Towards automated service-oriented lifecycle management for 5G networks

Rafia Inam, Athanasios Karapantelakis, Konstantinos Vandikas, Leonid Mokrushin, Aneta Vulgarakis Feljan, Elena Fersman
Ericsson Research, Sweden
Email: {firstname.lastname}@ericsson.com

Abstract—5G networks will be a key enabler for the Internet of Things by providing a platform to connect a massive number of devices with heterogeneous sets of network quality requirements. In this environment, 5G network operators will have to solve the complex challenge of managing network services for diverse customer sectors (such as automotive, health or energy) with different requirements throughout their lifecycle.

In this paper, we present current state of our work on automating part of the network service lifecycle management using knowledge management- and decision support techniques. We also present our ongoing implementation steps for such management function in 5G networks.

Keywords—5G, service management lifecycle, Internet of Things, network slicing, knowledge management, cloud computing.

I. INTRODUCTION

The Fifth Generation Mobile networks (5G) are seen as the network platform to support operation of reliable connectivity services for Internet of Things (IoT) applications, wherein billions of heterogeneous devices will be interconnected and connected towards the Internet [1]. IoT will cover a broad range of industry segments, from transportation networks and logistics, healthcare, energy and utilities to industrial automation, with different requirements, such as network coverage, latency and throughput [2].

In this environment, 5G network operators will have to deal with an unprecedented level of complexity to manage network services for multiple customer segments with different requirements. In order to achieve economies of scale, 5G networks will need to be designed with a greater degree of automation, because human involvement in network operator’s Operation and Business Support Systems (OSS/BSS) continues to decrease [3]. A paradigm shift from vertical mobile networks towards the horizontal implementation has been observed in 5G, as shown in Figure 1. It involves virtual slicing of the network and partitioning of network resources, as compared to previous and current generation networks. Network slice means a dedicated virtualized mobile network containing a set of network resources and providing a guaranteed quality of service. A slice can be allocated to a service layer application, for example, a logistics task involving a number of connected trucks, trailers and containers.

Driven by the IoT vision, 5G system will be able to support logical network slices and provide services. By service we mean a logical system of software, IT and telecom components organized to fulfil a given business objective that has specified requirements and outcome. Flexibility is a prerequisite for communication systems in IoT applications. The QoS requirements coming from the application domains are important to manage the network structure, its functionality, and the dynamic behaviour of such systems. This paper presents an approach for flexible (re)configuration in 5G network service lifecycle management to address these QoS issues.

Flexibility of 5G network management can be achieved by use of knowledge management techniques to generate and then assign/manage the network slices/resources to services. Knowledge management assumes a machine processable model (knowledge base) describing various concepts, their relations and behaviors combined with a set of methods such as logical inference [4], simulation [5] and verification [6]. Generally, the knowledge base contains digital representation of the domain of interest, i.e. generic- and domain specific models provided by experts and/or obtained automatically (e.g. using machine learning techniques). The analysis that knowledge management enables is done by automatically picking up relevant knowledge w.r.t. matters under investigation from the knowledge base, and applying the aforementioned methods to it.

This position paper contributes in addressing the challenge of automating the service creation and deployment parts of the network service lifecycle management as a first step towards automated management of the complete service lifecycle. Specifically, we focus on the following:

- Capturing of the customer network requirements in a formally-described Service Level Agreement (SLA). If the available network resources are not sufficient to fulfill the specified requirements, a negotiation process may be involved to refine the requirements.
- Creation of the network slice deployment instructions for the captured SLA in the 5G network in the form of a sequence of steps for network resource orchestration (a workflow).

