

SFC WG
Internet-Draft
Intended status: Informational
Expires: April 25, 2022

CJ. Bernardos
UC3M
A. Mourad
InterDigital
October 22, 2021

Service Function Chaining Use Cases in Fog RAN
draft-bernardos-sfc-fog-ran-10

Abstract

Fog Radio Access Networks (RAN) refers to the part of the RAN that is virtualized at the very edge of the network, even at the end-user device. Fog RAN support is considered critical for the 5G mobile network architectures currently being developed in various research, standardization and industry forums. Since fog RAN builds on top of virtualization and can involve several virtual functions running on different virtualized resources, Service function chaining (SFC) support for the fog RAN will be critical. This document describes the overall fog RAN approach and also gives some use cases. Finally it proposes some requirements to be considered in the development of the SFC architecture and related protocols.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 25, 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents

(<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Terminology	3
3.	Fog RAN Overview	4
4.	Applicability of SFC to Fog RAN	8
5.	Fog RAN requirements	11
6.	IANA Considerations	12
7.	Security Considerations	13
8.	Acknowledgments	13
9.	References	13
9.1.	Normative References	13
9.2.	Informative References	13
Appendix A.	4G (LTE)	14
Appendix B.	5G	16
	Authors' Addresses	16

[1.](#) Introduction

The telecommunications sector is experiencing a major revolution that will shape the way networks and services are designed and deployed for the next decade. We are witnessing an explosion in the number of applications and services demanded by users, which are now really capable of accessing them on the move. In order to cope with such a demand, some network operators are looking at the cloud computing paradigm, which enables a potential reduction of the overall costs by outsourcing communication services from specific hardware in the operator's core to server farms scattered in data centers. These services have different characteristics if compared with conventional IT services that have to be taken into account in this cloudification process. Also the transport network is affected in that it is evolving to a more sophisticated form of IP architecture with trends like separation of control and data plane traffic, and more fine-grained forwarding of packets (beyond looking at the destination IP address) in the network to fulfill new business and service goals.

Virtualization of functions also provides operators with tools to deploy new services much faster, as compared to the traditional use of monolithic and tightly integrated dedicated machinery. As a natural next step, mobile network operators need to re-think how to

evolve their existing network infrastructures and how to deploy new ones to address the challenges posed by the increasing customers' demands, as well as by the huge competition among operators. All these changes are triggering the need for a modification in the way operators and infrastructure providers operate their networks, as they need to significantly reduce the costs incurred in deploying a new service and operating it. Some of the mechanisms that are being considered and already adopted by operators include: sharing of network infrastructure to reduce costs, virtualization of core servers running in data centers as a way of supporting their load-aware elastic dimensioning, and dynamic energy policies to reduce the monthly electricity bill. However, this has proved to be tough to put in practice, and not enough. Indeed, it is not easy to deploy new mechanisms in a running operational network due to the high dependency on proprietary (and sometime obscure) protocols and interfaces, which are complex to manage and often require configuring multiple devices in a decentralized way.

Network function virtualization (NFV) [[etsi nvf whitepaper](#)] and software defined networking (SDN) [[onf sdn architecture](#)] are changing the way the telecommunications sector will deploy, extend and operate their networks. In this document we focus on one particular application of network softwarization: the radio access network (RAN), and in particular, how to run RAN functions on dynamic virtual resources very close to the users, the so-called Fog. We analyze the applicability of the SFC architecture to the Fog RAN use case.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

While [[RFC2119](#)] describes interpretations of these key words in terms of protocol specifications and implementations, they are used in this document to describe requirements for the SFC mechanisms to efficiently enable fog RAN.

The following terms used in this document are defined by the ETSI NFV ISG, the ONF and the IETF:

NFV Infrastructure (NFVI): totality of all hardware and software components which build up the environment in which VNFs are deployed

NFV Management and Orchestration (NFV-MANO): functions collectively provided by NFVO, VNFM, and VIM.

NFV Orchestrator (NFVO): functional block that manages the Network Service (NS) lifecycle and coordinates the management of NS lifecycle, VNF lifecycle (supported by the VNFM) and NFVI resources (supported by the VIM) to ensure an optimized allocation of the necessary resources and connectivity.

OpenFlow protocol (OFP): allowing vendor independent programming of control functions in network nodes.

Service Function (SF): a function that is responsible for specific treatment of received packets (e.g., firewall, load balancer).

