TRANSACTIONS ON EMERGING TELECOMMUNICATIONS TECHNOLOGIES
Trans. Emerging Tel. Tech. (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/ett.2915

INVITED PAPER

SDN-based 5G mobile networks: architecture, functions, procedures and backward compatibility

R. Trivisonno*, R. Guerzoni, I. Vaishnavi and D. Soldani
Huawei European Research Institute, Munich, Germany

ABSTRACT

In this paper, we describe an SDN-based *plastic architecture* for 5G networks, designed to fulfill functional and performance requirements of new generation services and devices. The 5G logical architecture is presented in detail, and key procedures for dynamic control plane instantiation, device attachment, and service request and mobility management are specified. Key feature of the proposed architecture is *flexibility*, needed to support efficiently a heterogeneous set of services, including Machine Type Communication, Vehicle to X and Internet of Things traffic. These applications are imposing challenging targets, in terms of end-to-end latency, dependability, reliability and scalability. Additionally, *backward compatibility* with legacy systems is guaranteed by the proposed solution, and Control Plane and Data Plane are *fully decoupled*. The *three levels of unified signaling* unify Access, Non-access and Management strata, and a *clean-slate forwarding layer*, designed according to the software defined networking principle, replaces tunneling protocols for carrier grade mobility. Copyright © 2014 John Wiley & Sons, Ltd.

***Correspondence**
R. Trivisonno, Huawei European Research Institute, Munich, Germany.
E-mail: riccardo.trivisonno@huawei.com

Received 13 October 2014; Revised 5 November 2014; Accepted 8 November 2014

1. INTRODUCTION

The evolution of mobile broadband networks is currently driven by the definition of next generation services and related requirements. Within mainstream research frameworks, key challenges for 5G are being identified [1–3]. Beyond system capacity and scalability requirements (thousand-fold capacity increase with respect to 4G and ability to provide connectivity to billions of devices), tight constraints in terms of latency and reliability will be also imposed by delay critical services, such as virtual reality office, teleprotection for smart grids, real-time remote computing for mobile terminals, and traffic safety and efficiency, which are regarded as key use cases for 5G networks [1]. Moreover, next generation systems are expected to enable brand new classes of services, required by emerging vertical markets, including Machine Type Communication (MTC), vehicle to X (V2X) communications and Internet of Things (IoT) traffic.

The most stringent requirements trigger ad hoc research activities, aiming at boosting performance of specific enabling technologies of future networks. For example, 5G capacity requirements call for spectral efficiency improvements, which stimulate research on interference management (e.g. see [4, 5] and [6] for device to device communications). Besides specific performance requirements, the *heterogeneity* of services, devices and access networks 5G shall support, will undoubtedly impose radical changes to network architecture. It is worth noting that 3rd Generation Partnership Project and ETSI have already addressed the issue, planning functional adjustments to 4G, to optimise the management of diverse use cases and devices (see [7] and [8], where ad hoc solutions for MTC communications and multicast-broadcast multimedia services are analysed). In order to manage such heterogeneity, *flexibility* is going to be the key feature of next generation networks. The required architectural flexibility can be achieved *by design*, leveraging Software Defined Networking (SDN), Network Functions Virtualisation, and cloud and edge computing paradigms.

Several prominent proposals of SDN-based architectures were recently published. In [9], authors described a generic Software Defined Wireless Network architecture, based on a common core network (CN) and a variety of radio access networks (RANs), orchestrated by a Mobile Network SDN Controller. In the proposed architecture, RANs are enhanced with programmability, and the transport network consists of programmable switches and routers. Both 'evolutionary' and 'clean-slate' approaches were proposed. The first approach allows incremental deployment in existing networks. In this case, the SDN controller implements the required standardised interfaces. In the latter clean-slate approach, Control Plane (C-Plane) functions are directly programmed into the SDN controller. Similarly, in [10], authors discussed a 5G mobile architecture made by two network layers: a radio network, providing a minimum set of L2 and L1 functionalities, and a network cloud dedicated to all higher layer functionalities. The proposal claims a lean protocol stack that can be achieved by consolidating redundant Access Stratum (AS) and Non-access Stratum (NAS) functionalities. Such integration allows a simplification of mobility management, session management and security procedures. On the user plane, dynamic network deployment and ability to scale are achieved by merging RAN L2 and gateway functionalities in the CN. Another notable proposal can be found in [11]. The paper describes an all-SDN network architecture featuring hierarchical control capability. More specifically, the article focuses on a 5G C-Plane aiming at providing Connectivity Management (CM) as a Service, with a 'unified' approach to mobility, handoff and routing management. According to the authors, 'Unified' relates to merging RAN and CN functions, which are implemented as network applications running on one or more hierarchical controllers.

Harvesting some high-level concepts presented in prior art, this paper develops an SDN-based plastic architecture for 5G networks. The term *plastic* relates to the key characteristic of our proposal: both C-Plane and Data Plane (D-Plane) are not prescribed; rather, the C-Plane and D-Plane are tailored according to the requirements of specific applications and devices. Any network function, and corresponding states, is dynamically instantiated and interconnected within the cloud infrastructure, according to application and device performance and relevant functional requirements. This approach aims at achieving efficiently the target 5G key performance indicators by optimising the physical infrastructure utilisation and allowing backward compatibilities with legacy communication systems. Indeed, compatibility with 4G systems is regarded by the authors as a strict prerequisite for the adoption and market uptake of the proposed concepts for the next generation networks.

Section 2 reviews briefly the key 5G performance metrics and highlights the design rationale we followed to conceive the proposed plastic architecture, which is described in Section 3. Section 4 presents in detail how the new procedures for C-Plane instantiation, device attachment, service request and mobility would efficiently work. Section 5 includes a proof of backward compatibility, as it shows how a 4G network could be instantiated using our plastic architecture as a platform. Finally, Section 6 concludes the paper.

