



ETSI White Paper #36

Harmonizing standards for edge computing - A synergized architecture leveraging ETSI ISG MEC and 3GPP specifications

1st edition – July 2020

ISBN No. 979-10-92620-35-5

ETSI
06921 Sophia Antipolis CEDEX, France
Tel +33 4 92 94 42 00
info@etsi.org
www.etsi.org



Authors

Nurit Sprecher <i>Nokia</i>	Jesus Martin <i>Telefónica</i>
Omar Elloumi (Editor) <i>Nokia</i>	Jane Shen <i>Futurewei</i>
Pekka Kuure <i>Nokia, Vice Chair of ETSI ISG MEC</i>	Jiangping Feng <i>Huawei</i>
Adrian Neal <i>Vodafone</i>	Juan Carlos Garcia <i>Telefónica</i>
Alan Soloway <i>Qualcomm Technologies Inc.</i>	Kenichi Yamamoto <i>KDDI</i>
Alex Harmand <i>Telefónica</i>	Hyeonsoo Lee <i>SK Telecom</i>
Suresh Chitturi <i>Samsung, Chair of 3GPP SA6</i>	Mona Mustapha <i>Apple</i>
Alex Reznik <i>HPE, Chair of ETSI ISG MEC</i>	Qian Huang <i>China Unicom</i>
Alice Li <i>Vodafone</i>	Ralf Rossbach <i>Apple</i>
Cristina Costa <i>FBK</i>	Rong Huang <i>China Unicom</i>
Dario Sabella <i>Intel, Vice Chair of ETSI ISG MEC</i>	Sami Kekki <i>Huawei, Vice Chair of ETSI ISG MEC</i>
Danny Moses <i>Intel</i>	Shamik Mishra <i>Altran, Vice Chair of GSMA OPG</i>
Deepak Gunjal <i>Altran</i>	Uli Kleber <i>Huawei</i>
Eddy Hall <i>Qualcomm Technologies International, Ltd</i>	Walter Featherstone <i>Samsung, Technical Manager of ETSI ISG MEC, Chair of ETSI MEC DECODE WG</i>
Gao Chen <i>China Unicom</i>	Mingeun Yoon <i>SK Telecom</i>
Himanshu Singh <i>Altran</i>	



Contents

Authors	2
Contents	3
1. Introduction	4
2. Considerations when deploying applications at the edge of mobile networks	6
3. A possible reference architecture leveraging existing/ongoing standards	8
3.1 Introduction	8
3.2 Synergized Mobile Edge Cloud architecture	10
3.3 Examples of deployments	11
4. Conclusions and recommendations	13



1. Introduction

The Information and Communications Technology (ICT) industry has been debating edge computing¹ for many years. With a growing market demand stemming from applications requiring low latency and recent developments in standards and products, the opportunity to deploy edge computing at scale is here to stay, fuelling substantial market growth. Mobile network infrastructures offer a fast track deployment path for edge cloud infrastructures. There are two factors contributing to this trend:

- **On the demand side**, with Industrial IoT going mainstream, there are several commercially viable use cases which call for processing data at the edge to address stringent latency requirements and avoid overloading networks. Other applications calling for edge computing capabilities include Augmented Reality (AR) / Virtual Reality (VR), gaming and Vehicle-to-everything (V2X) communication (such as Vehicle-to-Vehicle, Vehicle-to-Infrastructure or Vehicle-to-Pedestrian) – all are currently gaining momentum in the marketplace.
- **On the supply side**, building a distributed cloud computing infrastructure is proven to be a viable business opportunity, thanks to geographically dense points of presence of Mobile Network Operators (MNOs). Cloud providers are also looking for opportunities and partnerships to build edge cloud infrastructures. MNO points of presence are unique in addressing the proximity requirements of most demanding use cases with deployment options ranging from deep and far edge (up to 5km and 10km from end user resp.) to aggregated edge (up to 30km). With 5G being extensively deployed in several geographies, new features such as Ultra-Reliable Low Latency Communications (URLLC) and massive Machine Type Communication (mMTC) call for complementary edge computing capabilities to realize the full market potential of 5G.

