Mobile Networks and Applications

Introducing mobile edge computing capabilities through distributed 5G Cloud Enabled Small Cells

--Manuscript Draft--

Manuscript Number:	
Article Type:	Special Issue on Mobile Networks and Management, MONAMI 2015
Keywords:	centralized mobile networks; Light data centre; small cells; mobile edge computing; 5G
Corresponding Author:	Jose Oscar Fajardo University of the Basque Country Bilbao, Bizkaia SPAIN
First Author:	Jose Oscar Fajardo
Order of Authors:	Jose Oscar Fajardo
	Fidel Liberal
	Ioannis Giannoulakis
	Emmanouil Kafetzakis
	Vincenzo Pii
	Irena Trajkovska
	Thomas Michael Bohnert
	Leonardo Goratti
	Roberto Riggio
	Javier Garcia-Lloreda
	Michele Paolino
	Pavel Bliznakov
	Jordi Perez-Romero
	Claudio Meani
	Ioannis Chochliouros
	Maria Belesioti

Introducing mobile edge computing capabilities through distributed 5G Cloud Enabled Small Cells

Jose Oscar Fajardo
University of the Basque Country (UPV/EHU), ETSI Bilbao, Almda Urquijo s/n, 48013 Bilbao, Spain.
+34-946017202
+34-946014259
joseoscar.fajardo@ehu.es
http://det.bi.ehu.es/nqas/joseoscar.fajardo
Fidel Liberal
University of the Basque Country (UPV/EHU), ETSI Bilbao, Almda Urquijo s/n, 48013 Bilbao, Spain.
fidel.liberal@ehu.eus
Ioannis Giannoulakis
N.C.S.R. "Demokritos", Patr. Gregoriou & Neapoleos 15310, Agia Paraskevi, Greece.
giannoul@iit.demokritos.gr
Emmanouil Kafetzakis
ORION INNOVATIONS, Ioulianou 56 10439, Athens, Greece
mkafetz@orioninnovations.gr
Vincenzo Pii
Zurich University of Applied Sciences (ZHAW), Obere Kirchgasse 2, 8400 Winterthur, Switzerland
piiv@zhaw.ch
Irena Trajkovska
Zurich University of Applied Sciences (ZHAW), Obere Kirchgasse 2, 8400 Winterthur, Switzerland
traj@zhaw.ch
Thomas Michael Bohnert
Zurich University of Applied Sciences (ZHAW), Obere Kirchgasse 2, 8400 Winterthur, Switzerland
thomas.bohnert@zhaw.ch
Leonardo Goratti
CREATE-NET, Via alla Cascata 56D, 38123, Trento, Italy.
leonardo.goratti@create-net.org
Roberto Riggio
CREATE-NET, Via alla Cascata 56D, 38123, Trento, Italy.
roberto.riggio@create-net.org
Javier Garcia Lloreda
ATOS, Albarracín 25 28037 Madrid, Spain.
javier.garcial@atos.net
Michele Paolino
Virtual Open Systems, Grenoble, France
m.paolino@virtualopensystems.com
Pavel Bliznakov
Virtual Open Systems, Grenoble, France
p.bliznakov@virtualopensystems.com
Jordi Perez-Romero
Universitat Politecnica de Catalunya (UPC), Campus Nord, C/Jordi Girona 1-3, 08034 Barcelona, Spain
jorperez@tsc.upc.edu
Claudio Meani
ITALTEL, Via Reiss Romoli - Località Castelletto 20019 Settimo Milanese (Mi,)Italy
claudio.meani@italtel.com
Ioannis Chochliouros
Hellenic Telecommunications Organization S.A. (OTE), 1, Pelika & Spartis, Maroussi, Athens, Greece
ichochliouros@oteresearch.gr
Maria Belesioti
Hellenic Telecommunications Organization S.A. (OTE), 1, Pelika & Spartis, Maroussi, Athens, Greece
belesioti@oteresearch.gr

Abstract:

Current trends in broadband mobile networks are addressed towards the placement of different capabilities at the edge of the mobile network in a centralized way. On one hand, the split of the eNB between baseband processing units and remote radio headers makes it possible to process some of the protocols in centralized premises, likely with virtualized resources. On the other hand, mobile edge computing makes use of processing and storage capabilities close to the air interface in order to deploy optimized services with minimum delay. The confluence of both trends is a hot topic in the definition of future 5G networks. The full centralization of both technologies in cloud data centres imposes stringent requirements to the fronthaul connections in terms of throughput and latency. Therefore, all those cells with limited network access would not be able to offer these types of services. This paper proposes a solution for these cases, based on the placement of processing and storage capabilities close to the remote units, which is especially well suited for the deployment of clusters of small cells. The proposed cloud-enabled small cells include a highly efficient microserver with a limited set of virtualized resources offered to the cluster of small cells. As a result, a light data centre is created and commonly used for deploying centralized eNB and mobile edge computing functionalities. The paper covers the proposed architecture, with special focus on the integration of both aspects, and possible scenarios of application.

Keywords: centralized mobile networks, Light data centre, small cells, mobile edge computing, 5G

Acknowledgments:

The research leading to these results has been performed in the scope of the H2020 5G-PPP project SESAME. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 671596.

1. Introduction

In the framework of evolved 4G LTE architectures and future 5G mobile networks, different proposals are emerging aimed at overcoming the capacity limitations of current Radio Access Networks (RANs). The concept of Network Softwarization is being proposed to move some RAN-related functions back to shared hardware (HW) elements with high computational capacities, taking into account several trends such as Software Defined Radio (SDR), Software Defined Networking (SDN) and Network Function Virtualization (NFV). Complex future content centric Internet and the 5G paradigm suggest the introduction of intelligent network nodes that will enable more powerful adaptation and prioritization frameworks over the whole transmission chain, and especially at the edge of mobile network segments [1].

A Centralized RAN system concentrates different processing resources together to form a pool in a central data centre. This aggregation of HW resources in shared locations not only reduces deployment costs, but also leverages low latency connections between different RAN processing units enabling a series of enhanced capabilities such as advanced Radio Resource Management (RRM) coordination and interference management [2]. When these resources run over virtualized infrastructures, adding flexible and scalable HW resource management capabilities, the Centralized RAN becomes a Cloud RAN (C-RAN) [3]. According to 3GPP terminology, this high degree of centralization entails a network architecture where all the baseband processing is made by BaseBand Units (BBU) at centralized data centres, and radio signals are exchanged with the Remote Radio Heads (RRH) over high-speed low-latency connections that constitute the mobile fronthaul.

However, as explained in [4], the requirements of the fronthaul in terms of high data throughput and low delays makes this fully centralized solution nowadays feasible only when optical fibre connections are deployed. Since this technology is not available everywhere, and its complete deployment entails costly operations, other alternatives for partial centralization are also considered. The NGMN Alliance released a document concerning the fronthaul requirements associated to the centralization of the MAC entity [5]. In order to cope with the typical HARQ timing of 8 ms, only fronthaul delays of up to 250 μs (in terms of one-way latency) are acceptable. Higher latency fronthaul technologies require the application of HARQ interleaving, in order to delay the HARQ process. Similarly, the Small Cell Forum performed a comprehensive analysis including other possible functional splits between centralized and remote functions [6, 7]. Concerning the MAC-PHY split, the conclusion for the standard HARQ processing is the same than in the NGMN case: fronthaul delays of up to 250 μs are acceptable. Fig. 1 below illustrates different alternatives for the centralization of RAN functions.

Beyond the pure centralization of an evolved Node B (eNB) functions, the current trends towards “cloudification” of the RAN also allows reusing the available HW infrastructure for deploying service instances at the edge of the mobile network. One of the emerging technologies to cope with more personalized and user-centric service provisioning is the novel Mobile Edge Computing (MEC) industry initiative [8], a promising approach to solve these types of problems from an operator-supported perspective. This initiative proposes that mobile network operators would provide an API to third-party partners, offering them access to critical features such as location awareness and network context information. This information may be exploited to deploy proximity-enabled services with close-to-zero latency characteristics, in order to optimize the management of future mobile networks. Regardless of the adopted architecture for C-RAN, MEC-driven service instances must be deployed over the cloud resources available at the RAN side. Taking Fig. 1as a reference, these types of MEC services would be deployed as “APPS” features. Thus, the degree of coupling between service-level instances and other RRM functions may differ.

Besides the logical split of the RAN and MEC functions, the physical distribution of the HW resources must be also taken into consideration. Fig. 2 illustrates two different alternatives considering centralized and distributed HW resources. Focusing on the problem of deploying MEC services, only the upper layers of the protocol stack are considered for centralization in Fig. 2.