2. 5G MOBILE NETWORK DESIGN

2.1. 5G requirements

This paper focuses on 5G system requirements affecting the architecture definition according to the research challenges presented in [12]. The key requirement for 5G architecture is *flexibility*, as networks must support a *heterogeneous* set of use cases efficiently. Among those, V2X communications and IoT traffic will place challenging functional and performance requirements on network platforms and devices. In particular, V2V communication requires a significant latency reduction with respect to 4G, leading to a challenging 5 ms end-to-end delay. Additionally, reliable V2V services shall be provided in a variety of cases, including the out-of-coverage scenario, where one or both devices in communication cannot get network connectivity. IoT traffic requires an extended device ID space (enhanced addressing schemes), well beyond IPv6. Moreover, *flexibility* is also about backward compatibility: 5G architecture shall be able to support and coexist with legacy (e.g. 4G) systems.

2.2. 5G design rationale

To meet the requirements briefly reviewed in the previous section, we propose new fundamental design principles. First, 5G architecture shall feature dynamic definition and instantiation of both C-Plane and D-Plane, according to service and device types. The customised C-Plane and D-Plane shall enable ad hoc networking procedures (e.g. device attachment and service request) designed to achieve the target 5G key performance indicators. Second, the architecture shall support full (i.e. extended up to the access point) C-Plane and D-Plane decoupling, leveraging SDN technologies for transport network implementation and to allow radio dual connectivity. Finally, D-Plane shall be simplified by removing the tunneling concept and introducing *soft* packet data network (PDN) and mobility anchoring points.

3. 5G ARCHITECTURE

3.1. General principles

The proposed architecture stems from concepts presented in [13]. The architecture consists of a *unified* C-Plane, made by three *logical* controllers, and a clean-slate D-Plane, both represented in Figure 1.

The 5G unified C-Plane includes both AS/NAS signaling and management functions. The implementation of network functions can be either 'centralised' or 'distributed at the edge', depending on functional and performance requirements of the supported services. Additionally, the C-plane includes an SDN plane through which D-Plane is instantiated and operated. The selection of logical func-

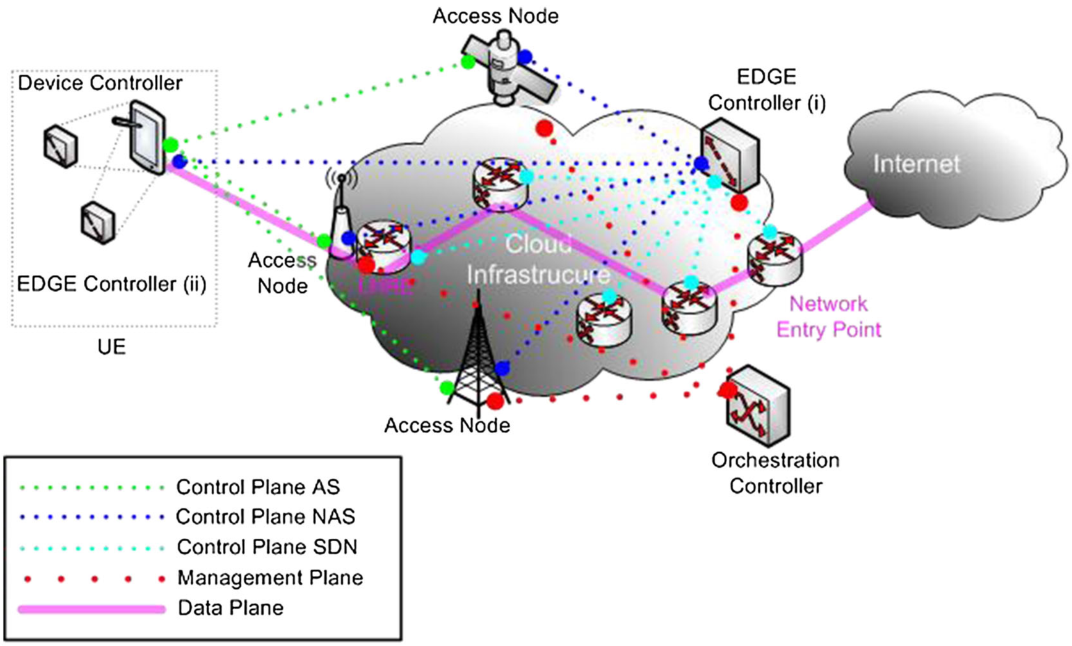

Figure 1. 5G High-level architecture—control and data planes.

tions to be supported has been made considering 4G control and D-plane logical elements (standardised in [14]) and decomposing them into sets of *applications* and *modules*, which can be dynamically instantiated in the cloud infrastructure according to network operation and service requirements. The D-plane consists of *forwarding paths,* which are set up in the cloud infrastructure, according to the SDN < *Match,Action*> principle [13].

3.2. 5G control plane

The unified C-Plane integrates AS/NAS and management functions in three logical controllers: the device controller, the EDGE controller(s) (ECs) and the orchestration controller (OC). The C-Plane encompasses all 5G system elements, from devices to access nodes and CN nodes. Specificities of devices and access nodes are out of the scope of this paper.

Device controller: The device controller (DCON) is responsible for the device physical layer connectivity to the 5G network. In particular, in a 5G scenario where devices are expected to have a plurality of radio and wired access capabilities, the DCON handles AS functions such as access selection and network selection [14]. As shown in Figure 2, DCON includes the user equipment (UE) radio access (RA) application.

