ATHONET

GSMA GLOMO AWARDS
THE GLOBAL MOBILE AWARDS
WINNER

2016

SGW-LBO solution for MEC
Taking services to the Edge

Executive Summary

A new wave of demanding services requires not just high bandwidth but lower latency and greater perceived user quality. Many of these services cannot afford to further overload often costly backhaul networks. This requires extending the mobile network architecture, and placing functionality at the edge (Multi-Access Edge Computing or "MEC") rather than increasing variability by hauling traffic back to the middle mile or centralised core.

Operators are seeing that a relevant part of their traffic, in some cases as high as 60-70%, is from common sources such as YouTube, Facebook and Netflix. Athonet's MEC solution allows this content to be cached and served locally. This improves the user perceived quality and reduces the amount of backhaul required in a network. It also enables new business opportunities with content providers and through other low latency services.

The SGW-LBO is a modified SGW function which has been enhanced by Athonet to allow traffic to be "broken-out" and steered locally, whilst maintaining 3GPP compatibility, to support i) caching of Facebook, Youtube and other content and ii) other applications that require low latency or local offload (e.g. smartcity or autonomous cars). The SGW-LBO is transparent to and does not require modifications in the Radio or Core Network. It can also be software upgraded to create a future-proof bridge to 5G.

This fully softwarised solution, deployed at the edge cloud on Commercial Off-The Shelf (COTS) servers and virtualized infrastructure, is easy to deploy and manage. It significantly reduces costs, and enables new services and revenue models, all of which are critical to the economics of edge deployments.

Unlike Athonet's SGW-LBO, there are alternative approaches to MEC that are functionally limited, increase cost and complexity or affect the existing mobile core network. One such approach requires the full PGW and SGW to be pushed out to the network edge. This has cost implications but it also means that an APN has to be fully, rather than selectively, off-loaded at the edge. Such a characteristic makes this approach generally unsuitable to a mobile operator network where the objective is to define finer granularity policies for offloading traffic (e.g. video). Another well-known alternative is "bump in the wire" which involves intercepting the interface between the eNB and the Mobile Core Network (S1 interface). This approach has drawbacks as it breaks the end-to-end 3GPP architecture affecting network security and lawful Intercept procedures. It is also incompatible with mobile paging which means that offloaded applications are unable to implement "push" services i.e. the application server cannot initiate data communications. This is a critical disadvantage for a number of applications where "push" is mandatory including IoT, messaging, voice etc.

It is clear that in order for MEC platforms to be widely adopted by Mobile Network Operators the MEC approach used needs to have i) minimal impact on existing 4G architecture and network processes ii) use standard 3GPP interfaces to the largest possible extent iii) provide a seamless software upgrade to 5G. Taken altogether, we believe that the SGW-LBO is the most deployable MEC solution for operators that are looking to deliver new demanding and high value services while saving backhaul costs, improving service quality and creating the foundation for the 5G edge-cloud.

Acronyms

5G NR	5G New Radio
API	Application Programming Interface
BIW	Bump in the Wire
CDR	Charging Data Record
CGW	Charging Gateway
CUPS	Control and User Plane Separation
eNB	evolved Node B
EPC	Evolved Packet Core
FTP	File Transfer Protocol
HSS	Home Subscriber Service
IMS	IP Multimedia Subsystem
ISP	Internet Service Provider
LBO	Local Break Out
LI	Lawful Intercept
LTE	Long Term Evolution
MEC	Multi-access Edge Computing
MME	Mobility Management Entity
MNO	Mobile Network Operator
NSA	Non Standalone
PGW	Packet Data Network Gateway
QoS	Quality of Service
RAN	Radio Access Network
SAE-GW	System Architecture Evolution Gateway. The SAE-GW includes both SGW and PGW
SGW	Serving Gateway
SGW-LBO	SGW Local Break Out

Table of Contents

1 What is MEC?

MEC refers to the concept of bringing application and computing capabilities to the edge of the network, where it is closer to the device consuming such resources. To understand the interest in this area of work, one needs to look at typical mobile network deployments today. Figure 1 shows a simplified network deployment.