Paper Outline: Section II identifies the challenges of the next generation mobile network management, and Section III gives an overview of the 5G networks as compared to the current networks. Our idea for automating network service lifecycle management using knowledge management techniques is presented in Section IV. Section V presents an example and finally, Section VI concludes the paper with a description of ongoing and future work.
II. CHALLENGES

A. Differentiated quality of service

The mobile networks of today are not capable of supporting applications with varying criticality and requirements on latency and bandwidth in a differentiated fashion. As an example, consider a case of a big event such as music concert or sport competition. During such an event there is a need in high bandwidth in the mobile network as the audience is typically actively using social media and video streaming. In the current generation mobile networks, performing a remote surgery in an ambulance passing by such an event would not be feasible as the network performance may not guarantee sufficient safety. Note that in this case there is a mix of criticality of two applications and there needs to be a mechanism to give higher priority to the data stream of the more critical application.

B. Changes in the mobile network affect the application

In some situations, especially when moving vehicles are involved, quality of service in the mobile network may heavily impact the application. An example of such service is a live video feed from the front camera of a long vehicle to the driver of a vehicle behind it to make overtaking safer. Such application requires high bandwidth and low latency of data transmission. When a vehicle is moving out of the area where the high bandwidth and the low latency can be guaranteed, the network layer should be able to notify the application so that it could switch to a simpler mode when only an assisting message is displayed to a driver in the rear vehicle to help overtaking. This application mode only requires low latency from the mobile network, and not the high bandwidth as the video stream is skipped.

C. Lifecycle management of the network slices

Service layer application, for example, a logistics task involving a number of connected trucks, trailers and containers, will have a network slice allocated to it. Similarly to allocating virtual machines in a cloud environment, the virtualized network functions that form a network slice need to be provisioned, monitored, possibly modified, and decommissioned along with the lifecycle of the corresponding service layer application.

Below we describe different approaches and technologies we intend to develop to meet the aforementioned challenges. We present 5G virtualization to achieve dynamic management of the network resources and better QoS control, and Service orchestration to allocate computational resources and enable automated provisioning of the services. We use knowledge management techniques to perform provisioning and lifecycle analysis of network slices.

III. NEXT GENERATION MOBILE NETWORKS

This section provides an overview on the next generation mobile networks, how they differ from the current traditional networks, and then briefly describes the concept of network slicing.

5G is the network for the next generation ultra-high broadband infrastructure that will support a fast, secure, and reliable services to billions of smart devices and cyber physical systems, such as automotive, robots and drones [7]. It aims to provide new services at low cost with a focus on providing a seamless and efficient communication capability. It will integrate the wireless access systems and IoT together, provide a seamless cooperation among them, and aims to fulfil very stringent requirements like high data rates, capturing and sending signals from a larger number of IoT devices, coverage area, latency, energy consumption, reliable communication, etc. Thus, it will create a universal communication environment and will address wider societal systems. An example could be connecting multiple systems like automotive, transport, safety, environment, and energy consumption in a bigger system of systems.

5G networks will need adaptation with intelligent computing and storage for IoT devices, efficient and controllable sharing of network infrastructure, better performance through supporting QoS aspects for diverse-natured services, and dynamic uploading/removal of services (also called service commissioning/decommissioning respectively) through the concept of virtualization of network slices and resources [8].

The paradigm shift from vertical implementations of current generation of mobile networks to 5G is shown in Figure 1. The current vertically implemented network architecture is presented in Figure 1(a), and the envisioned 5G horizontal implementation network architecture in Figure 1(b). The current structure supports year-long service lifecycle, while 5G aims to support both long- and short-term service lifecycle. Additionally, it aims to greatly reduce the service commissioning and decommissioning duration, e.g. reducing it from current 90-days for commissioning a new service to 90 minutes in future using 5G [8]. This is possible by automating the service lifecycle management process. Moreover, services and requirements will vary in their nature to much larger extent in 5G as compared to the current traditional network structure. This aspect puts more focus on SLA negotiation with a customer (before the start of a particular service) to guarantee certain levels of QoS requirements related to that service, and SLA renegotiation to change QoS requirements in case the required QoS could not be satisfied. For example, a service may require a particular bandwidth with a particular latency. The bandwidth and latency properties need QoS guarantees, and need to be a part of the SLA.