Service Function Chain (SFC): for a given service, the abstracted view of the required service functions and the order in which they are to be applied. This is somehow equivalent to the Network Function Forwarding Graph (NF-FG) at ETSI.

Service Function Path (SFP): the selection of specific service function instances on specific network nodes to form a service graph through which an SFC is instantiated.

Virtualized Infrastructure Manager (VIM): functional block that is responsible for controlling and managing the NFVI compute, storage and network resources, usually within one operator's Infrastructure Domain.

Virtualized Network Function (VNF): implementation of a Network Function that can be deployed on a Network Function Virtualisation Infrastructure (NFVI).

Virtualized Network Function Manager (VNFM): functional block that is responsible for the lifecycle management of VNF.

3. Fog RAN Overview

Virtualization is invading all domains of the E2E 5G network, including the access, as a mean to achieve the necessary flexibility in support of the E2E slicing concept. The ETSI NFV framework is the cornerstone for making virtualization such a promising technology that can be matured in time for 5G. Typically, virtualization has been mostly envisaged in the core network, where sophisticated data centers and clouds provided the right substrate. And mostly, the framework focused on virtualizing network functions, so called VNFs (virtualized network functions), which were somewhat limited to functions that are delay tolerant, typically from the core and aggregation transport. Yet in the access, virtualization of RAN functions could still be envisaged as one step forward for the Cloud-RAN concept, assuming an extremely low latency fronthaul transport

(typically over fiber) and again limiting to those RAN functions which delay requirements match for execution in a virtualized environment.

As the community has recently been developing the 5G applications and their technical requirements, it has become clear that certain applications would require very low latency which is extremely challenging and stressing for the network to deliver through a pure centralized architecture. The need to provide networking, computing, and storage capabilities closer to the users has therefore emerged, leading to what is known today as the concept of intelligent edge. ETSI has been the first to address this need recently by developing the framework of mobile edge computing (MEC).

Such an intelligent edge could not be envisaged without virtualization. Beyond applications, it raises a clear opportunity for networking functions to execute at the edge benefiting from inherent low latencies. Being in close proximity to the access, the edge becomes thus an attractive place for hosting C-RAN functions in particular, which could also be envisaged virtualized thanks to the virtualization capabilities now available at the edge. Transport and core networking functions could also take advantage of such a hosting environment, thus saving bandwidth in their respective domains and offering local breakout options where required. Furthermore, a rich set of context information from the RAN but also from other network domains could be offered as services through the edge for applications to consume and hence optimize their behavior or key performance indicators (KPIs).

Whilst it is appreciated the particular challenge for the intelligent edge concept in dealing with mobile users, the edge virtualization substrate has been largely assumed to be fixed or stationary. Although little developed, the intelligent edge concept is being extended further to scenarios where for example the edge computing substrate is on the move, e.g., on-board a car or a train, or that it is distributed further down the edge, even integrating resources from different stakeholders, into what is known as the fog. The challenges and opportunities for such extensions of the intelligent edge remain an exciting area of future research.

The computing and virtualization capabilities available down into the fog are of particular advantage to the virtualized-RAN/cloud-RAN, leading to what we refer to as "Fog RAN". Virtual networking functions (VNFs) related to the RAN may execute in the fog. The close proximity of the fog to the end user devices opens new possibilities for extending the scope of virtual RAN (VRAN) functions from just access nodes so far to end user devices, so that functions traditionally running on the end user devices may be moved to the

fog. In addition, an additional tier of intelligent VRAN functions could be envisaged to run across other functions of various nodes or devices and from different radio access technologies (RATs), supporting tight coordination between these nodes and devices, and enabling convergence between the RATs. VNFs from the transport and core could also be hosted in the fog so as to save bandwidth in their respective domains.

Figure 1 shows a diagram representing the fog RAN concept. The fog is composed by virtual resources on top of heterogeneous resources available at the edge and even further in the RAN and end-user devices. These resources are therefore owned by different stakeholders who collaboratively form a single hosting environment for the VNFs to run. As an example, virtual resources provided to the fog might be running on eNBs, APs, at micro data centers deployed in shopping malls, cars, trains, etc. The fog is connected to data centers deeper into the network architecture (at the edge or the core). On the top part of the figure, an example of user and control plane VNFs is shown. User plane VNFs are represented as "fx", and control ones as "ctrlx". Depending on the functionality implemented by these VNFs and the service requirements, these VNFs would be mapped (i.e., instantiated) differently to the physical resources (as described in [[I-D.aranda-sfc-dp-mobile](#)]).