In the first case (Fig. 2-a), the upper RAN functions are located in powerful data centres that are ideally connected to the RRHs through high-speed and low-latency fronthauls. According to [6,7], fronthaul delays of up to 30 ms can be tolerated for a proper operation of the LTE stack. Yet, high fronthaul delays may degrade the performance of certain novel edge services that require close-to-zero latencies as prescribed by 5G objectives.

In these scenarios, the second alternative in Fig. 2-b may become better suited for deploying mobile edge services. In that case, some processing and storage resources are placed close to the RRH, and thus the fronthaul delay is significantly reduced. Deploying huge data centres implies a series of requirements in terms of space, energy, etc. Hence, this second option envisages the deployment of a series of HW resources with limited capacity and requirements and in a distributed configuration, which may ideally collaborate to provide some edge computing capabilities.

This second solution is especially relevant to enable flexible deployment of Small Cells, and particularly attractive for targeting currently deployed network architectures and special limited-access scenarios. In the former case, a Small Cell operator may think about endowing its deployed network with novel mobile edge capabilities by gradually upgrading the Customer-Premises Equipments (CPE) without requiring to change wired connections. In the latter case, some special locations may face specific problem for upgrading the network access, such as rural areas. In both scenarios, the introduction of distributed computing and storage capabilities associated to the CPEs arises as an affordable and scalable solution.

This paper describes the solution developed in the scope of the European H2020 5GPP project Small cEllS coordinAtion for Multi-tenancy and Edge services (SESAME), and provides some insights on the deployment of mobile edge services over the proposed architecture.

The remainder of the paper is organized as follows. Section 2 presents the main concept of the SESAME architecture and defines the main functional elements. Section 3 focuses on the provision of mobile edge services over the SESAME architecture. Section 4 presents a series of target scenarios where the proposed solution introduces significant benefits, and states some open issues. Finally, Section 5 provides the conclusions to the paper, highlighting the foreseen benefits of the proposed solution and identifying several open issues.

2. SESAME concept and architecture

The main novelties and innovations of the proposed solution are achieved by placing intelligence in the network’s edge through the employment of Network Functions Virtualisation (NFV) and edge cloud computing directly to the Small Cell part. The consolidation of multi-tenancy in communications infrastructures is also in scope, allowing several operators/service providers to engage in new sharing models of both access capacity and edge computing capabilities. To that end, SESAME proposes the use of cloud-enabled Small Cells as a new multi-operator enabled Small Cell that integrates a virtualised execution platform (i.e., a micro-server) for deploying Virtual Network Functions (VNFs) and also to execute novel applications and services inside the access network infrastructure.

SESAME will consolidate the concept of delivering Small Cell coverage, coupled with a virtualised execution platform, according to the ‘as-a-Service’ fashion. By enhancing Small Cells with an execution infrastructure, the inclusion of mobile edge computing capabilities emerges and this leads to increasing responsiveness from the edge of the network. This allows enriching the end users’ experience and at the same time, operators can open the radio network edge to third-party partners, allowing them to rapidly deploy innovative applications and services.

Fig. 3 demonstrates the proposed concept. An important aspect of this approach is that a service provider may exploit SESAME's architecture to compose and operate on an end-to-end basis his own 'virtual' network, provided by different mobile operators and infrastructure owners on top of a set of diverse physical infrastructure. Another instance could be a third party (as e.g., a virtual mobile network operator, a data centre provider or another business entity) that would act as end-to-end provider, without actually owning any physical infrastructure. That part would sign agreements with physical infrastructure owners, for instance municipalities, big stadiums, etc., in the areas it wishes to provide network services without deploying infrastructure, in order to cover sporadic large scale events for example. The same would be the case for contracting data centre operators in strategic locations and in order to complete a full network solution.

Fig. 4 illustrates the main elements in the SESAME architecture, as well as a possible split of the different layers of the protocol stack in VNFs and Physical Network Functions (PNFs).

The different elements are defined as follows:

- Small Cell: Low-powered radio access node that operates in licensed and unlicensed spectrum with restricted range between some meters to 1 or 2 kilometres.
- Micro server (μS): Specific hardware that is placed inside the Small Cell and provides a virtualized computing infrastructure i.e., processing power, hardware accelerators, memory and storage capabilities. In other words, the μS provides the execution infrastructure.
- Cloud Enabled Small Cell (CESC): The Small Cell device which has been enriched with a micro server.
- Cluster of CESCs: A group of CESCs within a common intended coverage area (exchanging information between them, properly coordinated).
- Light Data Center (Light DC or LDC): Cloud entity composed by the micro servers of the CESCs forming a cluster.
- CESC Manager (CESCM): The architectural component in charge of managing and orchestrating the cloud environment of the Light DC and the radio access functionalities of the CESC. It can manage at the same time multiple clusters, a cluster or a single CESC.

CESCs are equipped with a virtualised execution environment, materialised in the form of the micro server, which allows the provision of mobile-edge computing capabilities to the mobile operators for enhancing the user experience and the agility in the service delivery. Although the Hardware platform is transparent to applications, Small Cell environment imposes severe physical constraints in term of power consumption, thermal dissipation and available space. For these reasons the SESAME's μS architecture is based on a very efficient System on Chip (SoC) with multi-core ARM CPU and internal HW accelerators. For high computational requirements the μS is able to host external standard PCI Express (PCIe) Cards equipped with different HW accelerators (FPGAs, DSPs, GPUs) or a Disk Controller with related disks in case of high capacity storage requirements. This feature adds the necessary flexibility for the HW platform to be used with maximum efficiency in various scenarios. The μS will also support the execution of VNFs for carrying out the virtualisation of the Small Cell access. In this respect, both network functions features, along with 'Service' VNFs, will be executed there. Moreover, the micro server VNFs will benefit from specific hypervisor extensions which enable computing and networking accelerations for virtual machines. Finally, backhaul and fronthaul transmission resources will be also part of CESC, allowing for the required connectivity.

Up to now, several attempts have been initiated targeting to find an optimal solution for the part of the small cell which needs to be virtualized, running as one or more VNFs, and the part of it that should remain to run as PNFs. The right-hand side of Fig. 4 shows one possible case of this debate. In this example, the Small Cell network functions above Packet Data Convergence Protocol (PDCP) layer have been transferred to the Light DC and are executed as VNFs. Examples of Small Cell VNFs may be related to Operations, Administration and Maintenance (OAM) and Self-Organizing Network (SON) capabilities within the cluster of CESCs. This may not however be confused with the case that the Light DC also executes other "APPS" or "Service VNFs", like for example, virtual firewall, virtual caching and so on.

The proposed execution platform of the µS is used to support the required VNFs that implement the different features/capabilities of the Small Cells, as well as the computing support for the mobile edge applications of the end-users. According to SESAME's proposed architecture, the virtual machines hosting the VNFs will be controlled by the local Hypervisor. Nevertheless, the overall management and coordination of VNFs will be assigned to higher layers. Depending on the actual virtualisation capabilities, clusters will be assigned to one or more Virtual Infrastructure Managers (VIM) – in line with the current ETSI NFV ISG approach and terminology [9] –i.e., entities that will be responsible for managing the virtualised resources required for proper VNF deployment. Monitoring data from all active VIMs will be combined for managing the whole process of VNF restructuring (e.g., migration, rescaling, etc.) in a dynamic and efficient way.

3. SESAME environment for mobile edge services

In this section we address how these types of mobile edge services can be implemented within the infrastructure of a mobile network operator, taking into account the current 3GPP LTE architecture and the possible path towards virtualization. Specifically, taking into account the possible protocol split presented in Fig. 4, this section focuses on the implications of the LTE data plane and the possible solutions to implement the data path.

According to [8], the main features to be covered for proper design of MEC services are:

- Traffic Offload Function (TOF), which allows the proper forwarding of data packets between the mobile data path and the MEC applications.
- Radio Network Information Services (RNIS), which makes it possible for MEC application to obtain Channel State Information (CSI) associated to the mobile users and performance statistics of the cell. This information is critical for the accurate provisioning of services tailored to the users' context of use.

The data path resulting from the standard 3GPP LTE network architecture is depicted in Fig. 5 (upper plot). All the user data traverses the Evolved Packet Core (EPC) through dedicated tunnels based on the GPRS Tunneling Protocol (GTP) protocol. These GTP-U tunnels exchange user data packets from the Packet Data Network Gateway (PGW) to the eNB, going through the Serving Gateway (SGW). The eNB is able to decapsulate the IP packets from the GTP-U tunnel and to forward them towards the corresponding User Equipment (UE).