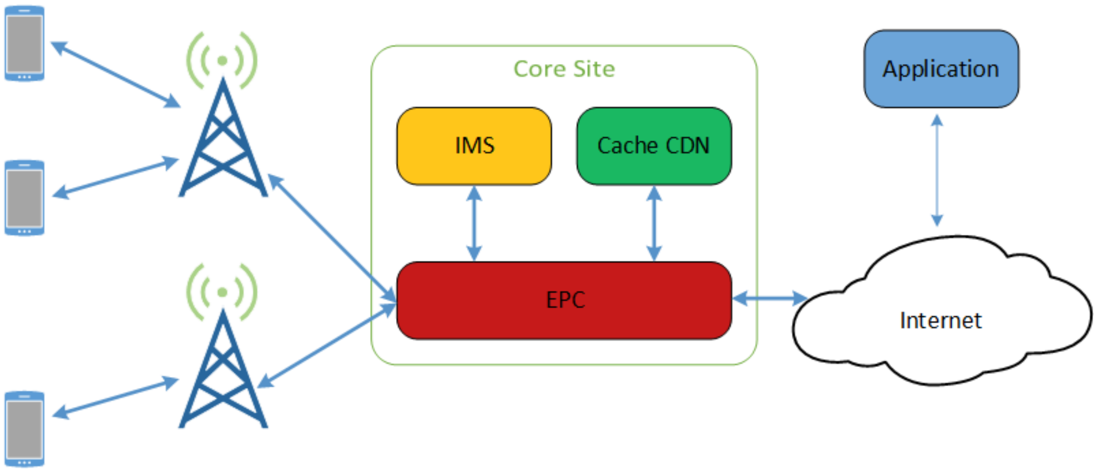

Figure 1. A typical mobile operator's deployment.

For devices communicating to application servers on the Internet, traffic needs to traverse the mobile network core, pass through transport networks and reach the application at the other end. The same happens in the reverse direction. This has been the case for decades; however, as mobile networks become more complex and their use cases increase to include ones that were rarely considered in the past, new demanding requirements arise.

A number of studies and business cases have shown that the existing model is inefficient and introduces non-deterministic or unacceptable delays for certain type of services. The following section presents an overview of use cases motivating the need for MEC.

2 The Case For MEC

The new wave of immersive video services (360 degree video) with AR/VR and the demanding needs of V2X for connected and autonomous cars, connected health and other vertical use cases require not just high bandwidth but low latency and highly reliable communications. Functionalities are required to be deployed at the edge rather than increasing variability by each additional point of failure introduced by hauling traffic back to the middle mile or centralised core. By highly distributing core network components (pure software VNFs) across the network, the operator and business partners are able to offer consistent QoS across thousands of network access points compared to what is available today.

MEC is driven primarily by three requirements: the economic need to reduce transit traffic between the edge and the centre, the delay-sensitivity of some services and the need to keep traffic local for security and regulation. We discuss these requirements below.

2.1 Reducing the cost of transit traffic

Operators like AT&T say that in 2016, video already accounted for more than 50% of mobile data traffic[1]. Other carriers globally are indicating that they are experiencing levels as high as 75%, most of it from popular video content providers (e.g. YouTube, Netflix, Facebook and Vimeo).

Given the consistency of content sources, it is natural that network operators would look for ways to reduce transit traffic on backhaul links and utilisation of the centralised resources by adding local caches of video providers. The closer the cache to the end device, the better. Non-mobile ISPs (Internet Service Providers) are already capable of bringing the content into a Point of Presence (PoP). MEC seeks to allow mobile operators to adopt a similar approach.

In other MEC cases, such as autonomous and cloud-assisted vehicles, where substantial data and traffic is generated at the edge, only traffic that is essential to be transported to the centralised cloud needs to be transited on backhaul links whilst the remainder can be off-loaded at the network edge.

Investing in economic edge computing facilities can significantly reduce opex and capex demands on more expensive centralised resources and backhaul networks whilst yet offering a superior service to end-users.

2.2 New Delay-sensitive Use Cases

V2X communications for autonomous and cloud-assisted vehicles and AR/VR demand reliable and low latency connectivity. These types of services are often referred to as Ultra Reliable Low Latency Communications (URLLC) services, where delay sensitivity is much higher compared to the traditionally delay sensitive voice. The ability to bypass the core network and deliver such services from a local compute facility will eliminate a significant delay component, where they compete with other traffic in the core and provide a more deterministic delay budget.