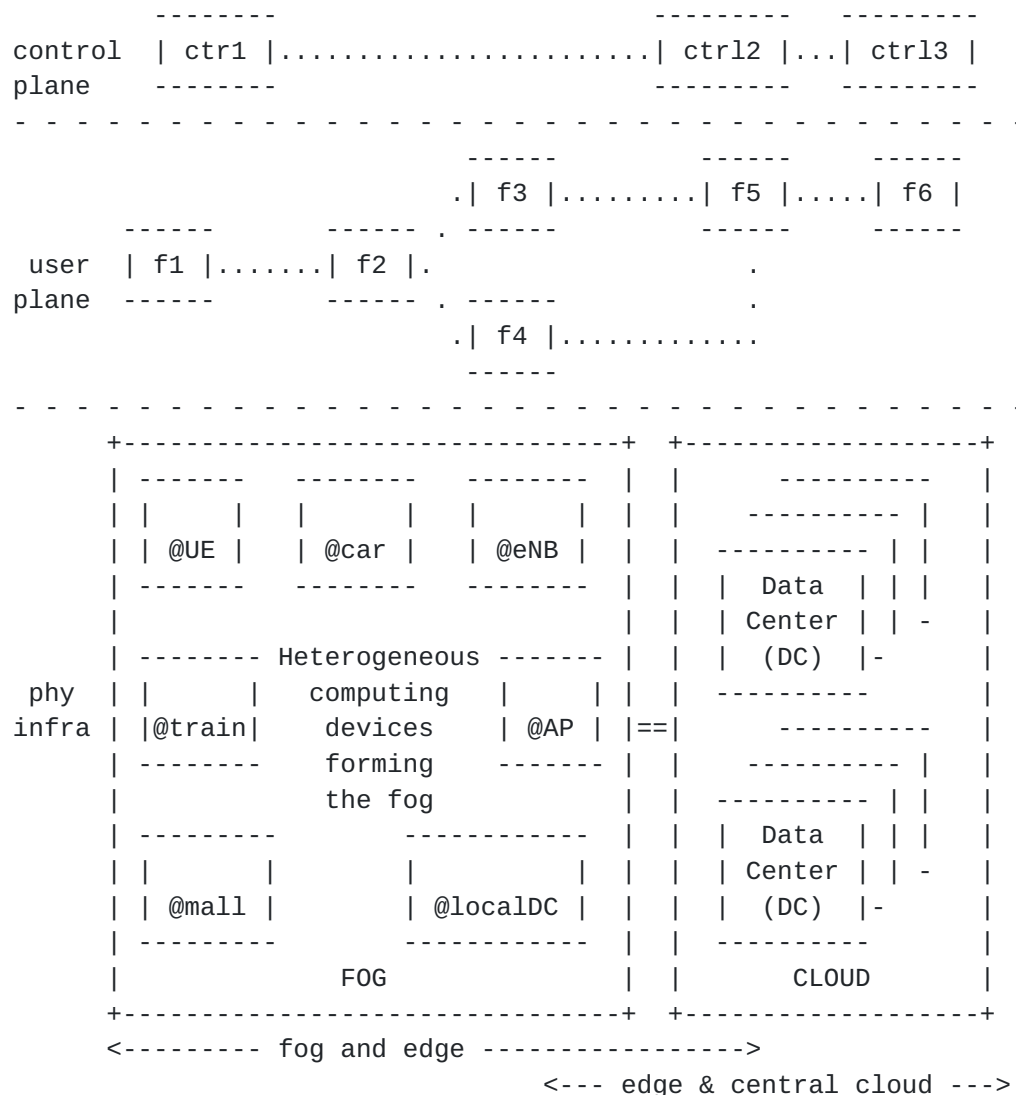


Figure 1: Fog RAN

The fog is also well suited to offer a rich set of context information noticeably (but not only) from the RAN that could be collected from the various RATs co-existing in the same service area. This information can be obtained either from networking resources (e.g., nodes and devices) or functions, some of which might be hosted in the fog. Such information may be exposed through services running in the fog or in the edge, for applications on top to consume and hence optimize their performance or behavior. The fog or edge will therefore be in charge of collecting and publishing context information towards network applications as well as end user/3rd party applications. This is accomplished by the so-called "services" which follow the concept of ETSI MEC ISG services and may run in the fog. Applications may also run directly in the fog and subscribe to one or more services provided in the fog. These fog applications

could also offer services to other applications residing inside the fog or outside, e.g., in the edge or central cloud. This enables different kinds of Over-The-Top (OTT) applications to utilize the RAN context information available at the fog.