The standard data path entails two main issues. First, all the IP data must traverse the EPC to the PGW for its forwarding to external applications. Second, taking into account the protocol split proposed in Fig. 4, the VNFs are not able to access the IP packets without breaking the GTP-U tunnel.

Different solutions exist to cope with both problems. For the first problem, the 3GPP introduced in Release 10 the concept of local breakout. Two main solutions are currently available: Local IP Access (LIPA) and Selected Internet IP Traffic Offload (SIPTO) [10, 11, 12]. Both options allow offloading some user traffic from the PGW through the addition of a local gateway with different alternative architectures. The SIPTO at the local network (SIPTO@LN) solution is the best candidate for Small Cells, the local breakout is not fully integrated into the Home eNB (HeNB).

In the scope of the SESAME architecture, the applications are designed to run within the Light DC following the MEC concepts. Therefore, using local breakouts within the Small Cells would entail significant advantages both from the management perspective and from the standpoint of reduced latency in the data path. The desired operation environment (lower plot in Fig. 5) would entail that the VNFs of the Small Cell (SC VNFs) are capable of handling the data path to/from the "Service VNFs" without going out of the CESC cluster.

The proposed solution for the local data path management in SESAME in illustrated in Fig. 6, and involves the coordinated use of the CESCM and a SDN controller.

The CESCM is the SESAME component that is in charge of the deployment, monitoring, configuration and administration of the mobile edge services instantiated over the hardware of the LightDC (resulting from the aggregation of the CESCs). A single CESCM can generically operate over a cluster of CESCs and consequently acquire an overall knowledge of the status of the virtual and physical resources across the different infrastructures.

The main functional components that build SESAME services are VNFs and it is therefore the role of the SESAME Orchestrator (in particular the VNF Orchestrator - located within the CESCM) to manage their deployment over the virtualized infrastructure offered by the VIM. This deployment of VNFs and their interconnection into composed edge services (the virtual infrastructure exposed by the VIM is made of the edge-deployed CESCs) is made with the awareness of different parameters and metrics that ensure an optimal allocation of compute resources and an efficient use of the radio access network in scenarios where end-users demand high performance and multiple network operators share the same infrastructure.

It is in this context that SESAME exploits the innovative aspect of RAN-aware cloud computing (a key concept for the convergence of mobile networks and cloud [13]), in which, together with processing, networking and storage, radio resources are also considered at the service level and as part of the execution platform.

Building over these principles, the CESCM obtains the ability to dynamically compose edge services and apply real-time monitoring procedures that allow the enforcement of the agreed SLAs between the different entities (CESC operators, mobile network providers, end-users). As a consequence of monitoring and taking advantage of the virtualized infrastructure, the CESCM can apply policies to mandate the reconfiguration, migration, and scale-out of the existing services in order to comply with the service agreements while dynamically maximizing the utilization of the resources.

As an input for the deployment of the requested edge services, the CESCM will draw from an infrastructure repository describing both the computational and radio capabilities of the CESCs of the cluster(s) and a service graph specifying the service to be deployed, its interconnection and dependencies with other services.

A fundamental role in the interconnection of the deployed VNFs is played by the SDN controller as a key entity to enforce the rules of traffic steering and chaining.

The emergence of the NFV concept and the expansion of VNF solutions have materialized service function chaining (SFC) among virtualized functions as a legitimate use case for a cloud data centre (DC). Yet in a premature state, applying SFC concepts in a fully virtualized environment requires changes and adaptations on the existing protocols in order to be able to apply the same concepts and achieve the desired behaviour on a network level. A typical burden in environments with fully virtualized functions running on virtual endpoints is the aggregated protocol encapsulation in the packets' headers added as they traverse those endpoints.

Currently there are two key approaches to implement SFC solution in virtualized scenario: packet based and flow based. The first requires manipulation of the packets, for instance by introducing some changes in the header field (packet tagging or rewrites) [14] or simply by applying protocols that introduce one more layer of abstraction on the top of the existing header fields – designed especially for this type of service. Such dedicated protocols has been ultimately introduced by Cisco and leveraged in the Open Daylight (ODL) community to support the ODL SFC integral project. In this case an additional header called NetworkServiceHeader (NSH), is introduced in order to enforce end-to-end traffic as an overlay connection above the service chain path. The problem of such solution is that it alters the datagrams and this can potentially cause a problem in the case where the VNF that runs on some of the virtual machine (VM) hops along the chain, requires the datagrams in their original structure. One example is a virtual function such as vDPI (virtual deep packet inspection) that requires the packets in their original structure in order to enforce a correct behaviour.

The advantage of SDN is that based on the Open Flow (OF) protocol, the routing can be steered over a specific networking path by programmatically applying OF based rules (flows) on the SDN controller or the virtual switch (OVS) inside the VM hosting the VNF. This is the second approach to implement SFC, based on flow programming rules, that leaves the packets untouched while applying actions on the OF ports of the switches, in order to gear the desired route of the packets in the chain. This routing logic is simplified compared to the first one, as it avoids unnecessary overheads on the top of the already existing ones (ex. in the scenario of inter tenant communication in OpenStack [15]) and the packets are left intact, completely agnostic of the existing chain. A solution based on this approach requires that the network environment is fully SDN capable in order to apply the chain rules along the full virtual network graph. The resulting routing flows need to be maintained to reflect alterations in the function chain (e.g. a VNF altering the packet header could invalid the end-to-end chain). Higher level chaining abstractions and programming languages are needed in order to allow service developers to programmatically declare the sequence the VNF traffic should follow, leaving to the underlying runtime system the actual implementation of such rules [16, 17]. For the chain routing to be deterministic, there has to be a field that keeps track of the chain hops. Using the VLAN ID as a workaround for this purpose could be one possible approach, since the chain routing does not follow the standard Ethernet routing.

Extending the concept of SFC for the scope of the SESAME project and in the context of the 5GPPP ecosystem, requires deeper understanding on the NFV concepts in such a scenario and carefully elaborating the requirements to establish the desired functionality in environment that is not fully prepared to support it. Today there exist VNFs deployments as substitutes for the EPC integral blocks as well as on the radio access front haul side. SFC concepts based on SDN can be applied in LTE environment to enforce the packets among a logical network graph and achieve certain service functionality among the virtualized components such as BSS, HSS, RAN etc [18]. The role of the controller in this holistic approach has been analyzed in the literature and some experimental and industry implementations already have been presented [19-23]. However SFC solution based on SDN in this scenario is an unexplored area and potentially one of the use cases to be addressed in SESAME.

From a single data centre point of view, the SDN controller stands in between the components such as CESCM / VIM and the light DC. In this case, the previously described approach for SFC applies if the SDN controller takes the charge of a single light DC deployment. The steering and rule enforcement policy are kept within the controller application logic and enforced over the network that hosts the light DC specific NFVs. If the routing happens across different light DCs, then the SFC approach may alter depending on the placement of the controller within the given architecture. This has to be in accordance with the networking protocols to be adopted in that case, for example MPLS and BGP as used today for intra data-centre routing.

4. Scenarios of applicability

Three initial target scenarios have been identified as promising fields for the applicability of the SESAME concepts, which can be used as the basis for the formulation of a number of specific use cases. A description of the three scenarios is given in the following, highlighting which are the main challenges, the applications and services in-scope and the SESAME components and capabilities.

4.1. Scenario 1 - Multi-tenant enterprise services

A typical scenario in which the SESAME system can be exploited is shown in Fig. 7. The figure depicts a situation of one CESC provider which owns, deploys and maintains the network infrastructure of Small Cells and Light DC (i.e. ensemble of micro-servers) inside the premises where different enterprises are hosted. In this case the CESC provider shall establish a Service Level Agreement (SLA) with each customer enterprise to enable enterprise users to access different services, including Internet access, voice communications, video conferencing, access to mail system and repositories, web browsing and open and closed subscriber groups with embedded high security credentials , just to name a few. The deployment of µSs can lead to achieve close-to-zero latency, with clear benefits for enhanced Quality of Experience (QoE) of media flows. Indeed this can be achieved resorting to content caching at the level of the Light DC, or in other words storing content at different µS locations. It is also worth stressing the fact that in a headquarter, hosting different enterprises traffic may fluctuate greatly depending on the time of the day and on the nature of specific events held, thus requiring a flexible system which can be scaled up and down depending on the situation. As shown in Fig. 7, the enterprise scenario will leverage on SESAME features such as intrinsic support of multi-tenancy since Small Cells operators can provide network services and connectivity over the network owned by the infrastructure provider. Furthermore, SESAME will optimize service delivery to the enterprise end users adapting the network behaviour by means of self-organizing network techniques.