EC(i): The EC implements the 5G network C-Plane, including network access control, packet routing and transfer, mobility and connection management, security, QoS and radio resource management functions ([14]). In other words, referring to legacy 4G networks, the EC inherits the AS/NAS functions performed by eNodeB and Mobility Management Entity (MME). The implementation of the EC is distributed over the cloud infrastructure via a set of interconnected control applications (C-Apps), as shown in Figure 2. Each C-App is dedicated to a subset of network control functions. Key C-Apps are CM, mobility management (MM), security (Sec), authorisation & authentication (AA), admission control (AC), flow management (FM) and RA applications. To allow full separation between C-Plane and D-plane even on the radio link, the RA App is further split into RA^D and RA^C Apps. For an attached device, RA^C App manages the C-Plane, while the RA^D App manages the D-Plane, possibly instantiated on a different point of presence (PoP). Note that EC(i) relates to a logical element composed by C-Apps instantiated in the cloud infrastructures.

EC(ii): As one of the key use cases for 5G is V2V communication, which has to be supported also in the out-of-coverage scenario (i.e. when one or all vehicles in communication cannot directly connect to the network), it might be required that some of the AS/NAS functions need to be located in the mobile devices. For this reason, the proposed architecture distinguishes between EC(i) and EC(ii), the first being composed by C-Apps instantiated in the cloud infrastructures and the second temporarily or permanently implemented in a mobile device. This allows some devices to act as *cluster head* in case of out-of-coverage scenario, taking over the C-Plane management on behalf of the network.

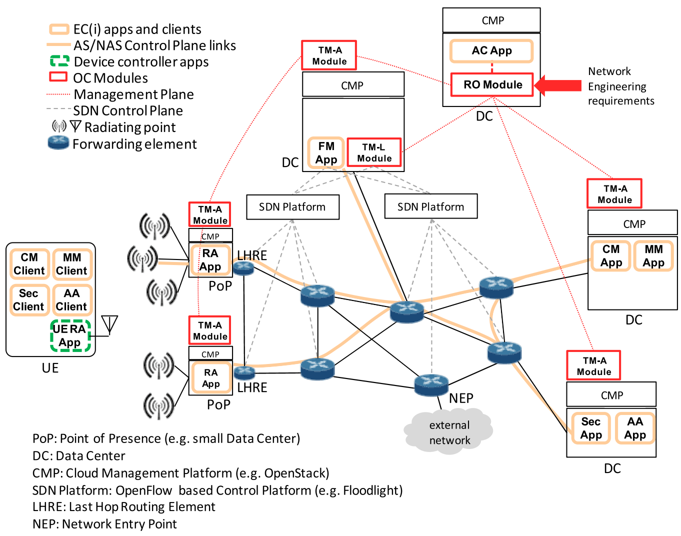

Figure 2. 5G control plane.

OC: The OC coordinates the utilisation of cloud resources. It inherits some of the 4G network management functions [14], as it is responsible for the allocation and maintenance of resources required to instantiate both 5G control and data planes. As shown in Figure 2, the OC is composed by the *resource orchestration* (RO) module and the *topology management* (TM) modules. The RO module decides how to allocate the available resources to instantiate and interconnect EC C-Apps. In other words, the RO determines the embedding solution for the virtual control and data planes to be instantiated within the cloud infrastructures. The TM module enforces RO decision: it handles the physical resources directly, and it is composed by the topology management apps (TM-A) and topology management links (TM-L), which handle virtual machines and virtual links, respectively, required to instantiate and interconnect the EC C-Apps. The RO is centralised and has visibility over the whole cloud infrastructure, while TM-A and TM-L are distributed modules interacting with SDN platforms (e.g. Floodlight), to instantiate data forwarding paths, and with cloud management platforms to instantiate the C-Apps. The instantiation of the C-Plane consists of embedding the EC(i) Apps while fulfilling network engineering and application requirements. Network engineering requirements include the required number, capacity and location of access nodes to provide coverage in the geographical area where 5G services are made available and the required backhauling capacity between access network and CN, according to the expected 5G traffic. The OC is the element guaranteeing *network efficiency* and *flexibility*: taking into account engineering requirements, tailored C-Planes can be instantiated according to application and device requirements. For example, for basic applications involving static devices and not requiring secure communication, a basic C-Plane will be instantiated, for example, not including MM, authorisation & authentication and Sec Apps. The intelligence placed in the OC is one of the key 5G features, which will guarantee efficiency in resource usage and fulfillment of 5G performance requirements, in terms of latency, reliability and scalability, as addressed in Section 2.

3.3. 5G data plane

A clean-slate design approach for the D-Plane has been followed, based on SDN technology. Neither dedicated D-Plane network elements (such as 4G SGW and PGW for instance) nor unique logical elements for the whole attached device population (such as gateways or mobility anchor points) are defined. Also, the tunneling concept and protocols are no longer required.

Simply, during network attachment, an address is allocated to the device. Additionally, a last hop routing element (LHRE) and a network entry point (NEP) is associated to it. Then, a *forwarding path* is established, to allow packets generated by (or directed to) the device to be routed from the LHRE to the NEP (or from NEP to LHRE). The establishment of a forwarding path requires the FM App to select the most suitable virtual link between LHRE and NEP, and the physical infrastructure controller to set forwarding tables on switches belonging to the SDN-based cloud infrastructure. The wireless connection between the RA point and the device is managed by the RA App of the EC(i). At service request, the required QoS is enforced over both the wireless connection and the forwarding path.