4. Applicability of SFC to Fog RAN

For a given service, the abstracted view of the required service functions and the order in which they are to be applied is called a service function chain (SFC), which is called network function forwarding graph (NF-FG) in ETSI. An SFC is instantiated through selection of specific service function instances on specific network nodes to form a service graph: this is called a service function path (SFP). The service functions may be applied at any layer within the network protocol stack (network layer, transport layer, application layer, etc.).

[RFC7665] describes the architecture for the specification, creation, and ongoing maintenance of service function chains (SFCs) in a network. The SFC architecture is composed of the following logical components: classifiers, service function forwarders (SFFs), service functions (SFs), and SFC proxies. These components are interconnected using the SFC encapsulation, as shown in Figure 2.

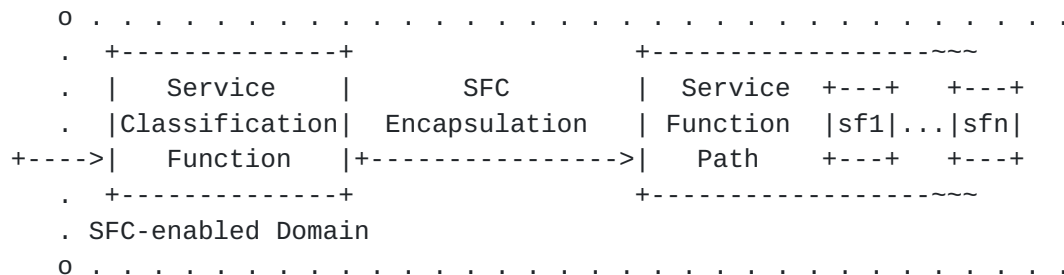


Figure 2: SFC architecture

An overview of how the SFC logical components interact after the initial classification is also shown in Figure 3 (reproduced from [RFC7665] for completeness).