4.2. Scenario 2 - Enhanced service experience on the move

In this scenario, a CESC provider manages three distributed CESC clusters deployed in geographically adjacent areas and supports a single mobile network provider that is offering services to his end users through the CESC infrastructures. The relationship between the entities is regulated by different SLAs which are established between the CESC operator and the service provider and between the service provider and its end-users. The main actors and interactions of this scenario are depicted in Fig. 8.

To demonstrate different SESAME capabilities in this setup, the mobility of a reference end user is taken into account together with his requirements for service continuity and quality as he handovers across the different CESC clusters. The type of traffic generated by the user is assumed to be high-definition real-time content that requires low-latency data access times.

The application of the MEC paradigm, implemented in the SESAME architecture through hardware and software components running at the edge of the network and in the proximity of the moving user, allows for an efficient monitoring of the user location and its related radio conditions, with real-time reporting and the actuation of coordination actions with close to zero latency.

The enhanced handling of the end-user mobility is thus a consequence of the SON or self-x features of the CESC clusters, which take advantage of edge monitoring capabilities to seamlessly manage the handover process across neighboring cells. As for decision making, limiting the processes within the boundaries of a CESC (or CESCs cluster), allows for the fast application of policies aimed at increasing the overall quality perceived by the user. In order to take into account more stringent requirements (for example across all users), some decision making processes resulting in the reconfiguration of services for the fulfillment of the SLAs can also be completed at the level of the CESCM, which has a wider view of the status of the resources across the CESC clusters.

With respect to user traffic, the scenario highlights how low-latency edge caches can be deployed across the CESC installations to allow the end user an uninterrupted access to content. Pre-provisioning cached data while anticipating the user handover (monitored through signal inspection) is an effective way to offload the user equipments from the task of storing data locally.

4.3. Scenario 3 – Management of flash events

Another relevant scenario addressed by the SESAME system is presented in Fig. 9 in which the sudden concentration of people at a specific geographical location and time of the day creates an unexpected hot spot zone in which a variety of different traffic types require proper management. The sudden gathering of crowds can be due to unexpected live events or emergency situations. This scenario is relevant to leverage the CESC cluster resources, which are essentially the collection of a number of CESCs (i.e. Small Cells with their micro-servers). In addition, the scenario allows showing that multi-tenancy can be considered a built-in function of the system. In the first place indeed, the CESC infrastructure deployed by an infrastructure provider shall support different mobile operators in serving their customers. In such a situation CESC cluster resources have to be provisioned to each tenant operator, and in order to efficiently handle the unexpectedly intense traffic generated by the users, self-organizing network techniques are required. At the beginning, the CESCM shall interface with an operator's OSS/BSS to retrieve tenant configuration parameters; afterwards, communications and QoS are supported at the CESC level. It is interesting to notice that this scenario is built on the edge computing capability of the system since computationally intensive tasks can be offloaded from the mobile terminals to the μSs, while at the same time optimizing the use of backhaul resources. The two main traffic types which need to be supported and optimized by the CESC cluster are live video streaming (e.g. users film and posting video contents in social media) and real-time group communications (e.g. small community of users exchanging files or videos). Deep packet inspection, video transcoding and data analytics can be enabled in the CESC cluster whereby hardware accelerators in order to optimize the management of the traffic from/to the CESCs.

5. Conclusions and further work

This paper presents a solution for the inclusion of mobile edge services in scenarios where an aggregation of small cells is able to share certain processing and storage resources in a local perspective. Therefore, rather than centralizing RAN and MEC functions into big data centres, the proposed solution creates a distributed light data centre by coordinating the micro-servers associated to different small cells in a cluster.

The proposed solution is based on the SESAME architecture, which integrates concepts of mobile edge virtualization and multi-tenancy to provide a unified coordination for the cluster of Cloud-enabled Small Cells. From this architecture, we analyze the problem of handling user data packets within the CESC cluster according to the standard LTE user data path, and especially when the virtualization split is made above the PDCP layer. The proposed solution includes a coordinated management of the VNFs from the CESC Manager element, and the implementation of the data forwarding functions through an SDN Controller. This solution provides the foundations for the implementation of advanced Service Function Chaining (SFC).

After the SESAME architecture and the possible introduction of mobile edge services are discussed, the paper presents a series of possible scenarios where the SESAME solution would provide significant enhancements over the current network management strategies. In order to exemplify the expected performance improvements of the SESAME solution, preliminary simulation studies have demonstrated the associated benefits for the problem of general flow scheduling [24] and for multimedia delivery [25] in multi-user scenarios.

In the framework of this work, several open issues are identified as further work.

The presented overall architecture provides a step forward towards the integration of the 3GPP LTE and ETSI NFV approaches. However, a detailed definition is required for the NFV management elements according to the standardization efforts such as ETSI NFV Management and Orchestration (NFV-MANO) [26] or the recently initiated 3GPP approach for management of mobile networks that include virtualized network functions (MAMO-VNF) [27].

The proposed multi-tenant virtualized infrastructure of cloud-enabled small cells brings new opportunities to the business ecosystem. For example, the role of the small cell provider that supports multiple mobile network operators shall be explored. In this sense, not only the management and performance aspects but also the security and privacy of the multiple tenants need to be carefully studied.

Finally, a series of research challenges that need to be properly addressed remain open for discussion, such as the Virtual Network Embedding (VNE) problem in mobile edge computing networks [28] and the dynamic coordination of CESC clusters in terms of virtualized RRM and SON.

References

1. Soldani D, Manzalini A (2015) Horizon 2020 and Beyond: On the 5G Operating System for a True Digital Society. IEEE Vehicular Technology Magazine 10(1):32-42

2. Chih-Lin I, Huang J, Duan R, Cui C, Jiang J Li L (2014) Recent Progress on C-RAN Centralization and Cloudification. IEEE Access 2:1030-1039

3. Wu J, Zhang Z, Hong Y, Wen Y (2015) Cloud Radio Access Network (C-RAN): A Primer. IEEE Network 29(1):35-41

4. Rost P, Bernardos CJ, Domenico AD, Girolamo MD, Lalam M, Maeder A, Sabella D, Wübben D (2014) Cloud technologies for flexible 5G radio access networks. IEEE Communications Magazine 52(5):68-76

5. Next Generation Mobile Networks (NGMN) Alliance (2015) Project RAN Evolution: Further Study on Critical C-RAN Technologies. https://www.ngmn.org/publications/all-downloads/article/project-ran-evolution-further-study-on-critical-c-ran-technologies.html

6. Small Cell Forum (2015) Document 106.05.1.01: Virtualization for Small Cells: Overview. http://scf.io/doc/106

7. Small Cell Forum (2015) Document 159.05.1.01: Small Cell Virtualization Functional Splits and Use Cases. http://scf.io/doc/159

8. ETSI Industry Specification Group Mobile-edge Computing, http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing

9. ETSI Industry Specification Group for Network Function Virtualisation, http://www.etsi.org/technologies-clusters/technologies/nfv

10. 3GPP TR 23.829: Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO)

11. 3GPP TR 23.859: LIPA Mobility and SIPTO at the Local Network

12. 3GPP TR 22.828 Study on co-ordinated Packet data network GateWay (PGW) Change for Selected IP Traffic Offload (CSIPTO)

13. Taleb T; Corici M; Parada C; Jamakovic A, Ruffino S; Karagiannis G; Magedanz T (2015) EASE: EPC as a service to ease mobile core network deployment over cloud. IEEE Network 29(2): 78-88

14. Service Function Chaining in Open Daylight using NSH protocol: https://wiki.opendaylight.org/view/Service_Function_Chaining:Main

15. Overlay encapsulation in DC network http://docs.openstack.org/developer/neutron/devref/openvswitch_agent.html

16. Anwer B, Benson T, Feamster N, Levin D (2015) Programming slick network functions. In Proc. the 1st ACM SIGCOMM Symposium on Software Defined Networking Research (SOSR '15), ACM, New York, NY, USA, article 14, 13 pages

17. Riggio R. et al (2016) Programming Wireless Network Functions. In Proc. IEEE NOMS 2016, Istabul, Turkey

18. Riggio R, Bradai A, Rasheed T, Ahmed T, Slawomir K, Schulz-Zander J (2015) Virtual Network Functions Orchestration in Wireless Networks. In Proc.11th International Conference on Network and Service Management (CNSM), Barcelona, IFIP/IEEE

19. Akyildiz IF, Wang P, Lina S-C (2015) SoftAir: A software defined networking architecture for 5G wireless systems. Computer Networks 85:1-18