LHRE: The LHRE chains the RA point of the device to the backhaul infrastructure. It represents a *soft mobility*

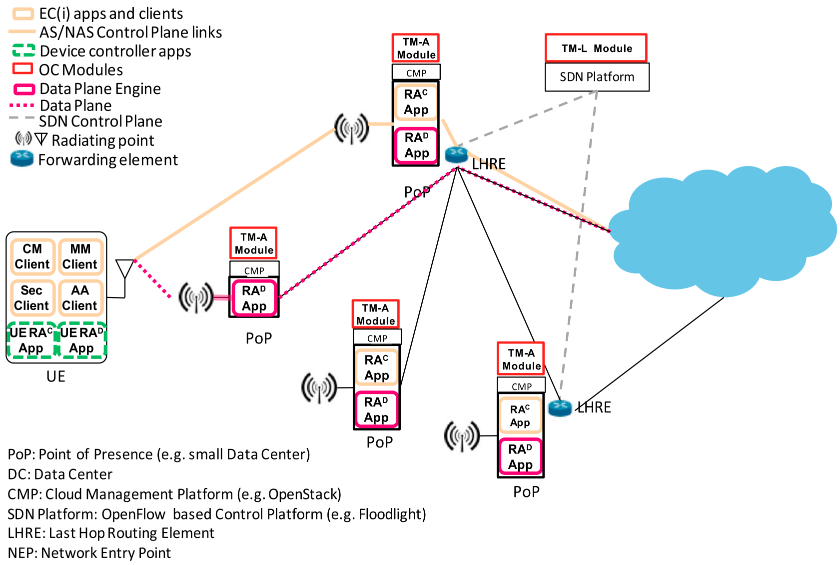

Figure 3. 5G access architecture—control and data planes.

anchoring point. When a device attaches to the network, a forwarding path from the selected LHRE to the selected NEP is established by the FM App. The forwarding path allows packets generated by (directed to) the device to be forwarded to (from) a NEP (to the device LHRE).

NEP: The NEP identifies the bound for physical infrastructure controlled by the OC. It represents a *soft PDN anchoring point*. Different attached devices may have different NEPs, depending on the attachment type and the related service, which might be requested. In some cases, the NEP might not be allocated, for example, in some V2V applications where the traffic is confined within the infrastructure owned by the service provider.

3.4. 5G access control and data plane

The separation between C-Plane and D-Plane is extended to the wireless access network. According to the dual connectivity principle, a device might be connected to two different access points simultaneously, one providing the (access) C-Plane and the other providing the (access) D-Plane. Several access network configurations might coexist in 5G systems, as shown in Figure 3. Each remote radio unit to which a device can connect is connected to a PoP, for example, a small edge data center. A PoP can implement only the RA^{D} App or both the RA^{D} and RA^{C} Apps. Finally, two PoPs might or might not share the same LHRE.

4. 5G PROCEDURES

4.1. Control plane instantiation

One of the key innovations introduced by 5G architectures is the adaptability and flexibility of the C-Plane, which can be dynamically reconfigured according to changing conditions related to network engineering requirements and service performance requirements. In our architecture, the RO module manages the allocation of resources to the EC(i) Apps. The RO has abstract knowledge of the underlying infrastructure, mediated by TM modules that handle the virtualised substrate. In multi-domain or multi-technology environment, the RO may request resources from TM modules belonging to other networks via cloud-to-cloud application programming interfaces. Each TM module handles the resource provisioning and monitoring within a single domain and single technology. As shown in Figure 4, the TM-A modules handles the cloud computing resources provisioned by cloud management platforms (e.g. OpenStack, just to mention an open source cloud management platform) and the TM-L module controls the virtual links maintained by SDN-based control platforms (e.g. Floodlight). The TM modules report resource state and availability to the RO module, which runs algorithms for embedding virtual infrastructures into cloud resources (some examples were presented in [15]).

The instantiation of the C-Plane consists of the following:

- Embedding the EC(i) Apps while fulfilling network engineering requirements (e.g. network planning, energy consumption or operational cost constraints

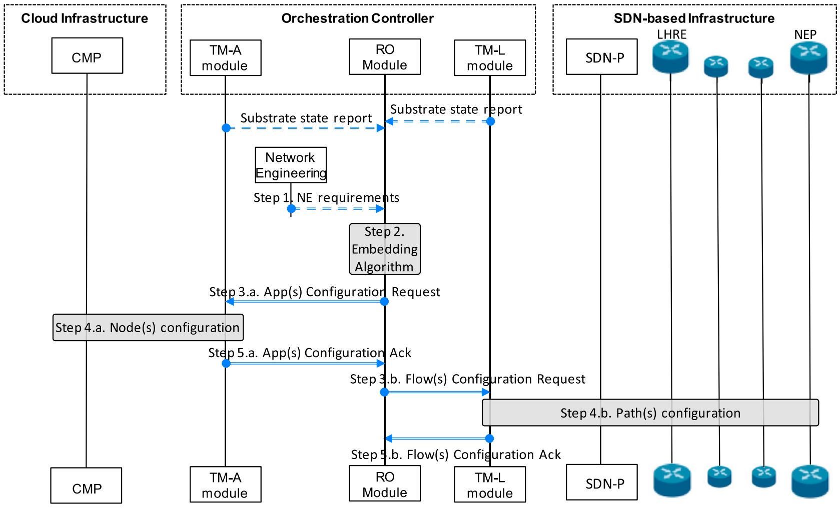

Figure 4. Control plane instantiation.