Today, deploying cloud computing at the edge is a market and a business imperative. The remaining questions are **how quickly will we get there? what use cases are emerging? which requirements can be derived from those use cases? what are the remaining roadblocks the industry needs to resolve to help the market further develop at scale? and last but not least, which standards and open source projects will help the market further develop?**

Let's examine some use cases that call for building edge computing capabilities combined with mobile networks' infrastructures.

Industrial IoT: Industry 4.0 is wide in scope and in the number of use cases it encompasses. Because of the low latency requirements introduced by these use cases and their need for trusted computing, using central clouds is oftentimes not an option anymore, leaving edge cloud as the only viable option. With 5G being deployed in public and private networks, combining 5G (and 4G) with edge cloud is a natural architecture option to support Industrial IoT.

Automotive: SAE International² defined several levels of automation leading to fully automated driving. There is a common consensus that reaching the highest levels of automated driving will call for connectivity and a high degree of cooperation, both between vehicles, on one hand, and between vehicles and highly digitized transportation infrastructure, one that supports high throughput and low latency communications and processing. Edge Computing will play an instrumental role in automated driving, with some advanced

¹ ETSI whitepaper on "MEC in 5G Networks"

(https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf) considers edge computing as an evolution of cloud computing brings application hosting from centralized data centres down to the network edge, closer to consumers and the data generated by applications.

² <https://www.sae.org/>



autonomous driving use cases requiring densely distributed edge computing capabilities. The 5G Automotive Association (5GAA)³ is also backing the vision of 5G combined with edge computing capabilities to support V2X applications.

AR/VR: Augmented and Virtual Reality provide rich and immersive experience to consumers and businesses. Despite their game changing potential, AR/VR services are still not mainstream yet because of limited user experience and high devices cost. Edge computing is expected to provide the capability to offload compute to a low-latency edge cloud – hence driving creation of low-cost and mass market devices. Enhancing AR/VR with edge computing provides the level of latency that would dramatically change the user experience and bring AR/VR into the mainstream.

Gaming: Mobile and gaming are increasingly two facets of the same coin. With an ever-increasing demand for immersive experience and interactivity; edge computing will be transformational for the gamers and the future of the mobile gaming industry alike. Its value is in providing a ubiquitous gaming experience without being tied to a specific and possibly high-cost and specialized hardware.

“There are two primary drivers for edge computing, lower latency and backhaul optimization. The killer app for edge services requiring lower latency is multi-player online gaming. Video resolution of 4 K and 8 K for AI-assisted video analytics and Augmented Reality/Virtual Reality (AR/VR) will drive up bandwidth, pushing the limits of a connectivity provider’s backhaul system. With edge computing, local data can be processed at the edge without transporting the data between the user and a centralized data center/core.”, DellOro report “Advanced Research Report – Multi-access Edge Computing”, Jan 2020

What is the role of standards and open source when edge computing is deployed in conjunction with mobile networks?

Cloud computing has existed for some time; therefore, a legitimate question is “what is the role for standards when deploying cloud computing at the edge in conjunction with mobile networks?”

Standards are often required when there is a need to build solutions involving multiple stakeholders. They are also beneficial to realize economies of scale, avoid lock-in and enable multi-vendor solutions leveraging best of breed components from each provider to be built. Examples of where standards are specifically required when deploying edge computing in conjunction with MNO infrastructures include:

- **Common infrastructure capabilities** to help developers/customers access and deploy their software (e.g. applications, platforms, etc.) over the edge infrastructure of the MNOs.
- **Smart application placement** to allow the optimized deployment of applications at the edge infrastructure based on criteria such as available resources, geographical areas, cost and latency requirements.

³ https://5gaa.org/wp-content/uploads/2017/12/5GAA_T-170219-whitepaper-EdgeComputing_5GAA.pdf



- **Discovery of, and optimal (re)-routing** to an edge cloud capable of serving application clients (running on devices). When an application client wants to reach a server application, there's a need to discover the optimal edge cloud, one which runs instances of the server application, has the necessary resources (CPU, GPU, etc.) and provides the lowest network latency.
- **Service continuity:** when mobility occurs, it will be beneficial to seamlessly transfer the context from a stateful application instance in an edge cloud to an instance of the same application in a target edge cloud, e.g. one that provides a lower latency.
- **Cloud applications would enhance the user experience if they could leverage services offered by the network:** accessing information and services provided by the edge services such as the device location or QoS are key to enhance the user experience.
- **Edge federation across multiple MNOs:** allowing MNOs to offer to the developers/customers the possibility to deploy their software across multiple domains and ensure service continuity when roaming on alternative networks.