20. Basta A, Kellerer W, Hoffmann M, Hoffmann K; Schmidt E-D (2013) A Virtual SDN-Enabled LTE EPC Architecture: A Case Study for S-/P-Gateways Functions. In Proc. 2013 IEEE SDN for Future Networks and Services (SDN4FNS), pp.1-7

21. Malik A, Qadir J, Ahmad B, Alvin Yau KL, Ullah U (2015) QoS in IEEE 802.11-based wireless networks: A contemporary review. Journal of Network and Computer Applications 55:24-46

22. Chourasia S; Sivalingam KM (2015) SDN based Evolved Packet Core architecture for efficient user mobility support. In Proc. 2015 1st IEEE Conference on Network Softwarization (NetSoft) pp.1-5

23. He J, Song W (2015) Smart routing: Fine-grained stall management of video streams in mobile core networks. Computer Networks 85:51-62

24. Fajardo JO, Taboada I, Liberal F (2015) Radio-Aware Service-Level Scheduling to Minimize Downlink Traffic Delay Through Mobile Edge Computing. In Proc. MONAMI 2015, LNICST 158, pp. 1–14
25. Fajardo JO, Taboada I, Liberal F (2015) Improving Content Delivery Efficiency through Multi-Layer Mobile Edge Adaptation. IEEE Network November/December 2015
26. ETSI GS NFV-MAN 001 V1.1.1 (2014) Network Functions Virtualisation (NFV); Management and Orchestration
27. 3GPP TS 28.511: Telecommunication management; Configuration Management (CM) for mobile networks that include virtualized network functions; Procedures
28. Beck MT, Maier M (2014) Mobile Edge Computing: Challenges for Future Virtual Network Embedding Algorithms. In Proc. The Eighth International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP 2014)

Fig. 1 Fully centralized vs. partially centralized RAN functional architecture

Fig. 2 Geographical alternatives for the data centre: centralized vs. distributed

Fig. 3 SESAME concept and impact on service deployment

Fig. 4 High level SESAME components

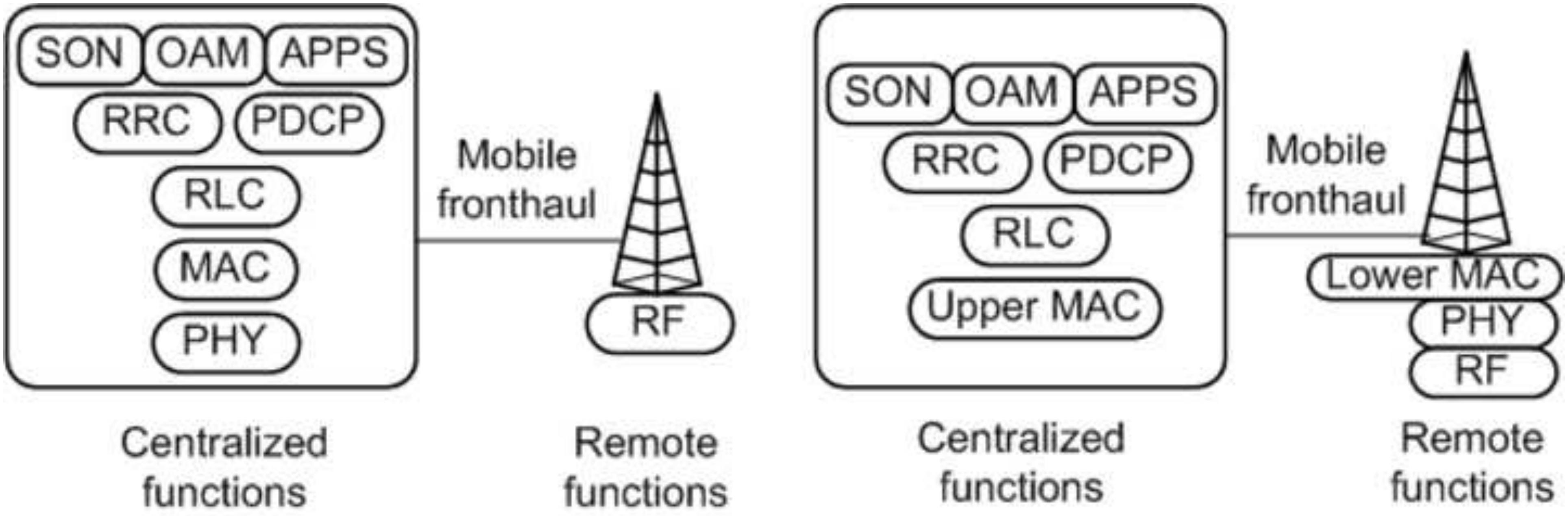
Fig. 5 Data path in typical LTE connection and in SESAME approach

Fig. 6 SESAME solution for traffic forwarding at mobile edge

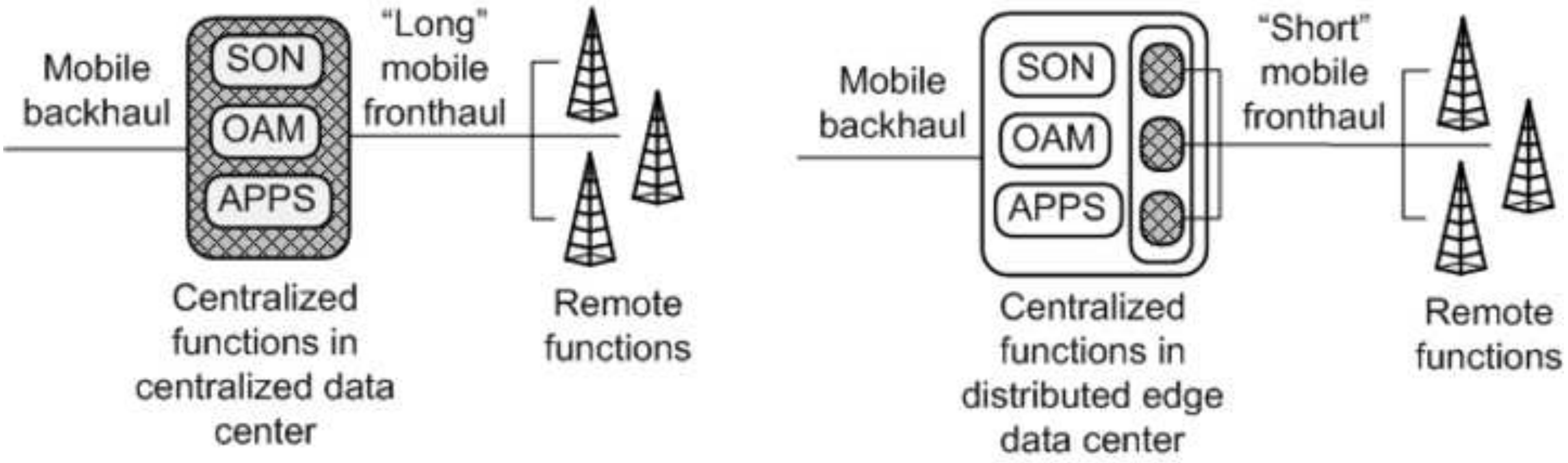
Fig. 7 SESAME system for multi-tenant enterprise services