2.3 Regulation and security

Some industries such as healthcare and financial services have stringent requirements for regulatory compliance which require data to be stored, transferred and processed on site. The ability to offload, steer and process such traffic locally opens up new opportunities not currently supported by the mobile network.

[1] https://www.bloomberg.com/news/articles/2017-02-27/at-t-boosts-5g-networks-rollout-as-online-video-demand-spikes

3 Current MEC solutions in the industry

Two different approaches to MEC have emerged over the years. One approach distributes the entire core or at least the SAE-GW (SGW + PGW) at the network edge ("SAE approach") and allows traffic to be offloaded, e.g., based on the APN configured in the PGW. This approach has been supported by Athonet to pioneer the LTE private networks that have been used all over the world with some notable "world firsts", such as the first use of LTE for the Smart-grids in 2011, the world first use of LTE for disaster recovery in 2012 following the Italian earthquake. The "private network" approach is very useful in the context of an enterprise that needs to create a business-critical dedicated network. However, this approach is limited by the fact that the entire APN traffic is locally offloaded. In other use cases the operator may need to have a finer granularity over the type of traffic that should be offloaded. Figure 2 illustrates such approach.

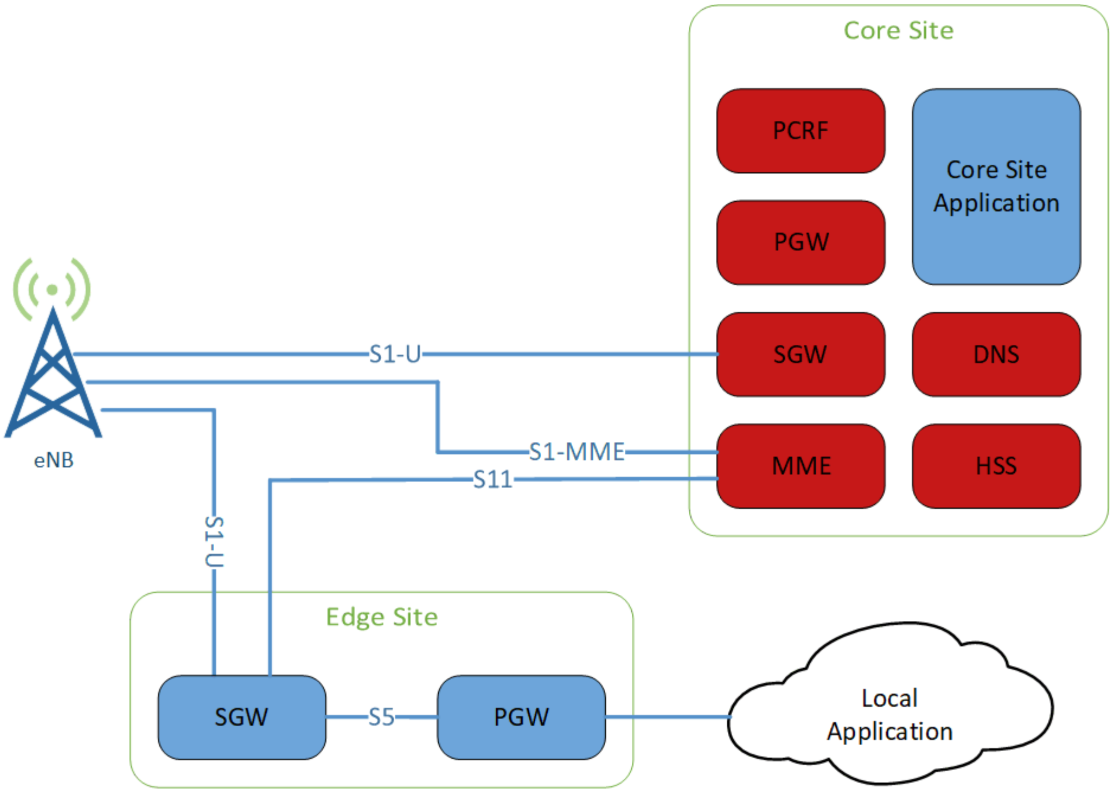

Figure 2. Distributed Core as a MEC solution

A second approach to MEC is "Bump in the Wire" (BIW) or "Bump in the Stack", which introduces a new function that intercepts signalling and data traffic on the S1 interface and steers it to the local MEC applications. Figure 3 illustrates a simplified BIW approach.