In order to offer these capabilities interactions between the edge cloud middleware (enabling applications running on the edge), the applications (whether running on devices or on the edge cloud) and the networks are required. Related to these interactions, this whitepaper highlights the relevant standards initiatives, their value propositions and their complementary natures.

Open source also plays an important role in shaping and accelerating deployments of edge clouds. A variety of open source projects pertaining to the cloud in general will equally apply to edge clouds. These include ones that offer high throughput, low latency, high availability, horizontal scalability, etc. Additionally, other open source initiatives are emerging to specifically support ongoing standards for edge cloud deployments in conjunction with mobile networks. Their value propositions will be about supporting developers and expediting the time to deployment.

Vertical specific open source components can also enhance edge cloud deployments through offering specific microservices that can be used by application developers through APIs.

2. Considerations when deploying applications at the edge of mobile networks

There is no one size fits all when it comes to deploying applications at the edge of a mobile network. There could be different **modes of operations** depending on the application and user needs. The text below highlights some important modes of operations which are of immediate relevance for ongoing and future standards and deployments of edge clouds in conjunction with mobile networks.

Applications clients and edge awareness

When deploying applications at the edge of mobile networks, initially application clients (running on devices) may enjoy being served at the edge without the need to change the application logic. Typically, application clients rely on the Domain Name System (DNS) to resolve a server's IP address. If DNS queries are augmented, through adding a DNS resolver at the core network, with location indication of the devices, the resolved IP address would allow serving the application clients from an edge cloud that runs a server instance and satisfies requirements for low latency. This mode of operation is often referred to as **edge-unaware**, meaning there is no need to change the application logic to enjoy edge cloud capabilities.



While using the DNS option allows several application requirements to be addressed, this mode of operation may have some limitations in cases requiring support of high user mobility in conjunction with the need to frequently transfer application context between servers. In this case, there is typically a need to inform, through notifications, the application about a change of the server IP address. This calls for a change in the application logic, therefore requiring applications to be **edge-aware**.

Server application edge awareness and application enablement

Edge awareness could also apply for server applications. The value propositions are three-fold:

- Access to network services such as location or traffic influence: by being deployed in conjunction with mobile networks, applications running at the edge could yield greater value if network services such as location, QoS or traffic influence can be used. By doing so, applications can offer an enhanced user experience.
- Application enablement at the edge: beyond accessing network services, edge applications may need to register and discover each other's services or access to a set of common services such as DNS or timing.
- Assist context migration in case of mobility: deploying applications at the edge allows latency to be reduced and saves network resources. However, to realise these benefits, highly mobile devices need to be connected to the most suitable edge cloud. For stateful applications, this results in a requirement for the edge cloud platform to assist in context migration to the target edge cloud. It must do this in coordination with the initial edge cloud, the target edge cloud, the network, the applications and the devices.

Discovery

An application client (i.e. running on a device) needs to discover the server application on the edge cloud. Broadly speaking there are two options for the discovery:

- DNS based: the most straightforward and friendly to application developers. With slight improvement at the mobile network, DNS queries may be enriched by the core network with the device location information. An authoritative DNS can, through IP address resolution, perform optimal routing to an edge cloud. However, the DNS options have limitations for applications running on devices that are highly mobile or when the edge cloud is highly distributed. The enhancements to support mobility when DNS is used are still under study in 3GPP.
- Device based: a device hosted client may also be used to facilitate the discovery of server applications and the optimal edge cloud. Such a client is termed an Edge Enabler Client (EEC) in 3GPP SA6 nomenclature. The same client can also assist service continuity because it can subscribe and receive information about mobility and possibly the decision to perform context migration during an application level session. Edge aware application clients on the device can directly interact with this client to benefit from all of these advantages.

Mobility and context transfer

While context migration can be performed over the top (through communication between edge cloud applications), it can profit from leveraging the network that provides the underlying mobility support. Examples include network exposure of mobility events, the network offered ability to influence traffic steering, etc. The whole subject of context migration is complex and may need a separate publication on its own – stay tuned!