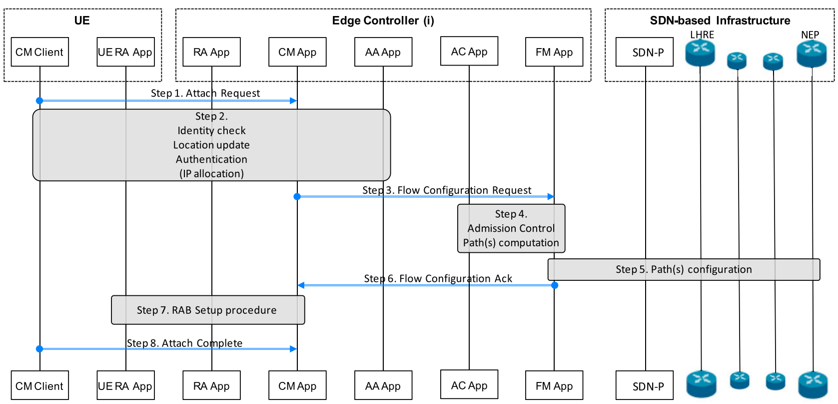

Figure 5. Initial device attachment.

and geographical distribution of the devices) and service performance requirements;

- Configuring virtual links to interconnect EC(i) control applications.

The procedure for the initial instantiation of the C-Plane is depicted in Figure 4. The TM modules update periodically the RO Module about availability and state of the substrate resources. In step 1, a network engineering requirement message triggers the embedding algorithm (step 2) in the RO module. The implementation of the NE requirements may result in instantiating (or de-instantiating) apps in the cloud infrastructure controlled by the TM-A module (steps 3.a–5.a) and new C-Plane links in the SDN-based infrastructure controlled by the SDN Platform (steps 3.b–5.b).

4.2. Initial device attachment

The 5G initial attach procedure is illustrated in Figure 5. This procedure is managed by the EC(i), assuming that an EC(ii) is not involved.

In step 1, the CM Client in the UE sends an attach request to the CM App in the EC(i). The CM client includes in the request the UE identity information as well as the reason of the attachment; for instance, to get assigned an IP address. The request triggers step 2 involving UE identity check, update of the UE location, authentication and, if requested, the allocation of an IP address. Note that the usage of IP is only one out of several alternatives for the addressing mechanism that could be employed in 5G systems. The implementation of these procedures may significantly affect the time elapsed to complete this phase. At the completion of step 2, the CM App requests the FM App to define and implement the default rules to handle

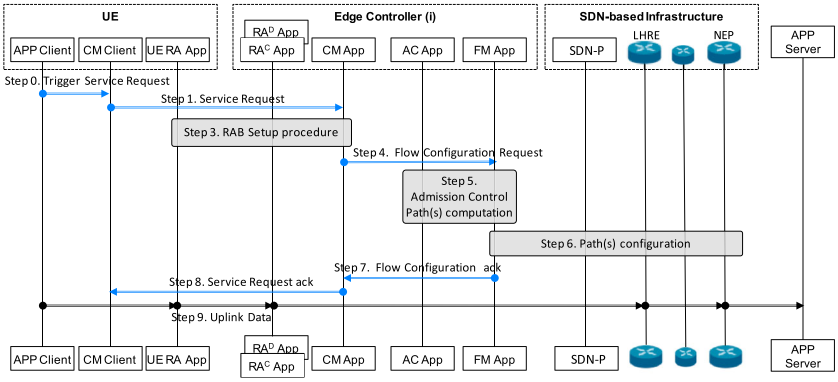

Figure 6. Device triggered service request.

traffic directed to or generated by the UE. This flow configuration request is handled by AC App and FM App in step 4. Depending on the type of attachment, the FM App determines the high-level rules for network configuration. For instance,

- Packets directed to the UE shall be routed to the LHRE connected to the access point that serves the UE;
- UE generated packets directed to an external device (not managed by FM App) shall be routed to FM App NEPs;
- UE generated packets directed to an internal device shall be routed to destination.

The rules are checked by the AC App, which determines how to realise the requested path based on the knowledge of the physical infrastructure topology and utilisation.

In step 5, the FM App deploys the rules in the SDN-based platform using the appropriate policy defined for the attachment type. For peculiar attachment types related to low latency services, a proactive deployment of the forwarding rules shall guarantee the required performance. If latency requirements are relaxed or undefined, the forwarding rules can be reactively dispatched by the SDN controller whenever an SDN switch needs to resolve the forwarding of the first packet of a data flow belonging to the UE.

The CM App, after receiving the acknowledgment to the flow configuration request (step 6), triggers the radio access bearer (RAB) setup procedure (step 7). The UE CM client completes the procedure by sending the attach complete message.

4.3. Device triggered service request

Figure 6 describes the procedure for a device triggered service request.

Triggered by an application client, the CM client in the UE sends a service request to the CM App including the identity information of the UE and, optionally, of the application. The CM App initiates the RAB establishment (step 3) and the flow configuration in the SDN-based infrastructure (steps 4–7). The RAB may involve different RA Apps from the one that received the service request; for instance, C-Plane and D-Plane could be managed by different RA Apps. The steps 3, 5 and 6 are similar to those introduced for the initial attach procedure, with the sole difference that radio bearer and forwarding paths are related to a dedicated bearer, which QoS requirements may depend on the device capabilities, on the specific service instance or on the application requesting the service. When the establishment of radio bearer and forwarding paths is completed, the CM app acknowledges the service request to the UE (step 8). Finally, in step 9, the application client can start sending data to the App server.