Fig. 8 SESAME system for enhanced service experience on the move

Fig. 9 SESAME system for flash events



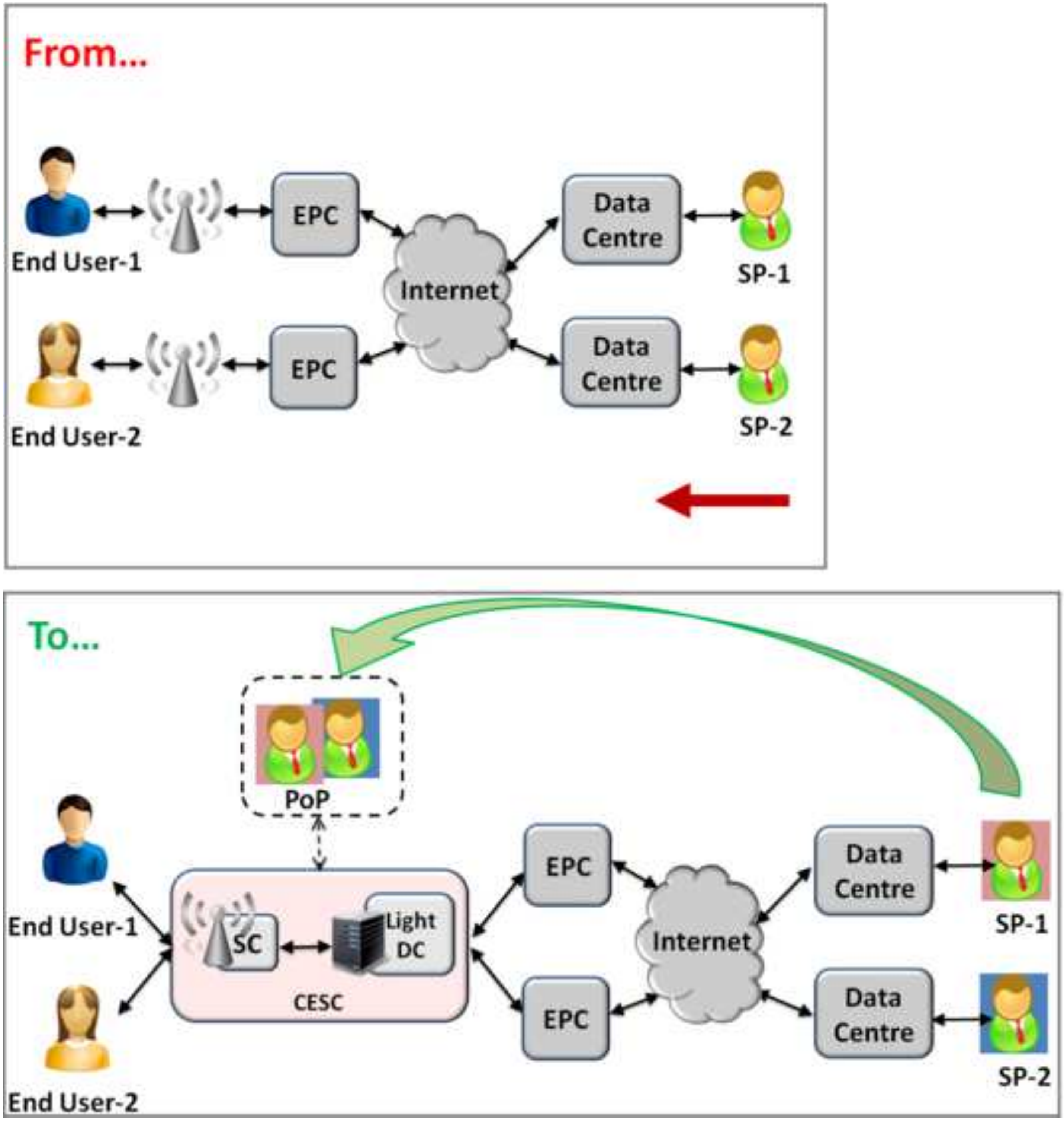
a) fully centralized RAN b) partially centralized RAN