Virtualization is a known technique within the operating system (OS) and real-time communities. It provides a method to consolidate multiple, diverse-natured applications on a shared hardware platform (e.g. a computer) [9], [10]. Virtualization is a resource-management technique that partitions the system resources, such as processor, memory or network, in a way that provides the illusion of a full resource but with a fractional capacity [11], [12]. Its main advantages are providing isolation among services by partitioning resources among services, and fulfilling QoS requirements associated with these services. Multiple hierarchical Time Division Multiple Access (TDMA)
Virtualization can be used to partition multiple network resources into customized network slices depending on the service requirements [8]. Thus, many parts of the backbone network can be virtualized, including network connections (software defined networks - SDN) as well as parts of the traffic management (“switching”), which previously required physical hardware (e.g. IP Multimedia Subsystem [15] - IMS core nodes such as the x-CSCF nodes and HSS). Virtualization will support isolation among independently developed services/applications, enforce fault containment within the services/applications, and provide better quality control for each service/application [16]. Thus, virtualization will facilitate the dynamic addition, removal, or updating of services, and dynamic mapping of different network resources to services. Creation of network slices, and partitioning and allocation of resources to these slices is an ongoing work at Ericsson AB. In this paper, our main focus is on automation of service lifecycle management process and guaranteeing QoS requirements using virtual network slices.

IV. PROPOSED ARCHITECTURE

A high-level, conceptual view of the 5G network architecture is presented is Figure 2. It mainly consists of the following distinct, but interconnected layers:

1) **User layer**: presents different types of the users and provides an interface between the users of the network and the rest of the network system.
2) **5G Service Lifecycle Management layer**: captures customer requirements, negotiates, and creates SLAs.
3) **Orchestration and Control layers**: orchestrates the SLA agreements depending on the customer and the nature of the requested service. It then transforms this information to the control layer which is responsible for controlling (allocating, monitoring, deallocating, etc) the physical resources.
4) **Radio Access and Core Network layer**: contains the radio access infrastructure (e.g. base stations, WiFi access points, capillary gateways, etc.) and core network functions (e.g. charging and billing, authentication, switching, etc.) of the system.
5) **Devices layer**: comprises all physical devices of system users that can range from long range (e.g. devices equipped with HSPA, LTE, GSM modems) to low-power short-range devices (e.g. 802.15-powered personal area networks such as 802.15.4-enabled devices using 6LoWPan, Xbee or 802.15.1 - Bluetooth network stack [17]).

The rest of this section describes the layers of Figure 2 in greater detail, with an emphasis on the 5G service lifecycle management layer, namely the customer SLA negotiation, creation or the network slice description that meets the requirements of this SLA, and transfer of this description to the orchestration layer for deployment. Since the main focus of the paper is on the management and orchestration of services in the 5G network, detailed description and management of the Radio Access and Core Network and the Devices layers is out of the scope of this paper.

A. **User Layer**

Here we define a set of user roles to interface with the service lifecycle management layer. We consider the following three roles:

1) **5G network operators** own the 5G service lifecycle management, orchestration and control components of
The reader should note that in this paper we specifically focus on the service deployment part of the lifecycle management (i.e. the components filled with white colour) and we expect to address the service operation and decommissioning parts of the lifecycle in upcoming work.

2) **Network resource vendors** own and maintain network resources in the network, such as radio equipment and/or network core equipment. They can be current mobile network operators (e.g. WiFi access providers, cloud platform providers), or software vendors. To use their network resources, the vendors can register resources in the network using the 5G Lifecycle Management layer.

3) **Customers** order services along with the service-requirements specifications from 5G network operators. The orchestration of the network slice and functionality of lifecycle management functions depends on the nature of the ordered service(s) and the role of the user accessing it. For instance, consider that a big organization (i.e. a customer) needs a service to interconnect its employees’ mobile devices in a large corporate network. The requirements could be a combination of an indoor wireless access (from a WiFi access provider) and an outdoor cellular connectivity (e.g. using a set of HSPA capable base stations from a network operator), and some standard switching software (by a cloud provider).

The network slice will include the configuration of radio- and cloud assets that interoperate in order to seamlessly serve the customer requirements.

