and analytics service chaining

A.18.1 Description

Consider the use case where traffic which is sent from the content caching application to a UE is steered first through the video compression application and then through the video analytics application.

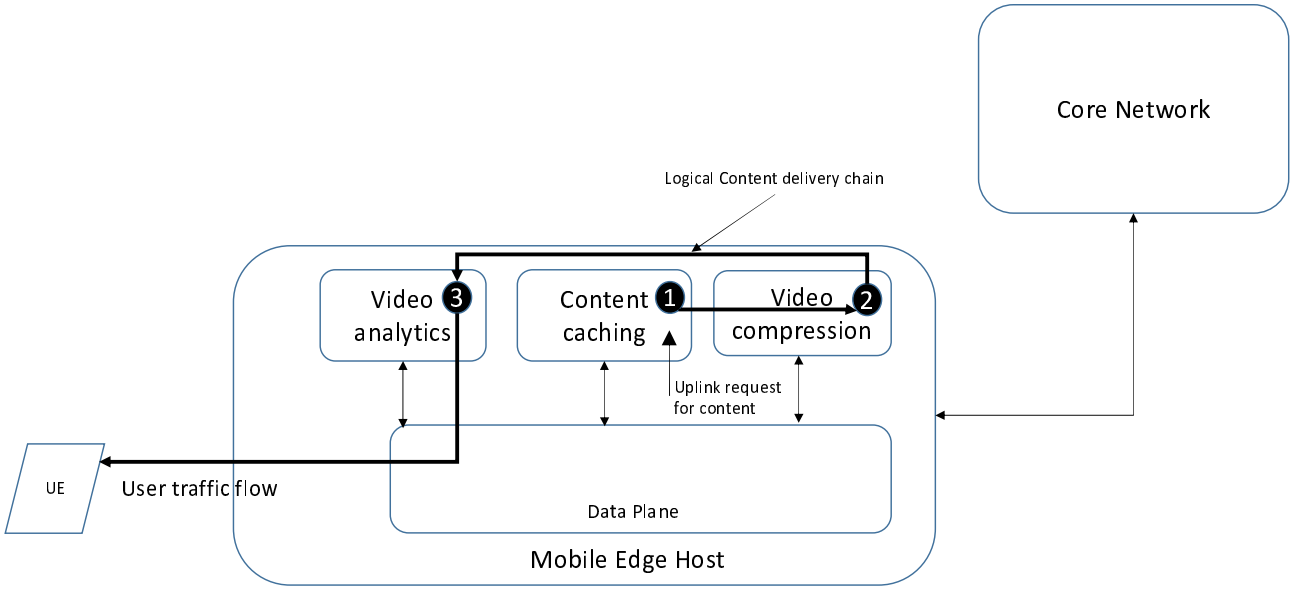

Figure A.18.1: Video content delivery

When the uplink request arrives at the mobile edge platform, the content request is routed to the Content caching application in order to retrieve the content.

Once the content is identified, the user traffic needs to be passed to the Video compression and Video analytics application before it be delivered to the end user.

The platform needs to support this scenario, whereby it will classify the traffic and then steer the traffic through multiple applications.

A.18.2 Use of Mobile Edge Computing

The mobile edge host is capable of hosting multiple applications.

Based on policy, classified traffic needs to be steered through multiple applications in a particular sequence.

A.18.3 Related requirements

- [Routing-09]

A.19 Radio access bearer monitoring

A.19.1 Description

Category: Network performance improvements

A UE can have multiple dedicated bearers in addition to the default bearer. A dedicated bearer is used to tunnel one or more specific traffic types (e.g. VoLTE, video, etc.). One of the dedicated bearers can be used for example to tunnel traffic that requires low latency and is assigned with a QCI value of 1. Another bearer can be used for high throughput traffic and is assigned with a QCI value of 9. The QCI value is used within the radio access network to control packet forwarding treatment.

Having the ability to route traffic based on bearers' information allows monitoring services on a particular bearer.

An application running on a mobile edge host can perform monitoring of traffic on specific bearers or bearers with specific QCI values.

A.19.2 Use of Mobile Edge Computing

User plane traffic which is tunnelled through a specific bearer and which can also be assigned with a particular QCI value is routed to a mobile edge application for monitoring services.

NOTE: This refers to the application level traffic rather than the transport level traffic.

A.19.3 Related requirements

- [Routing-12]

A.20 Mobile edge host deployment in dense-network environment

A.20.1 Description

Category: Network Performance and QoE improvements

This use case considers how to maintain the low latency requirement for gaming and new user experience under deployments intended to mitigate wireless network congestion.