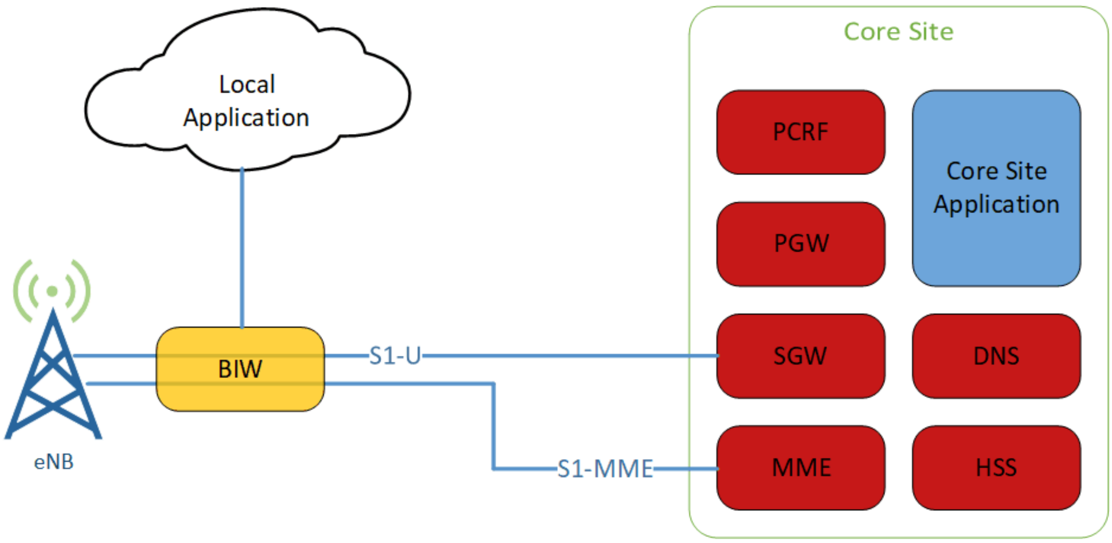

Figure 3. Overview of the “Bump in the Wire” approach

As seen in figure 3, the BIW function intercepts both signalling and user traffic and based on configured policies, decides to steer some traffic out towards the application outside the core network. This approach has several limitations as discussed below.

4 Bump in the wire limitations

The “bump in the wire” approach to MEC has several limitations which will hamper its ability to reach widespread adoption. The limitations of this approach are listed below.

<u>IPsec and security</u>: IPsec can be used to protect the S1 interface between the eNBs and the core network, thus increasing security by making it impossible to intercept messages. However, the BIW solution needs to inspect S1 messages. Therefore, the adoption of the BIW solution forces an operator to either disable IPsec or limit the BIW entity’s location to somewhere higher up in the network (e.g. behind the IPsec gateway) where the messages can be viewed in clear text. If the latter option is chosen, it limits an operator’s placement of the MEC platform and reduces their ability to distribute it. Such reduction in distribution limits a key benefit of a MEC platform, which is to be as close as possible to end devices. The alternative is to allow the MEC platform to “break” the IPsec tunnel which breaks the end-to-end 3GPP security architecture and therefore introduces new security risks.

<u>Idle user reachability</u>: A MEC application relying on bump in wire, at best, will add significant delays to the connection initiation with an idle device. At worst, the application will be unable to initiate a connection towards a device in power-saving (IDLE) mode. This is because a BIW solution does not support the paging procedure required to wake up a device in IDLE mode. It may achieve that, where possible, only by relying on a PGW node that will be at a different location thus introducing latency. This is an important limitation for an application that needs to be responsive and close to the user. An application would have no knowledge about the device’s status. That

is, whether the device is unreachable since it has gone out of the MEC domain or whether it has simply gone into IDLE mode.

Lawful Intercept of a selected user using BIW is possible only by adding complexities (e.g. new non-standard 3GPP network functions and interfaces) into the operators' network. The lack of standardized approach may pose problems with national authorities

Traffic charging: With BIW, it is difficult to produce Charging Data Records (CDR) for steered traffic. This is because the MEC platform does not own all the information such as IMSI, IMEI, IP address, APN, cell level user location, among others, which are necessary for producing CDRs. Charging can only be done by adding complexities (e.g. new non-standard 3GPP network functions and interfaces) into the operators' network.