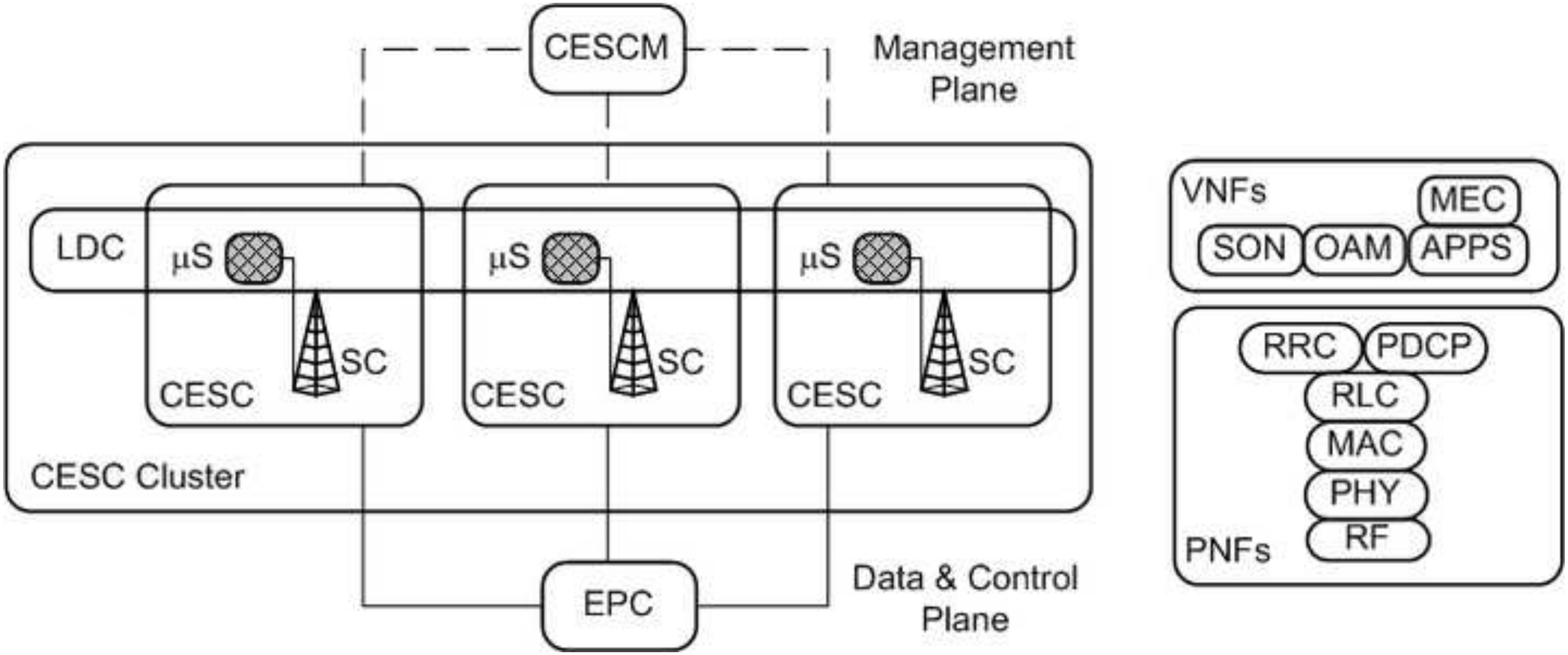


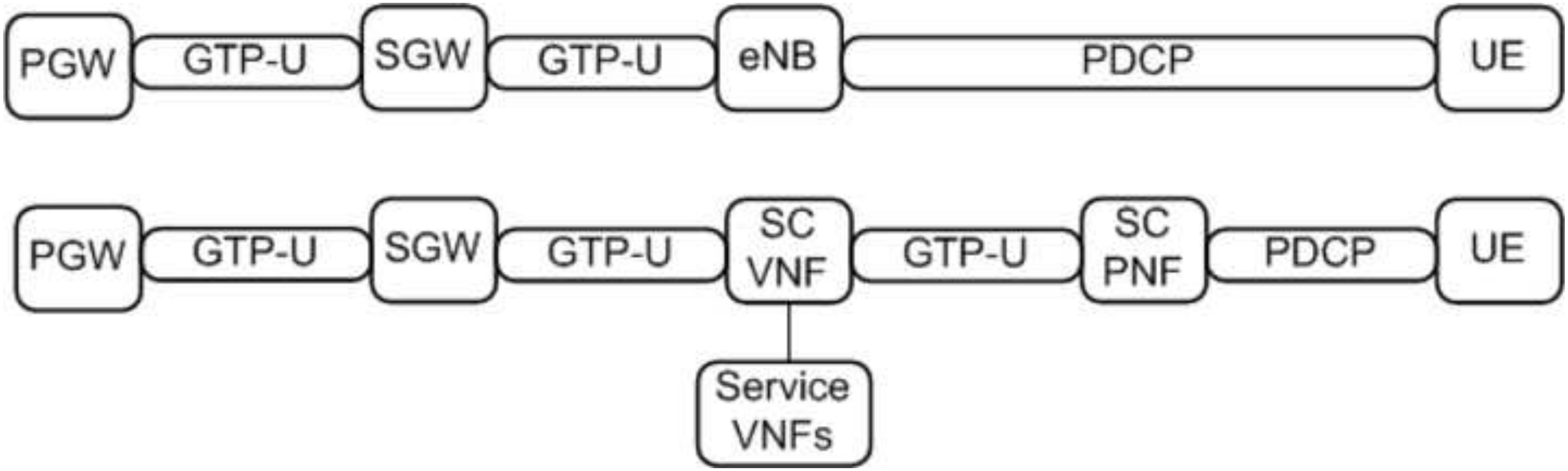
a) physically centralized data centre

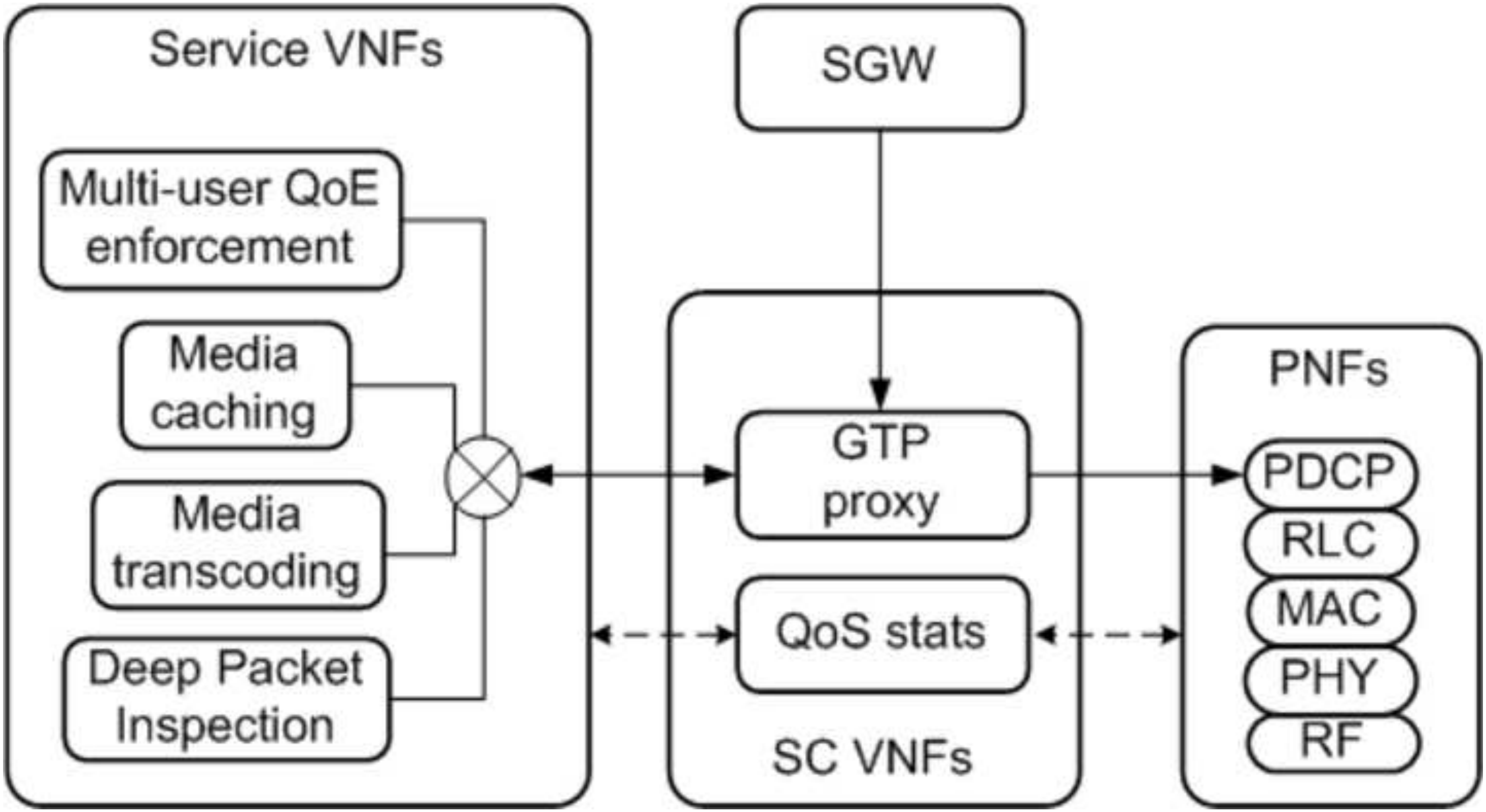
b) physically distributed data centre

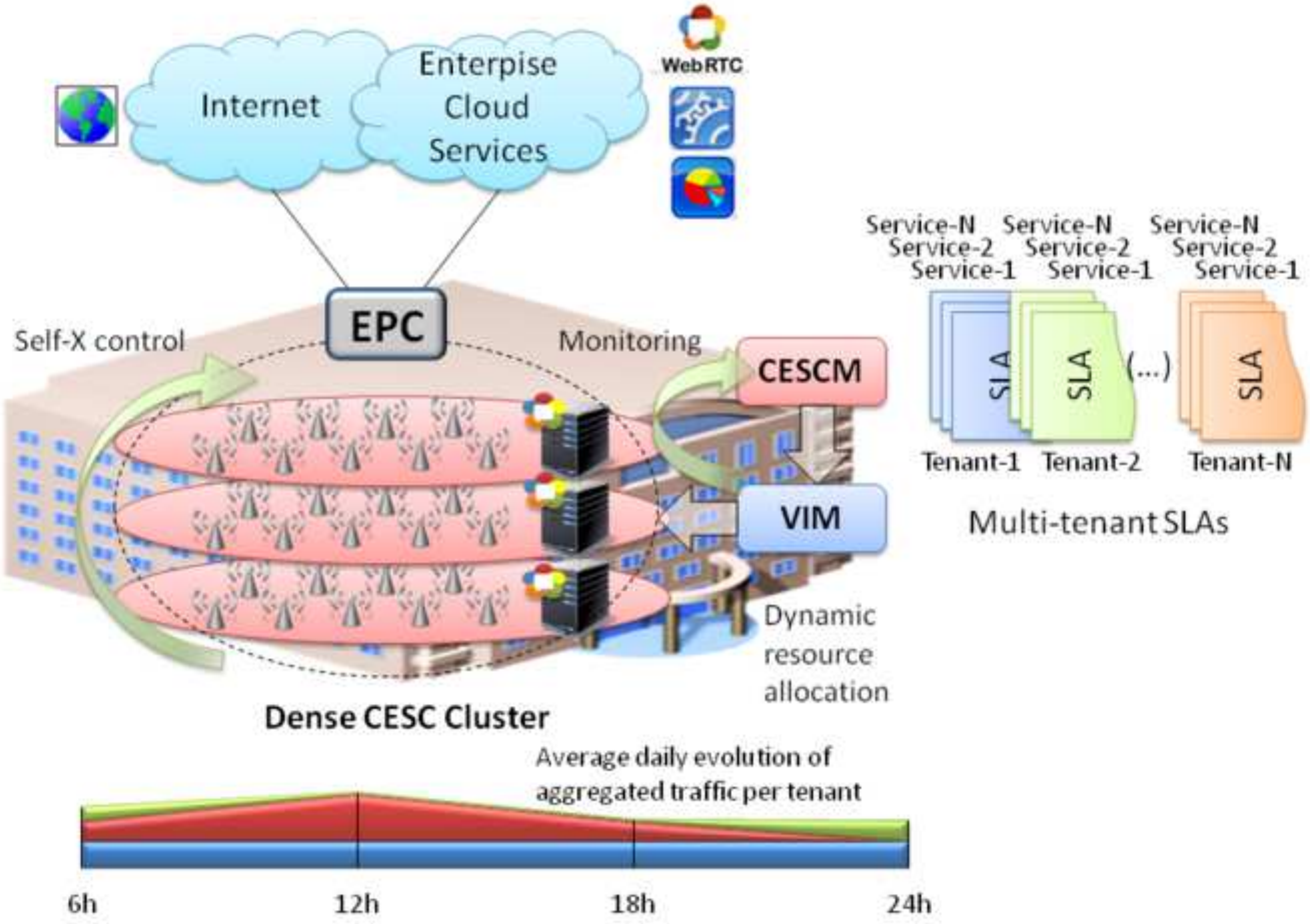
Fig3


From...
End User-1
EPC
Internet
Data Centre
SP-1
End User-2
EPC
Data Centre
SP-2
To...
PoP
End User-1
SC
Light DC
CESC
End User-2
EPC
EPC
Internet
Data Centre
Data Centre
SP-1
SP-2


CESCM
Management Plane
LDC
µS
µS
µS
SC
SC
SC
CESC
CESC
CESC
CESC Cluster
EPC
Data & Control Plane
VNFs
MEC
SON
OAM
APPS
RRC
PDCP
RLC
MAC
PHY
RF
PNFs


PGW
GTP-U
SGW
GTP-U
eNB
PDCP
UE
PGW
GTP-U
SGW
GTP-U
SC VNF
GTP-U
SC PNF
PDCP
UE
Service VNFs


Service VNFs
Multi-user QoE enforcement
Media caching
Media transcoding
Deep Packet Inspection
SGW
GTP proxy
QoS stats
SC VNFs
PNFs
PDCP
RLC
MAC
PHY
RF