Due to the nature of the wireless access network, the capacity of radio fluctuates based on numbers of users and types of applications. Mobile Edge Computing can participate to the definition of solutions for wireless network capacity issues. Inter-UE communication such as 'Device to Device Communication' can be used by applications deployed on UEs to help resolve specific situations of network congestion by limiting network traffic in addition to fulfilling low-latency requirement.

Another deployment alternative is the use of Relay Nodes. In that case, the mobile edge host can be deployed on Relay Nodes. The mobile edge system can manage these hosts on Relay Nodes similarly to other mobile edge hosts, allowing the system to have further options to fulfil the set of application requirements (e.g. latency, compute resources, storage resources, throughput, etc.).

A.20.2 Use of Mobile Edge Computing

In order to identify wireless network congestion, a mobile edge service available via the mobile edge platform provides radio network information to a dedicated application. When network congestion is identified, the mobile edge application can communicate with counterpart applications running on devices, to request them to activate direct device-to-device communication network capabilities through application-specific means.

If radio nodes such as Relay Nodes are deployed, a mobile edge application can collect radio network information including radio node location information and perform certain application-specific tasks to distribute optimally the processing of applications at the different locations.

Once the mobile edge application receives radio network information indicating that network congestion has been mitigated, it can deactivate application-specific support through direct device-to-device communication. Interaction resumes between UE applications and the corresponding mobile edge applications running on mobile edge hosts associated with their respective radio nodes.

The radio network information enables the selection of appropriate mobile edge hosts in order to fulfil a set of requirements (e.g. latency, compute resources, storage resources, throughput, etc.) of mobile edge applications.

A.20.3 Related requirements

- [Location-06]

A.21 Radio network information generation in aggregation point

A.21.1 Description

As different mobile edge applications might require radio network information to enable them to run in a more optimal way, a radio network information service is defined. In the scenario in which the mobile edge host is located at the radio node site it is understandable this information can be received by the mobile edge service from the radio node. When placing the mobile edge host in an aggregation point (and RNC) or in a Core Network gateway in which the radio network information is not available locally, it is more complicated to extract the required information.

This use case proposes to enable the generation of the required information independently also in mobile edge hosts located in other deployments such as an aggregation point or gateway.

In order to achieve the most accurate measurement for the radio network information, the mobile edge platform or a dedicated mobile edge application that provides the service can take into the calculation several parameters extracted from the traffic going over the network. The service can also involve a self-learning method to adapt the end results to changing conditions at each location and different radio nodes, which can be used by applications that register for this information and are authorized to receive it.

A.21.2 Use of Mobile Edge Computing

A number of applications running on a mobile edge host might require radio network information to optimize their performance. A mobile edge service could provide such information to such authorized applications, even in deployments where radio network information might not be available to the mobile edge service, by generating this information autonomously.

A.21.3 Related requirements

- [RNI-03]

A.22 Unified enterprise communications

A.22.1 Description

This description is based on SCF081 [i.8]. Mobile devices are gradually replacing fixed communications hardware, laptop software, and office services in the enterprise market by leveraging native desktop interfaces as well as additional value-added apps. This is a replication of trends in the consumer market. Since about 2010, the consumer market has seen many devices and physical world technologies replaced by smartphone and tablet hardware and software and internet cloud platforms.

Once robust coverage and capacity are available indoors, the enterprise can start to move towards a truly mobile office where the business tools are migrated into the mobile devices and there is ubiquitous access to cloud-based business tools. The presence of a mobile edge deployment of Small Cells on enterprise premises makes it a natural candidate for support of enterprise applications in the mobile edge.

Figure A.22.1 describes a Mobile Edge Computing-based breakout to an enterprise network, enabling employees using smartphones and tablet PCs to enjoy a fast broadband connection directly to the enterprise LAN.