Some of the above issues could be solved by adding new nodes to the operator's core network, adding costs, complexity and footprint to the solution which would inevitably diminish the attractiveness of the economics of edge deployments. Also, such new nodes would require proprietary interfaces resulting in vendor lock-in, thus limiting the ability to offer cost effective and efficient solutions.

6 Standardization Efforts

Standardization efforts aim to solve the following:

- Adjust data routing within the core and deliver traffic locally in some cases;
- Enable policy-based decision making for certain traffic patterns.

As part of the above two requirements, there is a need for applications to query the network about certain information pertaining to one or more users, or information about aggregate users within a particular location.

The above problems are essentially broken into two domains:

- Policy and configuration management requests for the MEC platform. This is executed by the MEC application or a policy engine; network configuration, traffic steering, QoS enforcement are based on the MEC application request.
- The policy engine and/or the application interfaces with the MEC Platform. This requires the definition of standard communication mechanisms between the application and the MEC platform. This problem is being addressed by the ETSI MEC group, which turns to APIs to define such communication.

The other requirement involves the need of the platform to configure the network based on application demands and to steer traffic accordingly.

The MEC standardization effort by ETSI MEC in phase 1 has focused on service APIs, application enablement and management interfaces mainly for application on-boarding and life cycle management. The ETSI MEC in phase 2 may address other problems, which includes topics such as:

- Further API definition of breakout criteria;
- Handover between MEC and national network;
- Application mobility between MEC platforms;
- Idle user reachability from MEC application;
- Security E2E and 3GPP;
- Lawful Intercept;
- Charging.

The solution presented in this paper addresses the need for a generic MEC platform which is designed with careful consideration of 3GPP standards and principles, which is compatible with current mobile network deployments. The platform needs to be regarded as an add-on to the operators' network, for which the integration causes little or no impact to the existing MNO's infrastructure.

7 A New approach to MEC: the SGW Local Break Out (SGW-LBO)

Whilst the mobile edge cloud has often been talked about, a clean solution to enable it in the mobile network is lacking. As seen above, solutions such as BIW are hampered by security concerns, charging, lawful interception limitations and lack of support for "push" applications. On the other hand, steering the entire APN traffic locally (with the SAE approach) may not be appropriate for many deployments.

In order to allow an operator to steer traffic flexibly based on either users' identifiers or uplink classifiers that may contain complex filters, we are proposing to position the SGW into each MEC platform.

This allows an easy introduction of the MEC platform into the operator network following these steps:

- Ensure S11, S5 and Bx (optional) network reachability on the Core Network side by the MEC platform;
- Ensure the S1-U network reachability on the RAN side by the MEC platform;
- Update the operator's DNS in order for the MME to select the MEC platform for the Tracking Area where the eNBs that need to be served are located.