3. A possible reference architecture leveraging existing/ongoing standards

3.1 Introduction

There are several ongoing standardization activities supporting the deployment of edge clouds in conjunction with mobile networks, enabling different business models to exist. These include those where MNOs operate the edge cloud infrastructure and offer cloud services to application service providers, but also others where the cloud service provider has a business agreement or a partnership with the MNO. While existing standards activities aim to provide support for different business models, the various modes of operations may be at a different maturity level. The most relevant standards activities to support edge cloud deployment in conjunction with MNOs are taking place at 3GPP and ETSI ISG MEC (Multi-access Edge Computing). Other groups such as GSMA and 5GAA focus on setting requirements and implementation agreements leveraging those standards where applicable. Different standards activities cover different aspects and applicability statements and complement each other to a large extent. This section suggests an architecture that supports different market-driven use cases and related requirements. The architecture leverages the existing standards and is referred to as a **Synergized Mobile Edge Cloud Architecture**. This synergized architecture provides common practices to developers such that they can create a single application software module running on common edge environments.

The standards bodies/groups involved in edge computing combined with MNOs are:

- **ETSI ISG MEC:** ETSI ISG MEC created an open and standardized IT service environment which allows third-party (both **edge-unaware** and **edge-aware**) applications to be hosted at the edge⁴ of the mobile network and which is capable of exposing network and context information. It specified a common and extensible application enablement framework for delivering services, specific service-related APIs for information exposure and programmability, as well as management, orchestration and mobility related APIs. These APIs facilitate the running of applications at the correct location at the right time and ensure service continuity. ETSI ISG MEC is currently studying MEC federations to enable shared usage of MEC services and applications across MEC systems in support of a multi-operator/multi-network/multi-vendor environment. ETSI ISG MEC has published several white papers describing deployment options and solutions for industries⁵.
- **3GPP SA6:** Defines an architecture (EDGEAPP) for enabling Edge Applications, specifically through the specification of an enabling layer to facilitate communication between application clients and applications deployed at the edge. Whilst offering indirect support to edge unaware application clients, EDGEAPP offers additional benefits for edge aware applications through direct interaction with the device hosted Edge Enabler Client. The architecture also enables the CAPIF (Common API Framework) to be leveraged as a standardized means of providing and accessing APIs in the Edge Cloud.
- **3GPP SA2:** Defines the architecture for mobile core networks including 5G. In the context of edge computing, 3GPP SA2 defines how user traffic is routed to the appropriate application servers in the edge clouds. It also provides the means for applications to provision traffic steering rules.

⁴ In the second term of the ISG MEC, the initial scope of Mobile Edge Computing based on 3GPP access-related technologies was extended to support Multi-access Edge Computing, including also Wi-Fi and fixed-access technologies.

⁵ https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf



- **3GPP SA5:** is responsible for management and charging aspects of 3GPP networks. In the context of edge computing, 3GPP SA5 is in the process of specifying life-cycle management of application servers in the edge cloud and charging aspects for edge services.
- **GSMA:** : GSMA is working to specify requirements and end-to-end high-level architecture for a unified Operator Platform that can help operators to make their assets and capabilities consistently available to the developers and the enterprise segment across networks and national boundaries. In phase 1 the work focuses on federating multiple operators’ edge computing infrastructure to give application providers access to a global edge cloud to run innovative, distributed and low latency applications through a set of common APIs. GSMA works to ensure that the industry, including the different standard bodies and open source communities, will deliver on this vision.
- **5GAA:** Defines requirements and implementation recommendations for Cellular V2X applications using Multi-access Edge Computing as per ETSI ISG MEC specifications.

Figure 1 shows the scope of edge computing work areas covered by relevant SDOs including ETSI ISG MEC and 3GPP working groups. In the horizontal ‘layers’ of the diagram, both the 3GPP network and application enablement layers include functionality to support the exposure of capabilities within that layer to the layer above. The application layer is shown for illustration purposes only, since it is out of scope of this paper. Several open source initiatives can also enable applications (big data, messaging, etc.) and also complement SDO specifications by providing supporting open source code for APIs. The GSMA is defining a transversal view of all capabilities required for building a platform allowing further market uptake of edge computing commercial services.