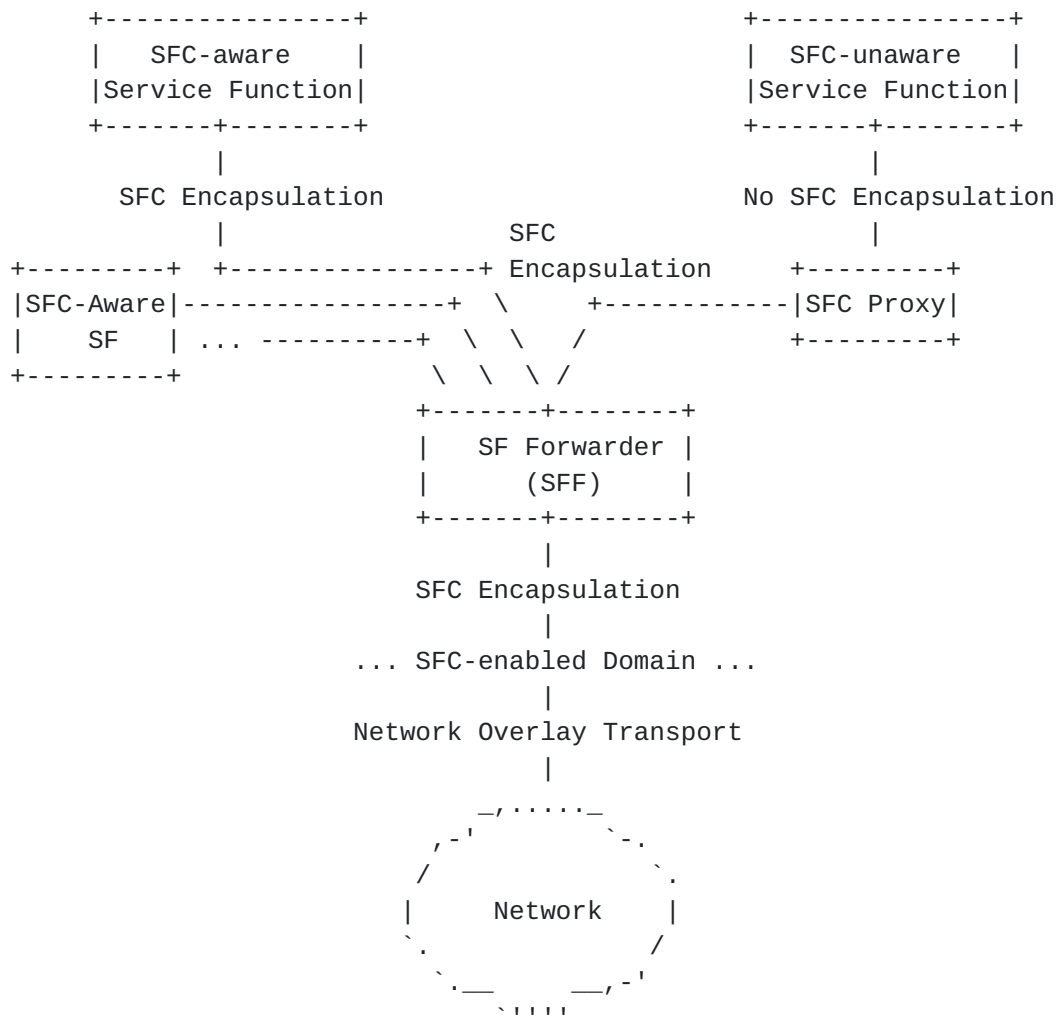


Figure 3: SFC architecture components

There are different use cases for service function chaining in mobile networks. [\[I-D.ietf-sfc-use-case-mobility\]](#) describes in general how to use SFC for mobile networks focusing on the core functions, while [\[I-D.aranda-sfc-dp-mobile\]](#) looks more at RAN aspects.

This document focuses on the specific use case of applying SFC concepts to the fog RAN environment, which is characterized mainly by:

- o The VNFs being chained implement mostly RAN functionality. The Cloud RAN (C-RAN) approach centralizes baseband processing and network resource allocation (which are today functions executed in a distributed way at the access nodes, such as eNBs). The strict latency and bandwidth requirements imposed by C-RAN triggered the evolution of the virtualization of the access infrastructure to the concept of RAN as a Service, where the centralization (i.e.

function virtualization in a data center) of processing and management in mobile networks is flexible and adapted to the actual service requirements and the network conditions. This is also referred to as flexible functional split of the radio protocol stack. Up to now, RAN virtualization approaches have only considered offloading of computation tasks in central clouds or regional data centers close to the network aggregation points. With fog RAN, virtualization resources are placed closer to users, from edge computing to fog computing at user devices, leveraging micro data centers at user premises, thus reducing end-to-end latency.

- o Virtualizing RAN functions at the fog could also enable new optimizations when information of (or available at) the access is used. Examples of this information include: RAN measured parameters, location information, etc. This information can be used for example to jointly optimize multiple RATs.
- o The fog computing environment can also be used to virtualize functions from the end-user devices, not only from the RAN. This can help facilitating a better convergence of multiple access technologies.

Fog RAN implementations will benefit from applying SFC concepts as virtual RAN functions will be executed on virtual resources from different stakeholders, meaning different data centers managed by different entities. In this environment, SFC encapsulation can be used to ensure proper data processing. Figure 4 shows an example of scenario of application of SFC in the fog RAN. On the top part we show a UE connected to the network. The eNB functionality is actually split, so part of it (e.g., RLC, PDCP and RRC) runs virtualized in the fog, while the rest (e.g., MAC and PHY) stay at the remote radio head (RRH). Other mobile network RAN functionality can also be running virtualized in the fog, such as a local virtual MME, multi-RAT authentication and RAT selection mechanisms, offloading solutions (such as local vEPCs), etc. The rest of the mobile network functions (see [Appendix A](#) for a short overview) runs in the core. On the bottom part of the figure we show a potential mapping of the virtual functions to the physical resources that compose the fog. SFC mechanisms would be used to interconnect the different functions in the right order and meeting the requirements (latency, compute, etc) needed.