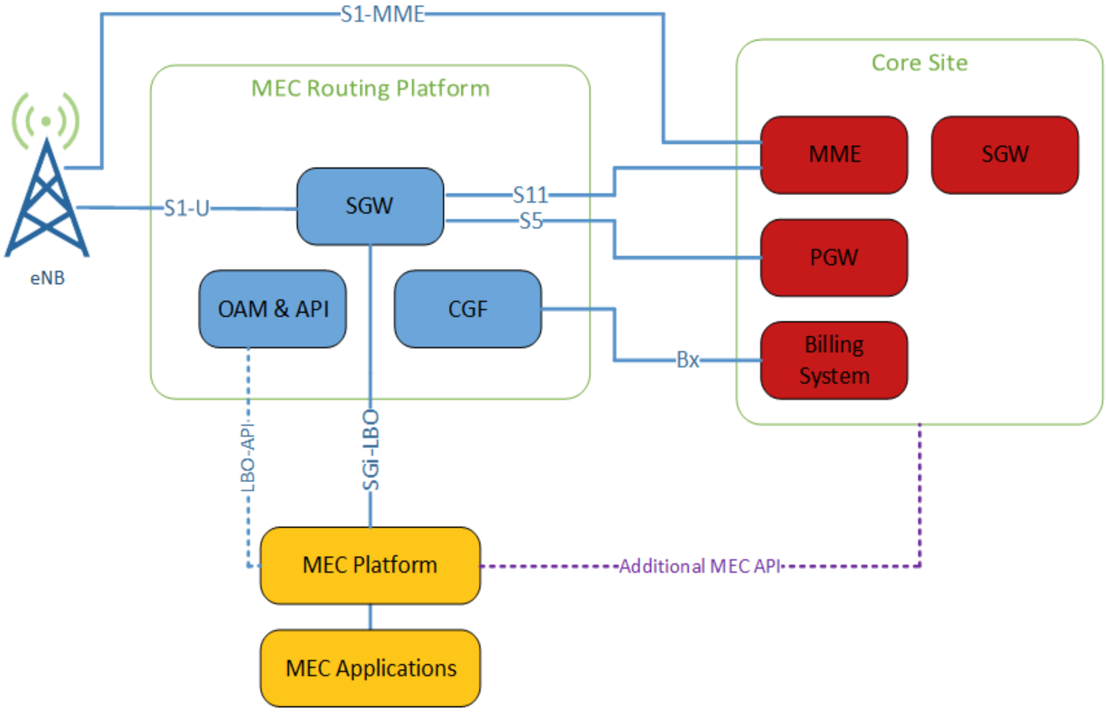

Fig. 4 MEC solution architecture using the SGW-LBO approach

The MEC application connects to the MEC platform through MEC API's. The MEC platform gathers data from various components in the network and uses them to respond to the MEC application's requests. The SGW-LBO routes the traffic and is the data plane of the MEC solution, which enables local breakout based on per-user or per-traffic stream policies provisioned via APIs.

The figure above illustrates the default 3GPP interfaces that should be supported by the MEC platform that uses the SGW with a special Local Break Out (LBO) functionality which allows to selectively steer the data traffic to a local application.

The SGW-LBO connects externally via the following interfaces:

- S1-U: GTPv1-U based interface used to connect the SGW to the eNBs;
- S5: GTPv2-C and GTPv1-U based interface used to connect the SGW to the PGW in the core site;
- S11: GTPv2-C interface used to connect the SGW to the MME in the core site;
- SGi-LBO: interface used to receive and transmit data to/from an external network including local private LAN (Intranet), Internet, or a services network;
- Bx: interface used for fetching the CDRs. This interface allows billing systems to get the CDRs for offline charging.

Other interfaces (not depicted) include:

- X1, X2, and X3, or H1, H2 and H3 interfaces for LI purposes;

- Configuration management to provision LBO rules based on parameters such as IMSI, APN and 5 tuples.

8 The SGW-LBO and the migration to 5G architecture

The SGW-LBO is consistent with 3GPP standards and maps into 5G functionality (CUPS, UPF, networks slicing). This approach allows mobile operators to drive down the risk and thus the cost of future 5G deployments at a time when 5G business models are uncertain. In fact our solution allows CSPs to serve new vertical markets, such as autonomous and cloud-assisted vehicles, before deploying 5G. It can be seen as an ideal instrument that allows operators to build the edge cloud and experiment with 5G-like use-cases ahead of full 5G deployments.