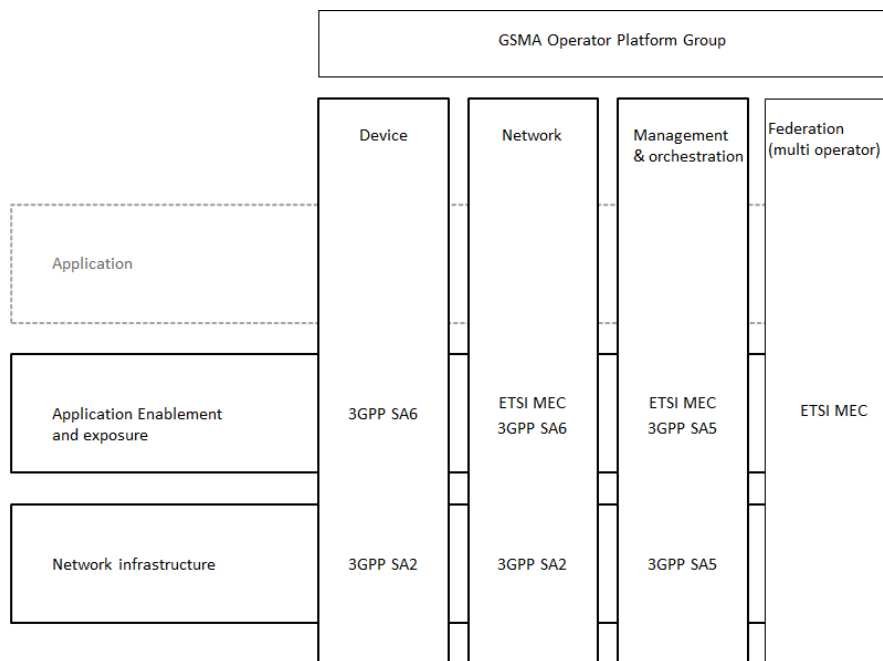


Figure 1: Overview of the involvement of standardization groups in the specification of edge computing in conjunction with mobile networks

GSMA is defining requirements about federation concepts, APIs, mechanisms and associated procedures between operator systems. ETSI ISG MEC is addressing those requirements in coordination with GSMA.



The northbound capabilities (e.g. for application deployment) are still being defined (GSMA, 3GPP, ETSI) and will support ongoing commercial platforms and services.

3.2 Synergized Mobile Edge Cloud architecture

Figure 2 provides a Synergized Mobile Edge Cloud architecture supporting different modes of operations and leveraging 3GPP and ETSI ISG MEC. This architecture is aligned with Annex C of 3GPP TS 23.558.

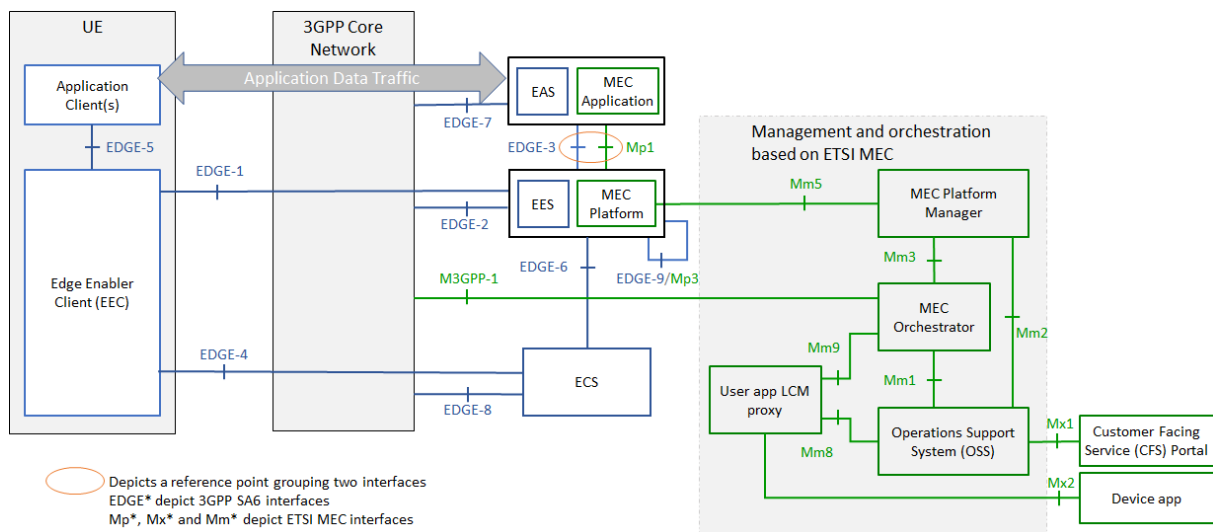


Figure 2: Synergized Mobile Edge Cloud architecture supported by 3GPP and ETSI ISG MEC specifications