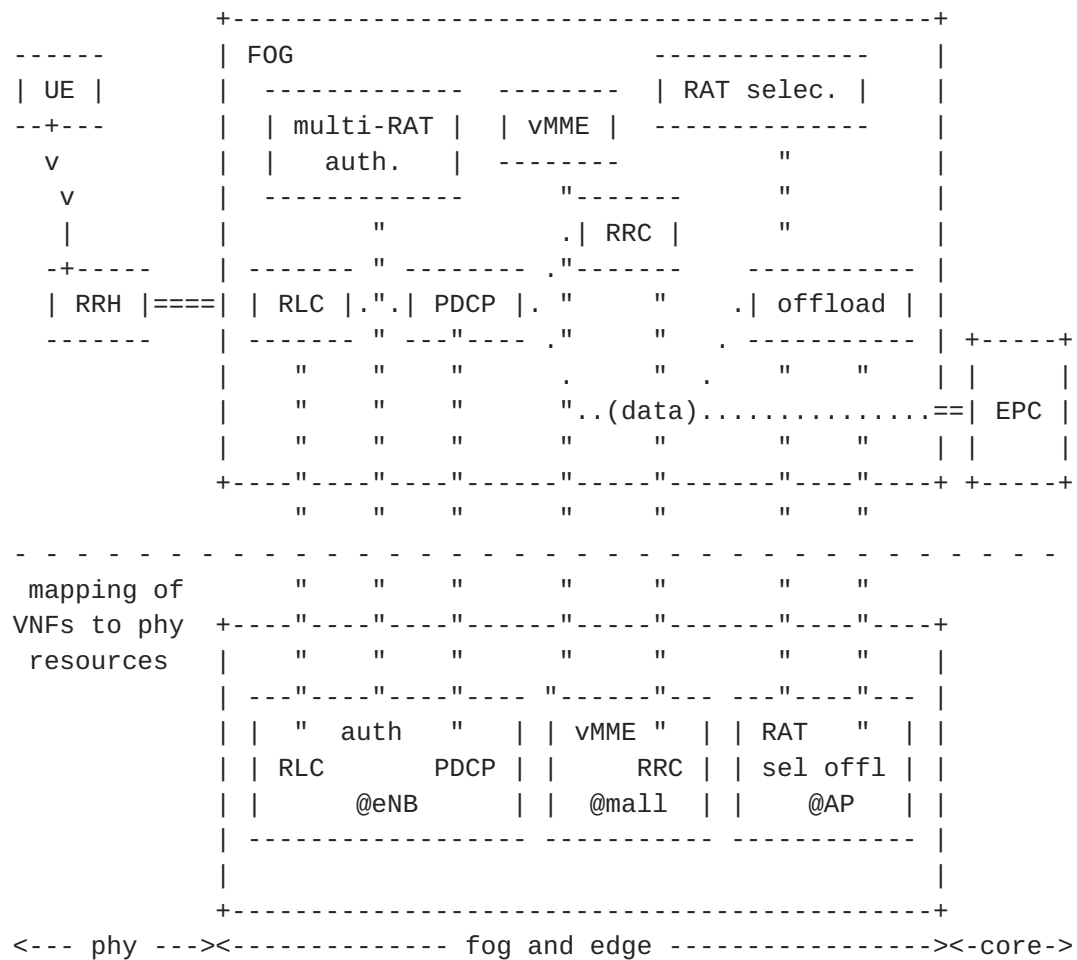


Figure 4: Fog RAN SFC example scenario

Since the virtual resources where virtual functions are potentially hosted on infrastructure managed by different entities, SFC encapsulation is a key tool to enable the fog RAN scenario. In the next section we enumerate some initial requirements to be considered, that will be extended and further developed in subsequent versions of this document.

5. Fog RAN requirements

The following requirements should be considered during SFC architecture and related protocol design to ensure that SFC can fully support fog RAN functionality:

REQ-R01: SFC MUST support user traffic flows with very low delay budgets (e.g., less than 1ms) at the edge of the network.

REQ-R02: SFC MUST support mobility of Service Functions (e.g., edge network node containing SF that is located on a high speed train).

REQ-R03: SFC MUST support Service Functions (e.g., MEC) that are located at the edge of the network and that perform L7-Application processing very early in the forwarding path.

REQ-R04: SFC MUST support metadata used to exchange information about the network and virtual resources status, so it can be used to decide about updates of the service function path.

REQ-R05: SFC MUST support metadata with context information (e.g., location, RAN information, etc.) from fog nodes (e.g., access nodes and end user devices) from multiple RATs.

REQ-R06: SFC MUST support chaining functions hosted at heterogeneous hosts (in terms of computing resources, availability, mobility, and other policies).

REQ-R07: SFC MUST support chaining functions hosted across various geographically spread physical networking and computing resources.

REQ-R08: SFC MUST support a secure and dynamic discovery of functions, potentially belonging to different domains.

REQ-R09: SFC MUST support dynamic mobility of the hosts where the service functions run, e.g., when the host is a moving device (user terminal, car, train, etc.).

REQ-R10: SFC OAM MUST support monitoring of SFs, SFFs and SFC proxies deployed on mobile platforms.