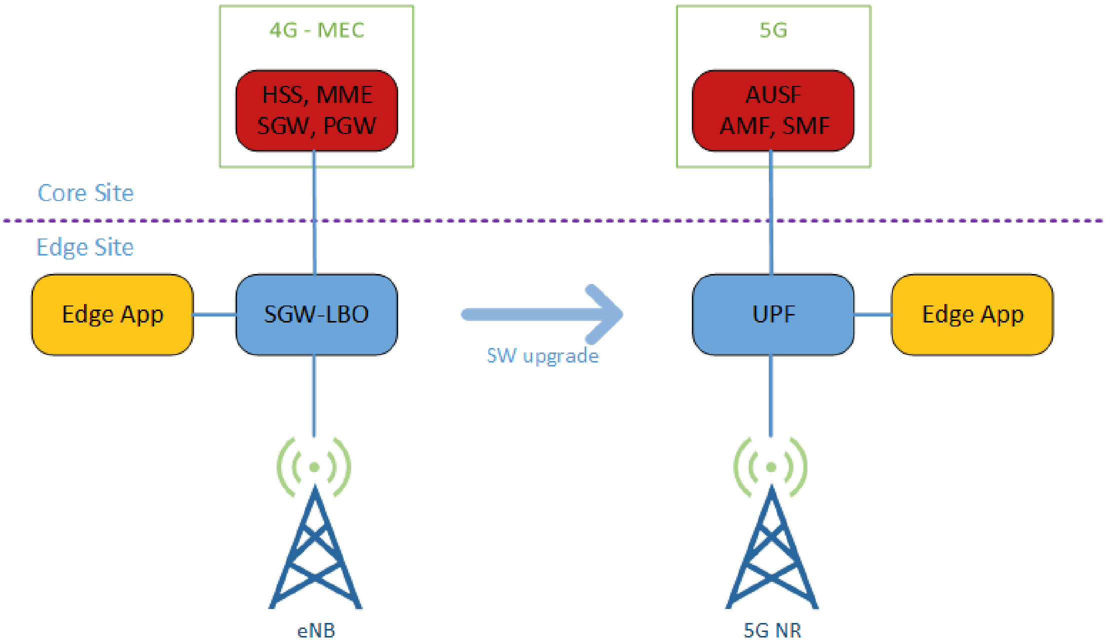

Fig. 5 MEC adoption and evolution to 5G

9 Integration with a mobile network operator

The SGW-LBO is distributed at the edge of the network and connects to the mobile network operator’s MME and PGW via the 3GPP standard interfaces S11 and S5, respectively. The SGW-LBO is a standard compliant 3GPP SGW node, part of the MEC platform that is controlled and coordinated by the operator from the central core. The SGW-LBO selection by the MME is performed using the standard 3GPP procedures which involve querying the DNS and selecting the SGW and PGW pairs based on user location (Tracking Area) and the requested APN.

In case the SGW-LBO is switched off, or the user goes out of MEC coverage due to mobility or the entire MEC platform becomes unavailable, another SGW, such as the one in the central core site, will be selected by the MME according to the 3GPP indications and the DNS priority ranking returned to the MME.

10 Benefits for network operators

The SGW-LBO uniquely solves the requirements imposed by new applications using existing LTE networks by:

- Extending traffic plane network functions of the SGW, coupled with highly selective traffic steering functionalities, to the network edge;
- Positioning content/application servers (e.g. CDN) close to end customers in radio sites and/or fibre cabinets for small cells;
- Enabling push applications (paging);
- Requiring low CPU power (2 vCPU) making for very light-touch deployments;
- Applying strict security, encrypting all transport interfaces (to/from eNBs and mobile core);
- Providing transparent support of standard functions such as mobility, paging, lawful intercept, charging;
- Integrating easily with existing mobile networks;
- Reducing costs potentially down to a fraction of the total cost of ownership of competing approaches;
- Simplifying deployment, management and reducing the hardware/software footprint thus making wide-spread deployment economically and technically feasible.

This solution represents an ideal bridge to offer 5G type of performance and use-cases with LTE and NSA 5G-NR radio, while building the network edge as a key stepping stone for 5G evolution.

11 About Athonet

Athonet provides a complete software-based mobile packet core solution (EPC) which also includes Policy and Charging Rules Function (PCRF), Home Subscriber Server (HSS), Voice-over-LTE (IMS for VoLTE), LTE Broadcast (eMBMS) and the CSGN for NB-IOT support. This efficient mobile core solution can be deployed in fully virtualised environments (NFV), enterprise data centres or on standard off-the-shelf servers (bare metal). The software solution is designed for centralised or highly distributed deployments in Tier 1 Mobile Operators, Challenger Telcos, Governments and Public Safety or mission critical applications.

Over the years, Athonet has deployed some of the world's first private LTE networks in order to satisfy the need of first responders, public safety, smartgrids, utilities and enterprises which require guaranteed service even in case of backhaul failure and low latency application support. Those deployments are now regarded as early MEC deployments of private broadband cellular networks which also host the applications (e.g. monitoring, voice, video and Push-To-Talk) at the network edge.

More info on Athonet
Web: www.athonet.com
Mail: info@athonet.com
Twitter: https://twitter.com/athonet_primo
LinkedIn: https://www.linkedin.com/company/athonet/