**B. 5G Service Lifecycle Management Layer**

We propose two different choices for the 5G lifecycle management system to interface with customers.

- The Service Order API that exposes a customer interface for creating Service Level Agreements (SLAs). These agreements are subsequently analyzed in order to create a network slice. The Service Order API can for example be a RESTful API [18] that uses secure authentication and JSON [19] representations for the customer requirements. Third party providers can call this API from their own systems.

- A generic web portal that uses the Service Order API described above and a Graphical User Interface (GUI) to showcase an idealized process of capturing customer requirements, negotiating, and creating an SLA. The GUI includes a self-service web portal where customers can specify a set of network requirements, such as the number of devices to be supported, latency and throughput of the network traffic, coverage and network access protocol(s), value added services based on analysis of their device’s traffic (e.g. relative levels of device activity), as well as financial parameters (e.g. pricing model and budget). Once the customer is satisfied with the service specification, the SLA creation and network slice deployment processes start to deploy the service.

**Fig. 3: Information flow in the 5G Network Lifecycle Management component for creating and deploying network slices.**

Figure 3 illustrates the “5G Service Lifecycle Management” layer in more detail, and in particular the process of network service deployment, from capturing customer requirements in an SLA to deploying a new network slice for this SLA.

The SLA creation process is assisted by a knowledge management and decision support system integrated to the 5G Service Lifecycle Management layer. The application of knowledge management in context of 5G networks is twofold. (A) It helps in automating the process of capturing high level user requirements of the system and translating them
to network requirements. For example, assuming a network service of connected vehicles, the high level requirements would relate to number of vehicles that will be connected through this service, area of coverage, etc., while the network level requirements would be the throughput/latency of the network connections, number of radio base stations that have to be configured to cover the desired area, etc. (B) It offers a white label optimization toolbox, a set of management functions that can be reused by different services. For example, a service routing buses (people logistics) and a service routing trucks (goods logistic) may have overlapping problem-solving approach, which leads to reuse same components from this toolbox.

The customer provides a set of requirements and then the decision support system checks whether the status of the available resources in the network can satisfy these requirements. If the requirements cannot be satisfied, then the system suggests an alternative set of requirements to the customer, as close to the original requirements as possible. For example, if a customer requests onboarding of 2000 devices with throughput of 1000 Mbps and the analyser within the 5G Service Lifecycle Management layer shows that it can only guarantee 1000 Mbps for 800 of these devices, it proposes this to the customer. Each device gets an access token, i.e. an encrypted string that identifies the device to the 5G network and lets it access network connectivity services under the requirements specified by the SLA. An example of such a service ordering interface is shown in Section V.

For monitoring purpose, the status of the network is continuously reported by the orchestration layer to the knowledge base, and is stored in a relational database.

Once an SLA has been created, it is stored in a database, and the deployment process can begin. The algorithm used for the deployment process, uses the SLA requirements (i.e. customer requirements) and Virtual Network Functions (VNFs) definitions as input and produces a network slice deployment instructions (see Section V).

C. Orchestration and Control layers

The orchestration and control layer receives network slice deployment instructions and executes them, i.e. deploys the slice on the physical resources. The network slices will mainly consist of VNFs. Additionally, a part of a network slice may also execute on a “bare metal”, i.e. use physically deployed infrastructure instead of virtual environment (e.g. to accommodate the legacy network functions).

A VNF is a self-contained functional component of the network infrastructure (for example core network nodes, policy control functions, load balancers, firewalls, etc.). In addition to pure network functional components, VNFs can provide common computing functionality such as storage (e.g. databases, filesystems), authentication, software development kits (SDKs), etc. Deployment of VNFs can span multiple nodes.