REQ-R11: SFC MUST support configuration mechanisms for mobile devices to be part of an SFC.

REQ-R12: SFC MUST support volatile resources (SFs), allowing for backup mechanisms in case a SF becomes unavailable due to the volatility of the resource.

6. IANA Considerations

N/A.

7. Security Considerations

TBD.

8. Acknowledgments

Authors would like to thank Akbar Rahman for his work on previous versions of this document.

The work in this draft has been explored under the framework of the H2020 5G-DIVE project (Grant 859881).

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.

9.2. Informative References

[etsi_nvf_whitepaper] ISG, E. N., "Network Functions Virtualisation (NFV). White Paper 2", October 2014.

[I-D.aranda-sfc-dp-mobile] Gutierrez, P. A. A., Gramaglia, M., Lopez, D. R., and W. Haeffner, "Service Function Chaining Dataplane Elements in Mobile Networks", [draft-aranda-sfc-dp-mobile-04](#) (work in progress), June 2017.

[I-D.ietf-sfc-use-case-mobility] Haeffner, W., Napper, J., Stiemerling, M., Lopez, D. R., and J. Uttaro, "Service Function Chaining Use Cases in Mobile Networks", [draft-ietf-sfc-use-case-mobility-09](#) (work in progress), January 2019.

[onf_sdn_architecture] (ONF), O. N. F., "SDN Architecture (Issue 1.1), ONF TR-521", February 2016.

[RFC7665] Halpern, J., Ed. and C. Pignataro, Ed., "Service Function Chaining (SFC) Architecture", [RFC 7665](#), DOI 10.17487/RFC7665, October 2015, <<https://www.rfc-editor.org/info/rfc7665>>.

[Appendix A](#). 4G (LTE)

This annex includes a high level summary of the 3GPP EPS architecture, commonly referred to as 4G (LTE), detailing both the EPC (core) and the RAN (access) parts.

The EPS architecture and some of its standardized interfaces are depicted in Figure 5. The EPS provides IP connectivity to user equipment (UE) (i.e., mobile nodes) and access to operator services, such as global Internet access and voice communications. The EPS comprises the core network -- called Evolved Packet Core (EPC) -- and different radio access networks: the 3GPP Access Network (AN), the Untrusted non-3GPP AN and the Trusted non-3GPP AN. There are different types of 3GPP ANs, with the evolved UMTS Terrestrial Radio Access Network (E-UTRAN) as the most advanced one. QoS is supported through an EPS bearer concept, providing bindings to resource reservation within the network.

The evolved NodeB (eNB), the Long Term Evolution (LTE) base station, is part of the access network that provides radio resource management, header compression, security and connectivity to the core network through the S1 interface. In an LTE network, the control plane signaling traffic and the data traffic are handled separately. The eNBs transmit the control traffic and data traffic separately via two logically separate interfaces.

The Home Subscriber Server, HSS, is a database that contains user subscriptions and QoS profiles. The Mobility Management Entity, MME, is responsible for mobility management, user authentication, bearer establishment and modification and maintenance of the UE context.

The Serving gateway, S-GW, is the mobility anchor and manages the user plane data tunnels during the inter-eNB handovers. It tunnels all user data packets and buffers downlink IP packets destined for UEs that happen to be in idle mode.

The Packet Data Network (PDN) Gateway, P-GW, is responsible for IP address allocation to the UE and is a tunnel endpoint for user and control plane protocols. It is also responsible for charging, packet filtering, and policy-based control of flows. It interconnects the mobile network to external IP networks, e.g. the Internet.

In this architecture, data packets are not sent directly on an IP network between the eNB and the gateways. Instead, every packet is tunneled over a tunneling protocol - the GPRS Tunneling Protocol (GTP over UDP/IP). A GTP path is identified in each node with the IP address and a UDP port number on the eNB/gateways. The GTP protocol carries both the data traffic (GTP-U tunnels) and the control traffic

(GTP-C tunnels). Alternatively Proxy Mobile IP (PMIPv6) is used on the S5 interface between S-GW and P-GW.

In addition to the above basic functions and entities, there are also additional features being discussed by the 3GPP that are relevant from a network virtualization viewpoint. One example is the Traffic Detection Function (TDF), which can be used by the P-GW, and in general by the whole transport network, to decide how to forward the traffic. In a virtualized infrastructure, this kind of information can be used to elastic and dynamically adapt the network capabilities to the traffic nature and volume.