4.4. Mobility management

The proposed architecture simplifies the current MM procedures. The device is initially camped on a serving (S) cell controlled by the S-RA^C App with data connection to S-RA^D App. The simplest mobility case occurs when a handover is required for the D-plane only, without the need of mobility anchor relocation, that is, when the PoP where the target (T) RA ^{D} App is instantiated is connected to the same LHRE as the PoP where the S-RA^D App is located. The procedure, shown in Figure 7, starts with the HO preparation phase (step 1): upon radio link quality measurements, the S-RA^{C}App triggers a 'RA^D App only HO'. The T-RA^{D}App is selected by the S-RA^C App, then the S-RA^C App requires the HO to the T-RA^D App. T-RA^D App reserves required resources for the HO and acknowledges the HO request to the S-RA^C App. The HO execution phase follows (step 2): the S-RA^C App executes the HO over the radio interface and notifies the HO completion to the T-RA^D App. Uplink data are now sent from the device to the T-RA^D App and from T-RA^{D}App to LHRE. Downlink data are sent from LHRE to S-RA^D App, forwarded from S-RA^{D}App to T-RA^D App and then transmitted over radio interface. Finally, forwarding path switch takes place on the last hop, as indicated in step 3. In this phase the

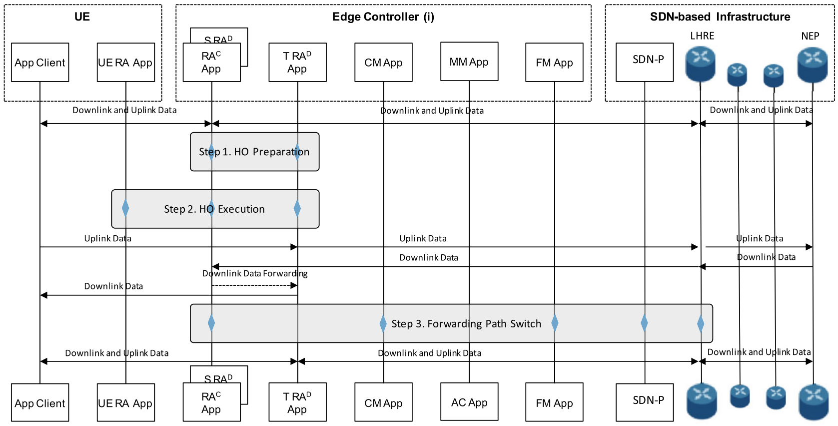

Figure 7. Handover without anchor relocation.

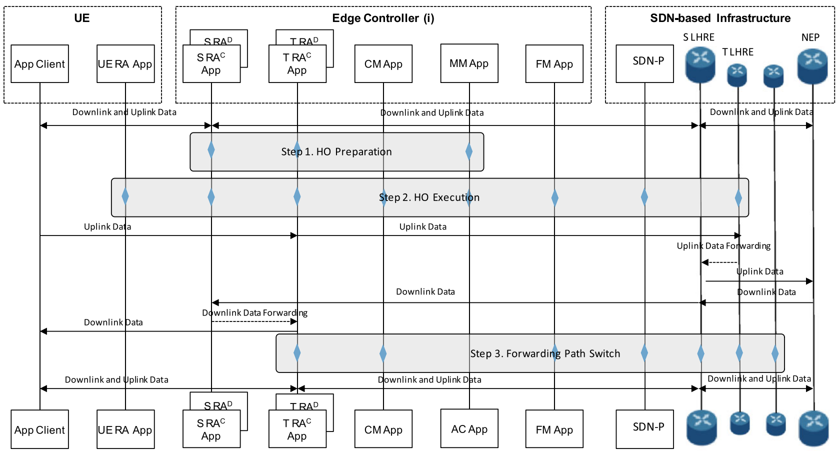

Figure 8. Handover with anchor relocation.

S-RA^C App notifies the HO to the CM App, which verifies the need for a forwarding path switch. Then, the CM App requires a forwarding path switch to the FM App, needed for the LHRE to forward data to the PoP where the T-RA^D App is instantiated. The FM App finds a new path for data to be forwarded from LHRE to T-RA^D App, and finally, the SDN platform enforces the new forwarding path, configuring forwarding tables in LHRE.

The handover procedure in case of mobility anchor relocation (i.e. LHRE change) is depicted in Figure 8. The device is initially camped on a serving (S) cell controlled by the S-RA^C App with data connection to S-RA^D App. The handover preparation phase (step 1) is initiated by S-RA^C App when radio link quality measurements indicate a better service cell exists. During the preparation phase, the target (T) cell and corresponding T-RA^D App and T-RA^C App are identified. The required resources are then reserved in the target cell. Then, the execution phase takes place (step 2). The S-RA^C commands the UE to camp on the cell controlled by the T-RA^C App. Radio connection is established to T-RA^C App for C-Plane and to T-RA^D App for D-Plane. The MM App updates information on UE location, the CM App selects the LHRE associated to T-RA^C and T-RA^D Apps (T-LHRE) and the FM App configures temporary uplink data forwarding from T-LHRE to S-LHRE. The handover procedure is completed with the forwarding path switching phase (step 3), where the CM and FM App trigger the reconfiguration of SDN-based cloud infrastructure to allow packets generated by the UE and forwarded to the T-LHRE to reach their destination, as well as to allow packets directed to the UE to be properly forwarded to the T-LHRE. The SDN-based cloud infrastructure is configured by the SDN platform controller upon trigger by FM App. Finally, it should be noted that in case IP is the addressing mechanism used, 'Mobile IP' is not required, as the IP address and the NEP associated to a device will not change when the access node changes because of device mobility.