Figure 2 depicts:

- **On the left**, devices run application clients which either use the DNS to discover application servers (Edge Application Server (EAS) in 3GPP SA6 terminology or MEC Application in ETSI ISG MEC terminology) or use the Edge Enabler Client (EEC) to perform the discovery according to the SA6 EDGEAPP architecture.
- **Towards the middle**, a platform (Edge Enabler Server (EES) in 3GPP SA6 and MEC Platform in ETSI ISG MEC) provide functionality pertaining to mediating access to network services, application authorisation, application's service registration and application's service discovery, context transfer, etc. A given implementation can combine functions specified by ETSI ISG MEC and ones specified by 3GPP SA6. The platform typically exposes APIs towards edge cloud applications (MEC application or Edge Application Server). EDGE-3 and Mp1 offer complementary API functions, therefore can be considered to be part of a single reference point from an application developer perspective.
- **Towards the right**, functionalities specified by ETSI ISG MEC include management and orchestration of the MEC platforms and OSS functions supporting access to portals offered to application service providers.



Below are additional considerations about this architecture:

- EDGE-3 and Mp1 provide service registration and service discovery features which allow an edge cloud application to register services exposed by this application and their subsequent discovery and use by other applications. The exposed services can be about network services, subject to their availability at the core or access network level. The common capabilities may be harmonised through adoption of the Common API Framework (CAPIF) as specified in 3GPP TS 23.222.
- EDGE-9 and Mp3 are both at early stage of development. Both are intended to assist in context migration.
- The following interfaces are about simple endorsement of SA2 interfaces (e.g. Network Exposure Function/Service Capability Exposure Function, NEF/SCEF): EDGE-2, EDGE-7, EDGE-8, M3GPP-1.
- According to 3GPP SA6 specification, edge services are exposed to the application clients by the Edge Configuration Server (ECS) and Edge Enabler Server (EES) via the Edge Enabler Client (EEC) in the UE. Each EEC is configured with the address of the ECS, which is provided by either the MNO or by the Edge Computing Service Provider.

Deployment options may implement all or a subset of the features of the synergized architecture as shown in subsequent sections.

3.3 Examples of deployments

This section provides different non-exclusive deployment options instantiating the synergized architecture presented in the previous section, which is a deployment option itself.

3.3.1 Deployment option 1: DNS based discovery

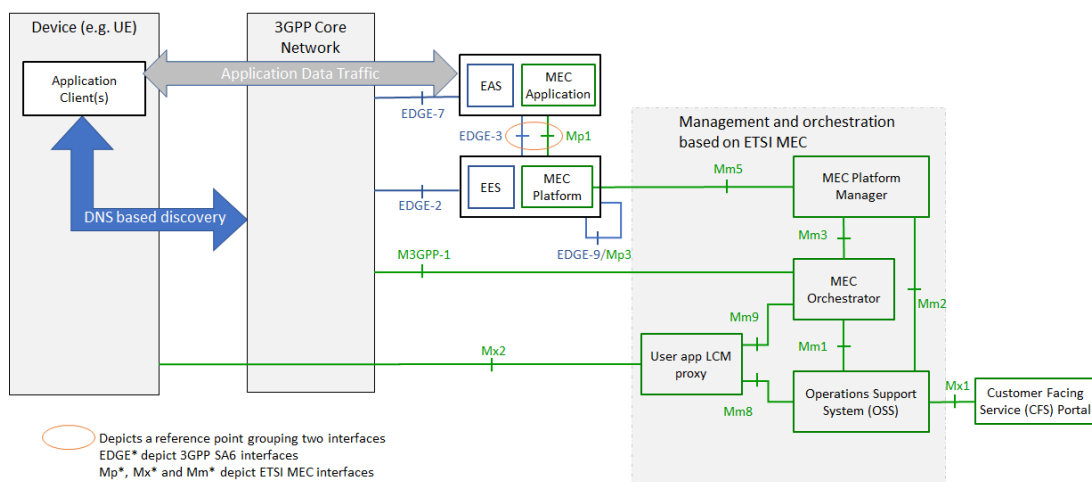


Figure 3: DNS based resolution in conjunction with harmonised 3GPP SA6/ETSI ISG MEC

In Figure 3, the DNS is used to discover the IP address of the edge cloud application. A possible approach is to extend the request with location to allow optimal resolution.