An important aspect of VNFs is “chaining”, wherein VNFs can be instantiatiated and topologically interconnected to form a service (chaining is explained in Section V) [20]. VNFs can be conceptualized as pieces of software executed in one or more virtual machines or containers, replacing the functionality that historically has been implemented as specific dedicated hardware components in the core network. We consider them as the building blocks of a network slice. VNF deployments can be complemented by Software Defined Networks (SDNs), which can be seen as analogy of physical network connections, but implemented in software and thus defining logical networks that can potentially share physical resources. The combination of these two virtualization technologies has numerous cost benefits, as it removes dependencies on specific hardware, reduces deployment time and the operations cost. A well known example, is the replacement of physical routers with physical network connections, with virtual router network functions and SDN connections [21].

The VNF descriptions are stored in the knowledge base and are used for reference when generating new network slice deployment instructions. A service description contains the following components:

- Deployment instructions: a dependency graph of VNFs the service depends on, each with a corresponding set of resource requirements and deployment instructions.
- Network configuration: a configuration of the networks that the service intends to use.
- Service parameters: an optional set of service specific functional and non-functional requirements.

An example of 3GPP (e.g. HSPA or LTE) mobile network service is shown in Figure 4. The real topology, with nodes representing different physical entities is illustrated in Figure 4(a). The virtual description of the topology Figure 4(b) consists of two VNFs, one for an IMS Core Service (networkCoreService type) and one for a 3GPP Radio Access Service (radioService type). Each of the aforementioned VNFs is chained to a set of other VNFs. In case of IMS, three VNFs encompassing the IMS core network (HSS, P-CSCF and S-CSCF nodes) and in case of Radio Access, one VNF representing the network controller (RNC), two VNFs representing base stations (Node-Bs) and two VNFs representing SGSN and GGSN nodes.

Both IMS Core and Radio Access services depend on a set of VNFs to be deployed for each, as well as a network configuration defining the networking between the VNFs. Nodes can coexist in a single container (or virtual machine), or can be spread over multiple containers, as dictated by the network configuration. We are considering to use Topology and Orchestration Specification for Cloud Applications (TOSCA) [22] as the language for expressing service deployments in the cloud and investigating different orchestrators such as Cloudify [23], Openstack Heat [24], Docker Compose [25], Canonical JUJU [26].

In the radio access case as illustrated in Figure 4, when the IMS Core Service is deployed, the orchestration creates a new instance in a virtual environment for each VNF. The boxes in the figure denote different components of IMS system. Note that the focus is on describing properties of IMS components,
V. A. Example Case

A. Prototype Implementation

In this section, we present our implementation of the architecture suggested in Section IV. Figure 5 illustrates the functional view of the implemented architecture. We describe the three parts of the user interface at the top layer. The service creation interface allows for creation of new services, and storage of these services in the service description database. A service consists of a textual description together with a set of requirements (for example numerical values or a range of values). The service description is currently stored in a relational database, but we will be transitioning to a more sophisticated description format such as XML-based USDL [27] or business TOSCA.

The customer interface allows for negotiation of SLAs and subsequent deployment of services as virtual resources in the cloud infrastructure. The cloud infrastructure consists of a set of servers forming a cluster controlled by Openstack [28], an infrastructure as a service provider, while Cloudify is set up on top as a platform provider. Cloudify accepts the descriptions of topologies as input, e.g. a number of servers and their network connections and interfaces with Openstack to deploy these topologies. The topologies are described in TOSCA language, as mentioned previously.

The service monitoring and decommissioning interface allows to monitor the service throughout its lifetime and decommission it upon request. The monitoring aspect of the service utilizes Ceilometer [29] which is part of Openstack.
Ceilometer collects metering information about CPU, Memory, Disk and Network utilization of both physical and virtual resources. The information is later on stored and processed by the ELK [30] stack. The decommissioning allows for decommissioning of services by interfacing with cloudify API.

Note that the prototype implementation includes a stub for radio assets (i.e. GGSN, SGSN, RNC and Node-B as presented in the previous section), where these assets are virtualized, using the NS-3 network simulator [31]. This allows for testing of the implementation without using real Node-Bs which are expensive and difficult to procure (due to licensed spectrum issues). The authors plan to extend the prototype system to using real Node-Bs in a controlled test site in the suburb of Kista, Stockholm, Sweden in the future.