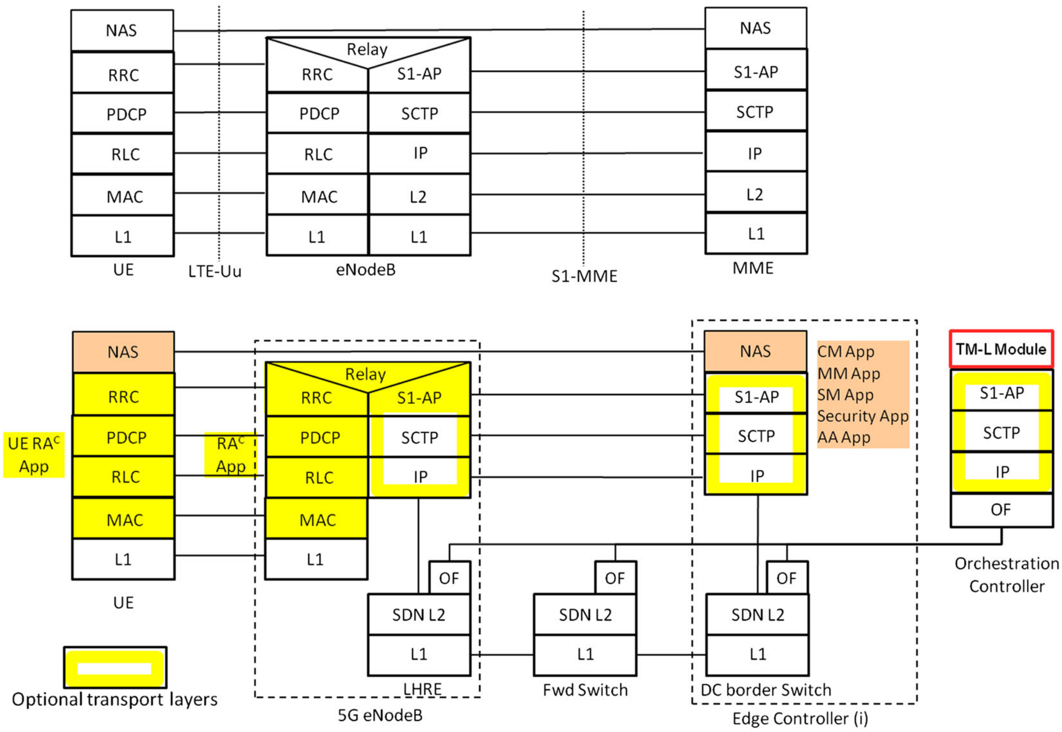

Figure 9. 4G control plane instantiation within 5G architecture.

5. BACKWARD COMPATIBILITY TO 4G NETWORKS

This section describes an instantiation of 4G control and user plane protocol stacks in our SDN-based 5G architecture. In Figure 9, LTE NAS is implemented by EC(i) Apps connected via forwarding paths, instead of manually configured C-Plane paths. The protocol stacks specified in [14] for the LTE-Uu and S1-MME interfaces are reported as reference in the upper part of the figure. At the bottom, the S1-MME interface uses an SDN-based transport network layer. As shown in Figure 2, the links between the EC(i) Apps are configured and maintained by the TM-L module in the OC. The TM-L deploys the forwarding paths between the LHRE and the data center hosting the CM App, as well as the connectivity between the other EC(i) Apps. For instance, in Figure 9, the forwarding elements are configured by the TM-L module by means of OpenFlow controllers. The S1 Application Part may either run on top of an Stream Control Transmission Protocol layer or dispatch the C-Plane messages directly to the SDN layer. In the former case, the SDN layer forwards IP messages and, in the latter case, S1-AP messages. This implementation provides a high degree of flexibility to the C-Plane, which can be automatically reconfigured to cope with evolving traffic engineering requirements or to react to possible network failures. Another innovation enabled by this architecture is that the NAS messages could be potentially forwarded to EC(i) Apps located in other networks: UEs camped on a eNodeB may request a service to an app unknown to the LHRE connected to the eNodeB. In this case, the TM-L module must reactively configure the C-Plane path: the LHRE forwards the S1-AP message to the TM-L module, which resolves the destination of the message and configures the C-Plane path accordingly.

Another possible innovative approach, not depicted in the figures, relates to splitting the RA^C App in a Radio Link Control (RLC) App and a Radio Resource Control (RRC) App. Medium Access Control and RLC are normally co-located with the radiating point for performance reasons, while Packet Data Convergence Protocol and RRC may be embedded in edge computing PoP or in centralised data centers. Also in this case, the paths between the RLC unit and the RRC unit of the eNodeB may be implemented through forwarding paths configured by the TM-L module.

Figure 10 shows an implementation of the LTE User Plane that replaces GPRS Tunneling Protocol (GTP) tunnels with SDN forwarding paths. The eNodeB receives the application layer messages originated by the UE and forwards them to the LHRE. The LHRE is an SDN forwarding element. If the forwarding path to reach the application server was already pre-configured by the FM App, the LHRE forwards the IP message to the next forwarding element. Alternatively, the IP messages that cannot be resolved by the LHRE are forwarded to the FM App. In this case, the FM App resolves the location of the application server and reactively configures the appropriate forwarding path.