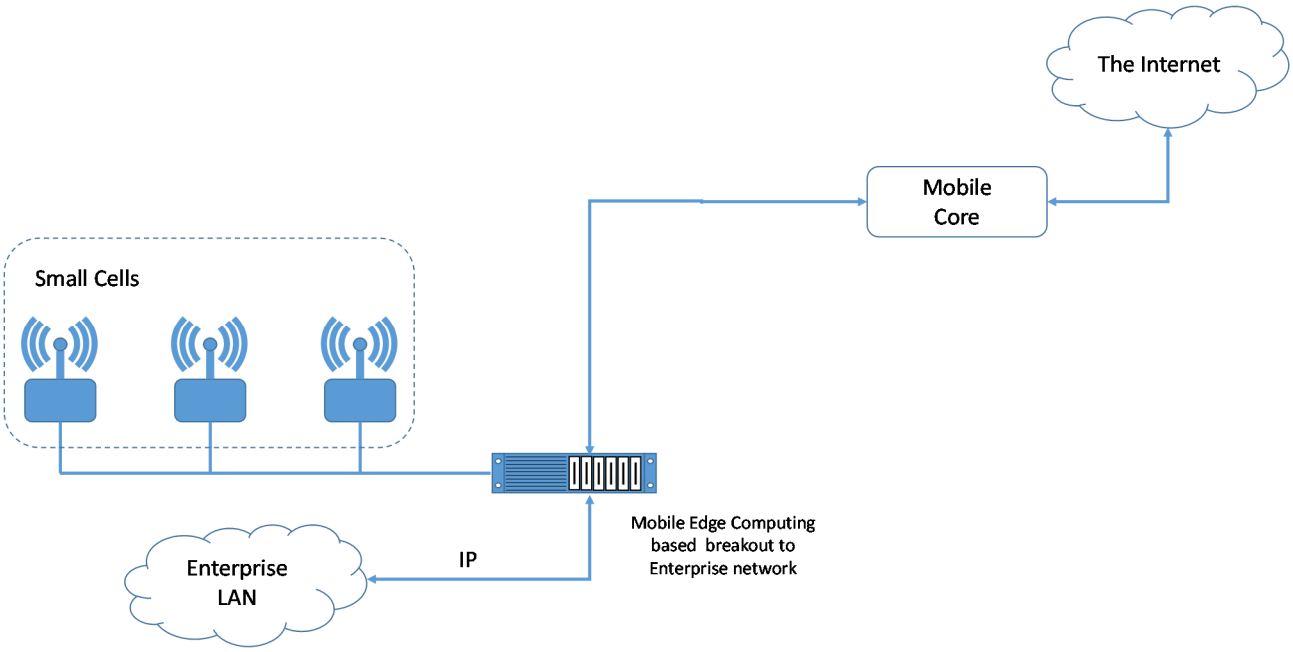

Figure A.22.1: Mobile Edge Computing based breakout to an enterprise network

An example of such a unified service which is in direct market demand today is that of unified communications with the enterprise PBX, allowing the enterprise user's BYO Device's to be used for enterprise communications.

Based on the discussion in SCF081 [i.8], an example (and very incomplete) set of features required for the unification of Mobile Edge Computing with PBX is as follows:

- **internal call re-routing:** support of routing and processing of IP-PBX (internal extension call) traffic for enterprise employees;
- **identity:** association of the user's enterprise identity with the mobile network traffic to support traffic rules based in the enterprise users identity;
- **time of day routing:** the capability to set rules/policies so that UE traffic can be handled in different ways based on the time of day;
- **enterprise messaging:** support selective re-routing of mobile messaging application traffic based on enterprise user's identity for the purposes of integration of enterprise IM and SMS, for example, to implement enterprise paging features.

A.22.2 Use of Mobile Edge Computing

The mobile edge platform provides functionality which facilitates the association of IP traffic flows with a particular UE to which an external network identifier (e.g. via active directory) is associated using an externally defined token. This association is performed using industry-standard techniques that preserve user privacy and the secrecy of the identity information at both the mobile network and the enterprise.

EXAMPLE: The service could involve facilitation of the use of a trusted 3rd party identity management service (e.g. an OpenID provider) with which the operator and the external network (e.g. enterprise) have the necessary trust relationships in place.

The mobile edge platform provides the capability to route user plane traffic of enterprise users between the operator network and the enterprise network without having to pass through the application. The application is responsible for managing issues associated with access control, integrity, etc. of the user content.

A.22.3 Related requirements

- [UEIdentity-01], [UEIdentity-02], [UEIdentity-03], [UEIdentity-04], [UEIdentity-05]