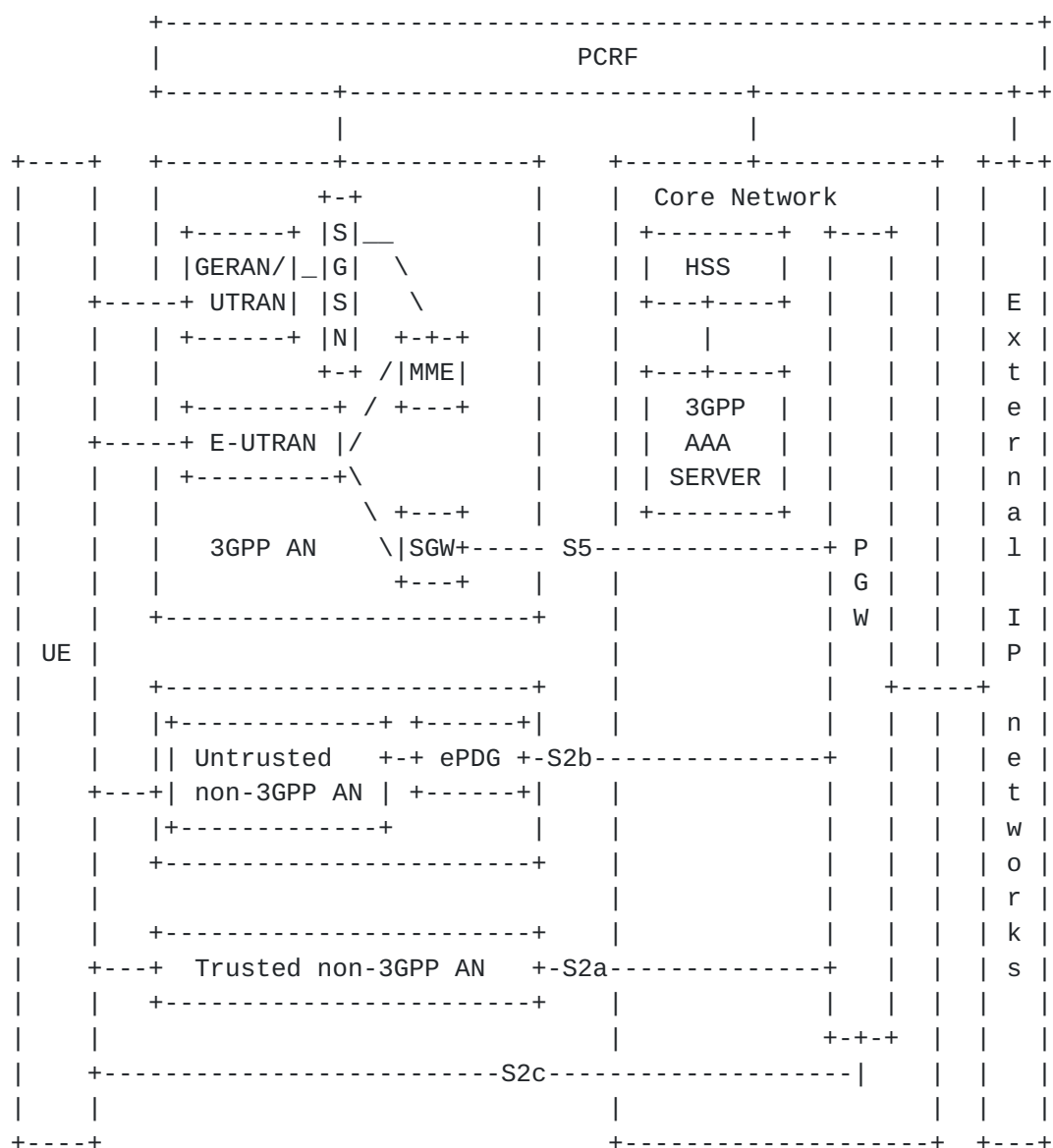


Figure 5: EPS (non-roaming) architecture overview

Appendix B. 5G

TBD. This section will include a description of main 5G building blocks.

Authors' Addresses

Carlos J. Bernardos
Universidad Carlos III de Madrid
Av. Universidad, 30
Leganes, Madrid 28911
Spain

Phone: +34 91624 6236
Email: cjbc@it.uc3m.es
URI: <http://www.it.uc3m.es/cjbc/>

Alain Mourad
InterDigital Europe

Email: Alain.Mourad@InterDigital.com
URI: <http://www.InterDigital.com/>