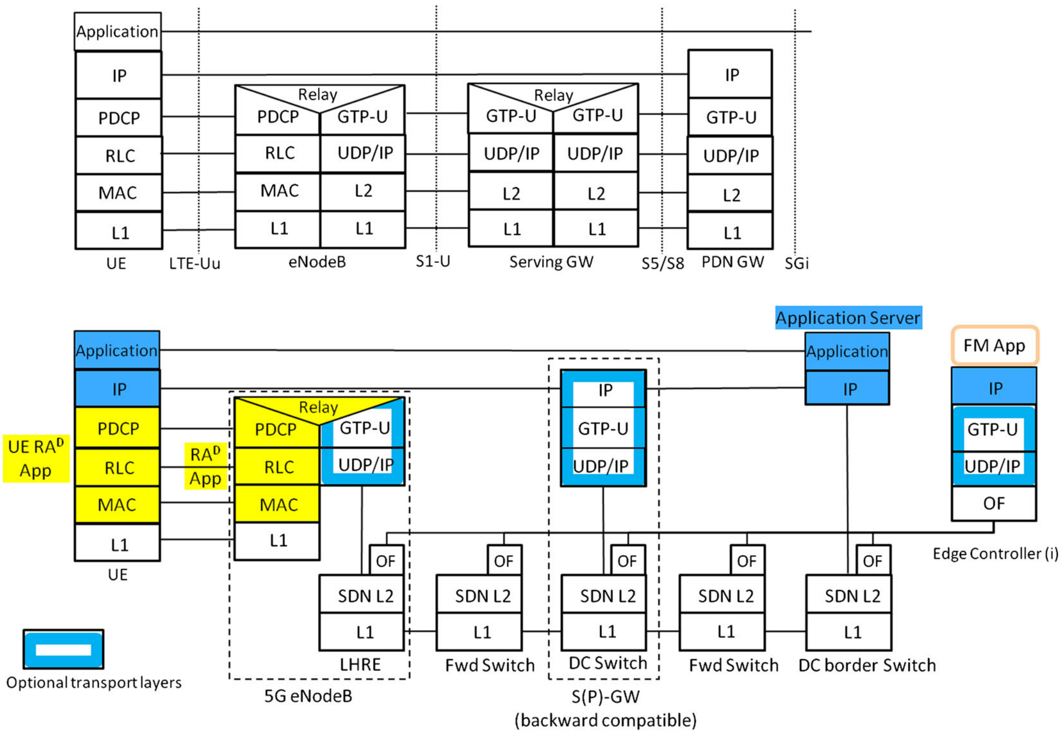

Figure 10. 4G data plane instantiation within 5G architecture.

For backward compatibility with legacy network elements, GTP tunnels could be established on top of the SDN layer. This solution can be necessary to integrate LTE eNodeB with virtualised Serving Gateways (GWs) and PDN-GWs or, viceversa, to connect 5G eNodeBs with legacy S(P)-GWs.

6. CONCLUSIONS

In this paper, we proposed a novel *plastic architecture* for the advanced 5G network infrastructure by harvesting latest advances of SDN, network functions virtualisation and edge computing.

The proposed architecture, functions and procedures have the potential to become the 'de facto' solution for 5G, which is expected to be the 'Nervous System' of the Digital Society and Digital Economy ([13]). Extremely low latency, ultra-high reliability and scalability are the most stringent performance targets that need to be reached to realise this vision. Beyond this, flexibility and backward compatibility are fundamental features of the novel SDN-based 5G architecture, which has thus the potential to be standardised by relevant organisations as the next generation infrastructure, supporting compelling services and enabling a plethora of new sustainable business models.

Future developments of the presented work, ongoing at the time of writing this paper, include case studies showing how the proposed architecture applies to key 5G services such as IoT and V2V, and a proof of concept aiming at an empirical demonstration that the proposed architecture can achieve the expected performance targets.

Ultimately, we would like to state clearly that the views expressed herein are solely those of the authors and do not necessarily represent the ones of their affiliate.

REFERENCES

1. METIS. Scenario, requirements and KPIs for 5G mobile and wirelss system. Available from: https://www.metis2020.com/documents/deliverables/ [20 September 2014].
2. Osseiran A, Boccardi F, Braun V, *et al.* Scenarios for 5G mobile and wireless communications: the vision of the METIS project. *IEEE Communications Magazine* 2014; **52**(5): 26, 35.
3. ITU towards IMT for 2020 and beyond.
4. Font-Bach O, Bartzoudis N, Pascual-Iserte A, *et al.* Interference management in LTE-based HetNets: a practical approach. *Transactions on Emerging Telecommunications Technologies* 2014. DOI: 10.1002/ett.2833
5. Cai Y, Yu FR, Bu S. A decision theoretic approach for clustering and rate allocation in coordinated multi-point (CoMP) networks with delayed channel state information. *Transactions on Emerging Telecommunications Technologies* 2014. DOI: 10.1002/ett.2831

6. Xu X, Wang H, Feng H, Xing C. Analysis of device-to device communications with exclusion regions underlaying 5G networks. *Transactions on Emerging Telecommunications Technologies* 2014. DOI: 10.1002/ett.2854
7. ETSI MCC Department. Standardization of Machine-type Communications, V. 0.2.4 (2014-06), June 2014.
8. ETSI MCC Department. 3GPP Multimedia Broadcast and Multicast Service (MBMS), V. 0.0.2 (2013-04), April 2013.
9. Bernardos CJ, De La Oliva A, Serrano P, *et al.* An architecture for software defined wireless networking. *IEEE Wireless Communications* 2014; **21**(3):52–61.
10. Agyapong PK, Iwamura M, Staehle D, Kiess W, Benjebbour A. Design considerations for a 5G network architecture. *IEEE Communications Magazine* November 2014: 65–75.
11. Yazici V, Kozat UC, Sunay MO. A new control plane for 5G network architecture with a case study on unified handoff, mobility, and routing management. *IEEE Communications Magazine* November 2014: 76–85.
12. European Commission. Advanced 5G Network Infrastructure for the Future Internet – Public Private Partnership in Horizon 2020, H2020 LEIT, Work Programme, ICT 14, 2013.
13. Soldani D, Manzalini A. A 5G Infrastructure for Anything-as-a-Service. *Journal of Telecommunications System & Management* 2014, DOI:10.4172/2167-0919.1000114, in press.
14. 3GPP TS 23.401. General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access.
15. Guerzoni R, Trivisonno R, Vaishnavi I, *et al.* A novel approach to virtual networks embedding for SDN management and orchestration. In *IEEE/IFIP Network Operations and Management Symposium (NOMS)*, Krakow, May 2014.