3.3.2 Deployment option 2: Edge Enabler Client discovery

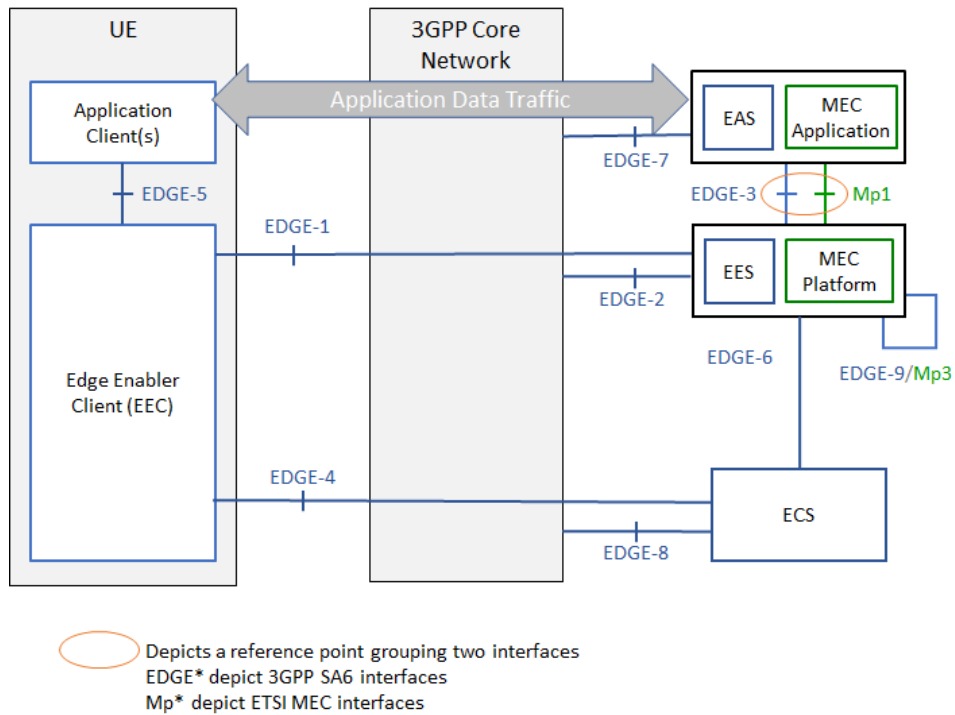


Figure 4: EEC based discovery

Figure 4, provides an instantiation where discovery is performed using the EEC with vendor-specific management and orchestration.