A.23 Application computation off-loading

A.23.1 Description

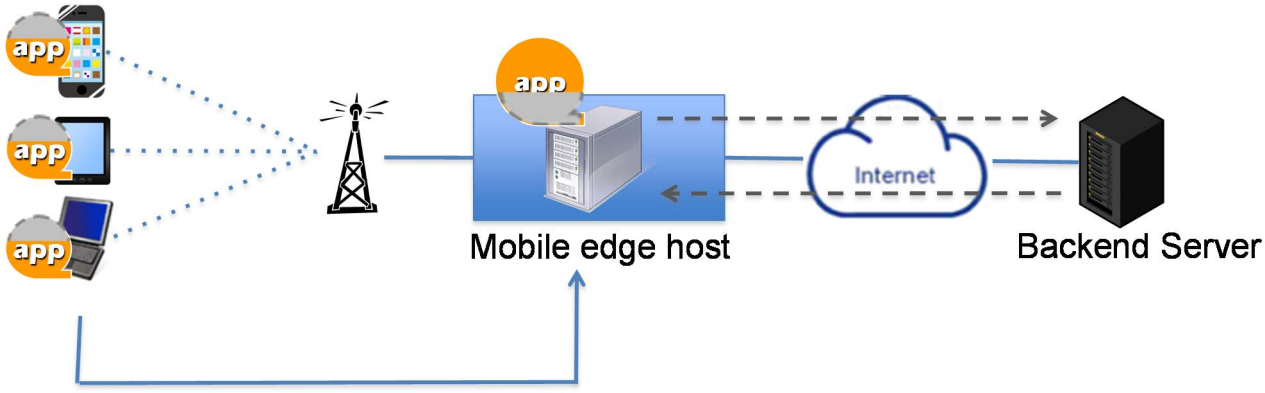

Figure A.23.1: Application computation off-loading using Mobile Edge Computing

In the application computation off-loading use-case, the mobile edge host executes compute-intensive functionalities with high performance instead of mobile devices as shown in figure A.23.1. By providing rich computation resources on a mobile edge host, application computation can be off-loaded to the mobile edge host to be accelerated even if a user uses relatively low performance devices, and user experience can be satisfied regardless of the type of UE.

This use-case is effectively used for especially computation-hungry applications such as graphical rendering (high-speed browser, artificial reality, 3D game, etc.), intermediate data-processing (sensor data cleansing, video analysing, etc.), and value-added services (translation, log analytics, etc.). One example of application computation offloading is the Edge Accelerated Browser (EAB). Most parts of the browsing functions, such as Web contents evaluation, rendering and optimized transmission, are off-loaded to the mobile edge application, while the UE just renders reconstituted browser graphics on its display. This can transfer a compute-intensive process from a UE to a mobile edge host to accelerate an application and make rich applications available on various types of mobile devices.

A.23.2 Value proposition

Application computation off-loading provides the following values to both service provider and end users.

- service providers:
  - deliver high-performance applications regardless of end device capability;
  - off-load computation resource required by an application by using rich compute resource in mobile edge host;
  - process application data collected from end devices (M2M, video, etc.) on mobile edge host to mitigate both end device and centre server load;
  - provide value added services by adding extra value to application data on mobile edge host.
- end users:
  - improve user experience through an off-loaded application on their mobile devices;
  - get low cost end devices by off-loading compute capacity to mobile edge host;
  - get new types of service offerings like auto-translation and recommendation based on log analytics by application linkage on mobile edge host.

A.23.3 Use of Mobile Edge Computing

In response to a request from the user, a new instance of a specific application (or part of application function) needs to be started on an appropriate mobile edge host fulfilling the latency and resources requirements of the application.

In response to requests from other users, connectivity needs to be established between their UEs and a specific instance of an already running application.

The application might have a set of requirements (e.g. latency, compute resources, storage resources, location, network capability, security condition etc.) that needs to be fulfilled by the mobile edge host. The mobile edge system needs to select a host fulfilling all the requirements.

When all the users connected to a specific instance of an application have disconnected, the application instance might be terminated.

A.23.4 Related requirements

- [Connectivity-04]
- [UserApps-02], [UserApps-03], [UserApps-04], [UserApps-07], [UserApps-08]

Annex B (informative): Operator trusted mobile edge applications

Operator trusted mobile edge applications can be viewed as extensions of the mobile edge platform functionality, such as allowing a tighter integration with the core network, by providing services to the mobile edge platform.

In addition to the mobile edge application capabilities, the operator trusted mobile edge applications will have advanced privileges to provide information to the mobile edge platform securely.

Due to the privileged role, it is required that such applications and the platform are mutually authenticated and authorized and that the communication between these entities cannot be eavesdropped on by third parties.

The role of the operator trusted mobile edge applications is not specifically to report on or modify the performance of the radio network but also to extend the functions of the existing trusted core network functions. The interface between such application and the core network elements are external to the platform. Some examples might be:

- distributed policy control;
- distributed quality of service;
- providing information on transport tunnel allocation.

History

Document history		
V1.1.1	March 2016	Publication