Internet
Enterprise Cloud Services
WebRTC
EPC
Self-X control
Monitoring
CESCM
VIM
Dynamic resource allocation
Dense CESC Cluster
Service-N
Service-2
Service-1
SLA
(...)
Tenant-1
Tenant-2
Tenant-N
Multi-tenant SLAs
Average daily evolution of aggregated traffic per tenant
6h
12h
18h
24h

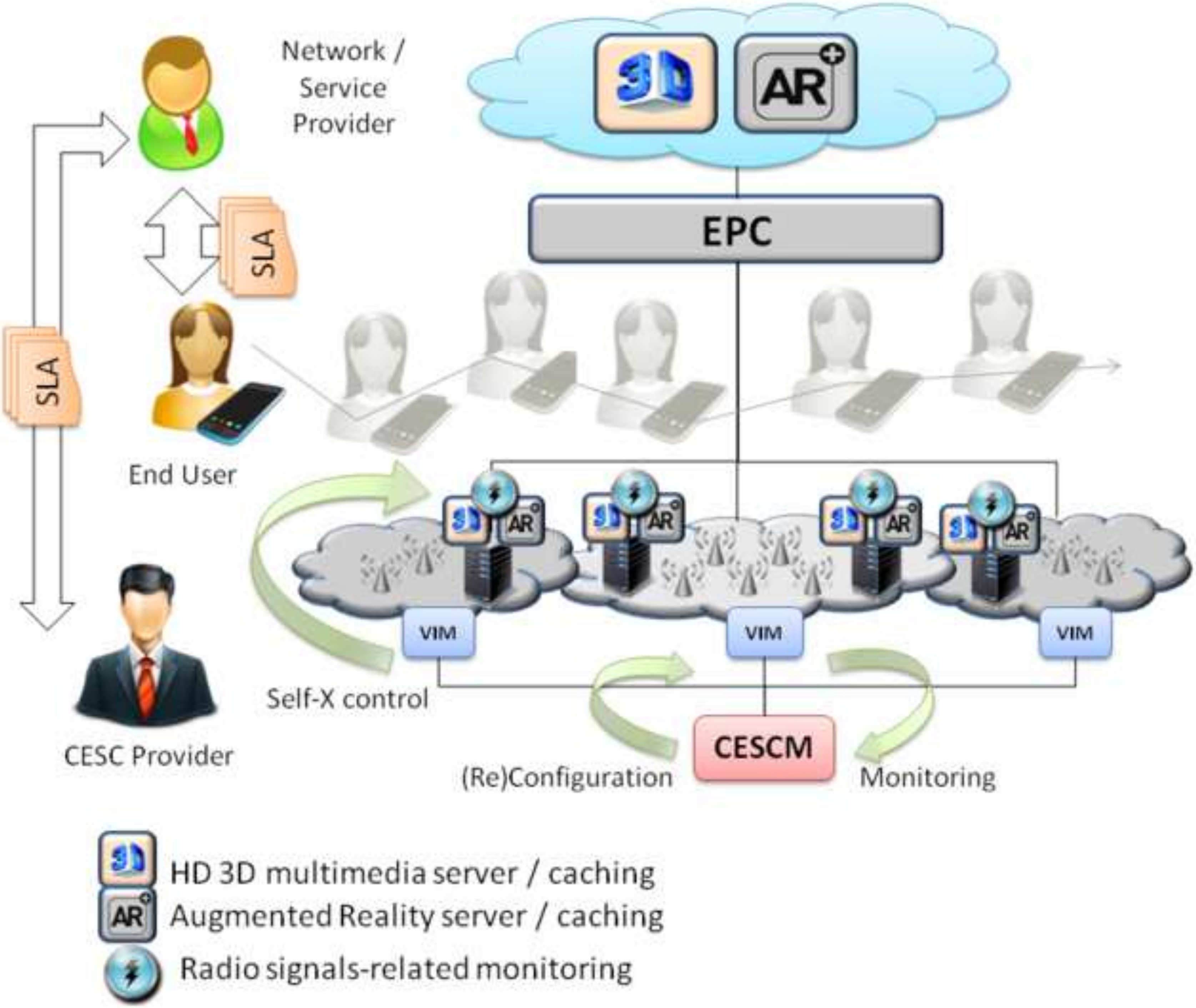
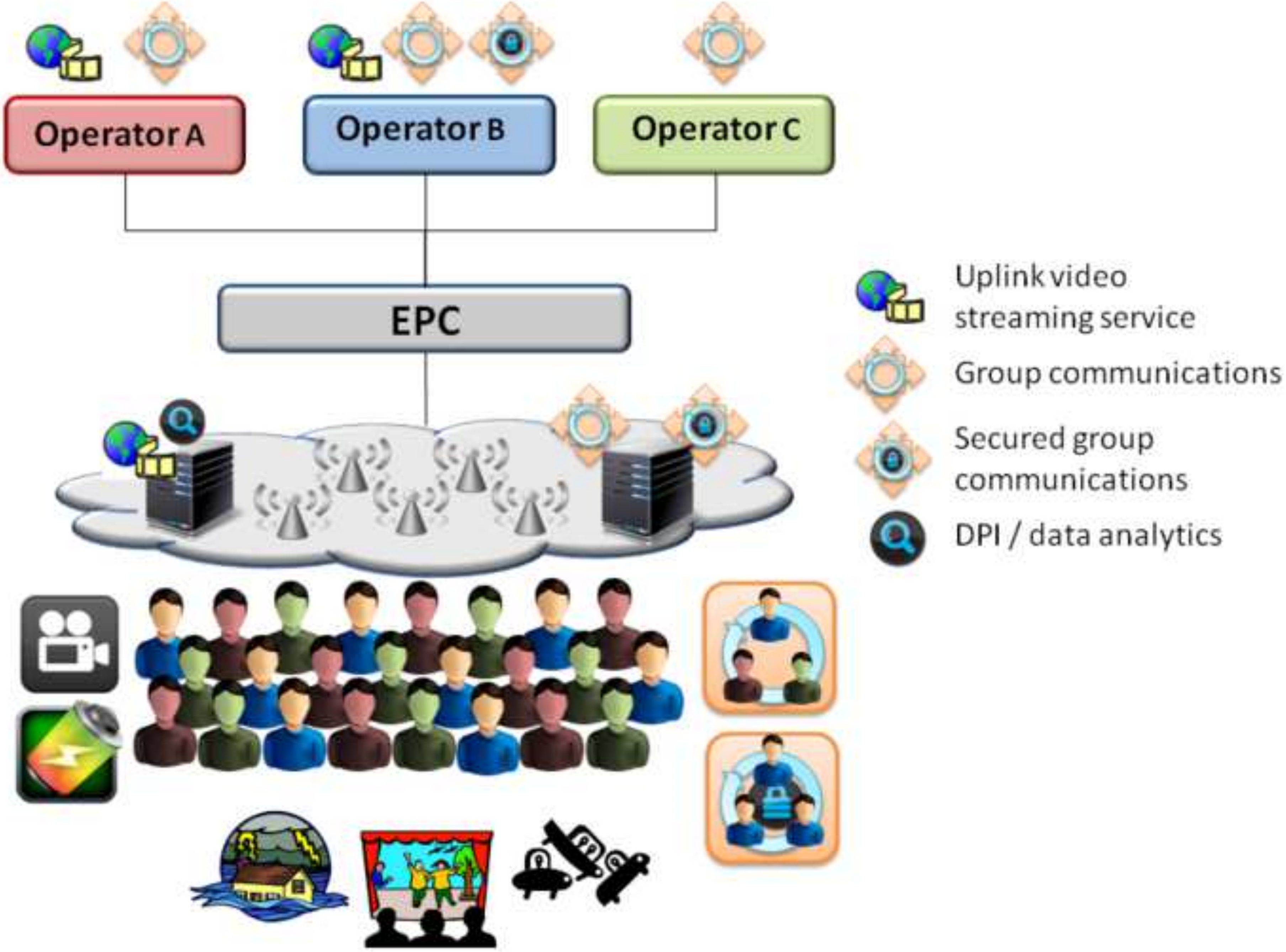

Network /
Service
Provider
3D
AR
EPC
SLA
SLA
End User
VIM
VIM
VIM
Self-X control
CESC Provider
CESCM
(Re)Configuration
Monitoring
HD 3D multimedia server / caching
Augmented Reality server / caching
Radio signals-related monitoring

Fig9


Operator A
Operator B
Operator C
EPC
Uplink video streaming service
Group communications
Secured group communications
DPI / data analytics