3.3.3 Deployment option 3: Edge Enabler Client discovery – Management and orchestration based on ETSI ISG MEC

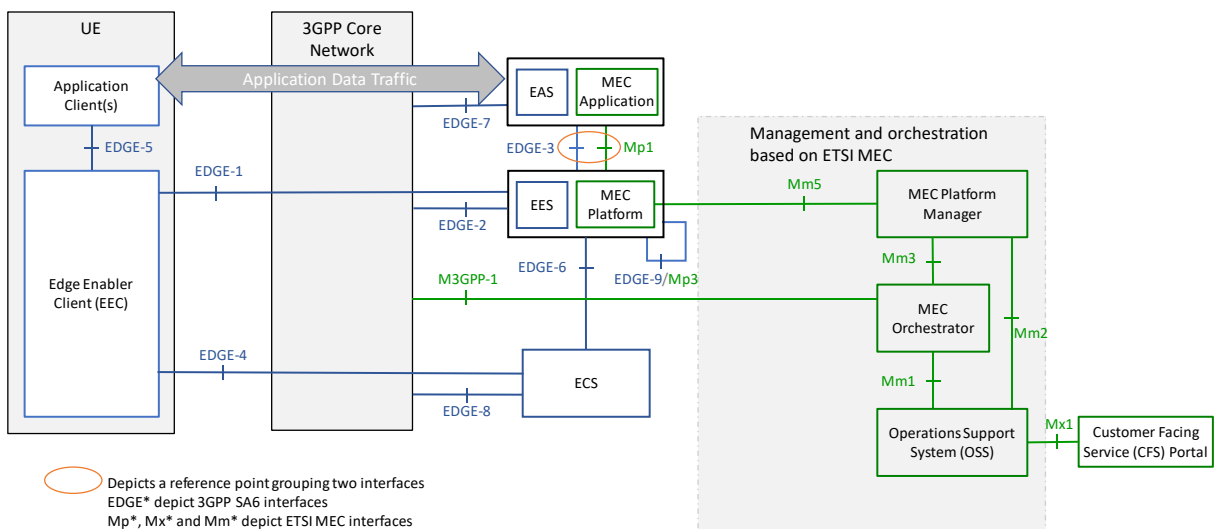


Figure 5: EEC based discovery – ETSI MEC management and orchestration



Figure 5, provides an instantiation where discovery of the Edge Application Server/MEC application is performed using the EEC as defined in the 3GPP SA6 specifications, and the orchestration and lifecycle management of the MEC application/Edge Application Server uses the standardized architecture specified by ETSI MEC. .

4. Conclusions and recommendations

ETSI ISG MEC and 3GPP have both worked on their own architectures for edge computing within the boundaries of their different scopes. Their common purpose is to create an open and standardized IT service environment for hosting and supporting third-party applications in edge environments.

3GPP SA6 is specifying the 3GPP UE and network's relationship with the EDGEAPP system. 3GPP SA2 is contributing to edge computing for 5G, with a main focus on DNS optimization for optimal discovery of/routing to edge clouds.

At the heart of the ETSI ISG MEC and 3GPP SA6 architectures lie the MEC Platform/Edge Enabler Server and the MEC Applications/Edge Application Servers respectively. There is a great degree of synergy apparent in the two architectures on these aspects, and in the information carried between these functional entities.

Further developments are still required in standards to cover business scenarios as defined in the GSMA for federating MEC capabilities among multiple operators and harmonising the access of developers/customers to edge computing capabilities.

ETSI ISG MEC is working to support those requirements for federation by enabling interactions among MEC systems, supporting multi-operator, -network, - vendor environments.

Involvement of open source communities has the potential to provide additional developments supporting ongoing standards, alignment and technology availability.

This paper highlights the value proposition of different standards streams and how those standards may be combined when it comes to deployments.

The authors of this paper believe, based on their experience in multiple organisations related to edge computing, that it is in the industry's best interests to define how a standardized end-to-end architecture can be created from a combination of complimentary parts.

The edge cloud ecosystem is still being developed; we are witnessing a rapid expansion of edge-related initiatives. To accelerate time-to-market and promote industry adoption, it is essential to ensure alignment, leverage synergies and offer common practices and tools for the developers. This will also help to simplify innovation co-creation with agile, plug-and-play value integration.



The Standards People

ETSI
06921 Sophia Antipolis CEDEX, France
Tel +33 4 92 94 42 00
info@etsi.org
www.etsi.org

This White Paper is issued for information only. It does not constitute an official or agreed position of ETSI, nor of its Members. The views expressed are entirely those of the author(s).

ETSI declines all responsibility for any errors and any loss or damage resulting from use of the contents of this White Paper.

ETSI also declines responsibility for any infringement of any third party's Intellectual Property Rights (IPR), but will be pleased to acknowledge any IPR and correct any infringement of which it is advised.

Copyright Notification

Copying or reproduction in whole is permitted if the copy is complete and unchanged (including this copyright statement).

© ETSI 2020. All rights reserved.

DECT™, PLUGTESTS™, UMTS™, TIPHON™, IMS™, INTEROPOLIS™, FORAPOLIS™, and the TIPHON and ETSI logos are Trade Marks of ETSI registered for the benefit of its Members.

3GPP™ and LTE™ are Trade Marks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.

GSM™, the Global System for Mobile communication, is a registered Trade Mark of the GSM Association.