### B. End-to-end service provisioning

This section describes the process of service deployment from capturing high-level user requirements, to matching those requirements to an NFV topology in the 5G Service Lifecycle Management layer and eventually sending a description of this topology to be deployed as a network slice in the underlying orchestrator. The purpose is to illustrate currently ongoing work and ideas on how service deployment may look like from an end-user perspective on the architecture described in the previous sections.

As an example use case, we are considering a radio access service which can be parametrized by different traffic priority classes. Instead of specifying network-level QoS requirements, customers use a web portal to specify expected quality of experience (QoE), which includes a high level indicator of the intended use of the service (e.g. high-definition video streaming, best effort web browsing, etc.), coverage, as well as value added services such as basic data analytics (starting from simple storage of data to identification of peak data transfer hours, most active devices per day, etc.).

Figure 6 illustrates a conceptual interface, currently under implementation, presented to customers of this service. The map illustrated on the right hand side of the figure can be used as canvas for users to designate the desired area of radio coverage, while on the left hand side, the users can describe the quality desired from the network using high level requirements. Additionally, basic data analytics services can be orchestrated and deployed in the corresponding cloud infrastructure. Finally, a slider can be used from the users to indicate the scale of the deployment. The service can thus support users of different size and requirements, from private individuals, to small and medium enterprises up to large corporations. Table I illustrates potential users of the service that may have different requirements.

<table>
<thead>
<tr>
<th>Customer Type</th>
<th>Description of Service</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private Individual</td>
<td>Basic connectivity for personal, mobile devices while roaming in the area.</td>
<td>One traffic class, Best Effort, no additional services, 2 devices.</td>
</tr>
<tr>
<td>City bus transport authority</td>
<td>Scheduling of 50 buses based on demand of passengers waiting in bus stops rather than fixed schedules.</td>
<td>One traffic class, mission critical data, including subscriber mobility analytics and storage of critical data for troubleshooting, 50 devices.</td>
</tr>
<tr>
<td>Mining Company</td>
<td>Mining company would like to automate vehicles carrying mined ore from the mining site to a nearby depot. The network to support these vehicles includes both prioritized traffic (mission critical data and media streams of cameras attached to vehicles for monitoring purposes).</td>
<td>Two traffic classes, media streaming and mission critical data, including basic data storage of critical data for troubleshooting if needed, 100 devices.</td>
</tr>
</tbody>
</table>
VI. CONCLUSIONS AND FUTURE WORK

The 5G networks will connect billions of devices and will provide connectivity services to Internet of Things (IoT) applications from different industry segments with varying criticality and heterogeneous sets of network quality requirements (i.e., latency and bandwidth). As such, 5G network operators will need to address the problem of simultaneously managing network services for diverse customer segments with different requirements throughout their lifecycle.

In this position paper, we have presented current state of our work on automating the network service lifecycle management for such networks using knowledge management- and decision support techniques. This automation will provide flexibility in service deployment and decommissioning, reduce deployment time from months and weeks to minutes and seconds, and support better QoS control. Specifically, we have focused on the part of the management process that includes: capturing of service driven network requirements by establishing Service Level Agreements (SLAs). This includes SLA negotiations, handled by a knowledge management system, when the consumer requirements can not be fulfilled due to insufficient network resources. Further, it includes the creation of network slice deployment instructions for a given SLA in the 5G network in the form of a workflow, i.e. a sequence of steps for the network resource orchestration.

Currently, we are implementing our ideas using orchestrators (such as Cloudify [23], Openstack Heat [24]), and a language to express the network slice descriptions (e.g., TOSCA [22]). In the future, we plan to develop the knowledge management by translating service description and requirements specified by a user into exact configuration of resources and service chains of VNFs. We also intend to evaluate the complete implementation on a real network in the context of a real use case. Once the implementation is completed and the basic functionality is evaluated for correctness, the next step will be to extend the work towards other layers of the lifecycle management, such as network operations monitoring and assurance of SLA fulfillment. Adding support for more user roles can form another direction of the future work.

REFERENCES


