

Dynamic 5G Slices for IoT applications with diverse requirements

Evgenia Kapassa, Marios Touloupou, Panagiotis Stavrianos, Dimosthenis Kyriazis
Department of Digital Systems
University of Piraeus
Piraeus, Greece
{ekapassa,mtouloup,pstav,dimos}@unipi.gr

Abstract — The wide deployment of Internet of Things (IoT) applications raises the need for underlying environments and infrastructures that can both support the traffic data requirements and enable the process of distributed IoT data. 5G - the fifth generation of mobile, cellular technologies, networks and solutions, provides the enabling environments to fulfil the aforementioned IoT applications diverse requirements. 5G promises to bring the reliability, latency, scalability and adaptability that would be needed for several services in the IoT space and beyond. To satisfy these demands, network slicing has been envisioned as the promising solution in an IoT-oriented 5G architecture. In this paper we propose an efficient IoT-oriented architecture supporting network slicing for 5G-enabled IoT services over the 5G Core, in order to meet the requirements for establishing an efficient network with high capacity, while ensuring the maximum Quality of Service (QoS) to the end users-applications.

Keywords— IoT, 5G, Slicing, QoS enforcement, NFV/SDN

I. INTRODUCTION

Next-generation 5G networks will cater for a wide range of new business opportunities. Such opportunities will provide support for advanced mobile broadband services with optimal spectral and energy efficiency. Remote operation of machinery, Telesurgery, and Smart Metering comprise a subset of applications requiring reliable connectivity, but with vastly different characteristics and requirements [1].

Internet of Things (IoT) is another broad area for development with 5G wireless network becoming its backbone to its evolution. The IoT refers to the interconnectivity of billions of physical devices around the world through the Internet, collecting and sharing data [2]. Analyst Gartner calculates that around 8.4 billion IoT devices were in use in 2017, up 31 percent from 2016, and this will likely reach 20.4 billion by 2020 [3], the exact point of history when standardized 5G networks will start to hit the market and deliver unprecedented levels of connectivity [4]. Therefore, by 2021 5G's broad enablement of IoT use cases will drive 70% of G2000 companies to spend \$1.2 billion on connectivity management solutions as highlighted in IDC [5]. IoT enables the critical development of a large number of applications in different domains, including smart cities, smart healthcare, intelligent transportation, industrial automation and disaster response. Thus, based on this vast introduction of IoT services, the production of large volumes of data is rendered as a predicted outcome. To satisfy the requirements of new IoT services, 5G aims to achieve the enhanced circulation of exchanged data than 4G LTE standard, in terms of speed and quantity and also to

skyrocket the support of connected devices up to 10–100 times as well as 5 times reduced latency [6-7].

In parallel, Edge Computing comprise the basis for introducing the term of “intelligence” at the terminal nodes of the infrastructure. Through its advent, data are processed as close as possible to the collection point and services function are decentralized. This topology provides the ability of taking operational decisions at the terminal nodes with semi-autonomous rights. Equally, the Edge Computing topology is in contrast with rebuilding the data feed to centralized data center and processing them before it pushes business decisions back to the edge platform. One main advantage of delivering intelligence on the edge is the optimum spectral efficiency on IoT platforms (mobile networks, UAVs, etc.). Since the quantity of circulated data is skyrocketed, the existing spectral ratio is deficient to efficiently transfer the data to the centralized storage. Secondly, the propagation delay in making data processing and decision-making far from the edge is too high for many applications. Several IoT applications are operational in nature, demanding the ability to operate with fast decisions. Consequently, giving the ability to the IoT services to adopt the aforementioned functionalities on the edge, it extends the plethora of applications being developed.

Thus, 5G networks aim to meet various Quality of Service (QoS) requirements, in different application scenarios (e.g. in terms of data transmission rate and latency) by offering full connection to all “things” [8]. At the same time, the ever-increasing traffic demand is pushing network operators to detect new cost-efficient solutions toward the deployment of future 5G mobile networks. Network slicing is a promising paradigm yet realizing it is not without challenges [9]. Mobile operators are facing tremendous traffic increases with the introduction of smartphones and tablets, especially due to content rich multimedia and cloud applications, and the upcoming vertical IoT Services as mentioned above [10].

To address the aforementioned challenges, we propose a IoT-Oriented Architecture that enables the facilitation of concurrent IoT applications and services with diverse requirements based on dynamic 5G slices allocation. In detail, it introduces a dynamic slice manager that allows an automatic selection of customized network slices while providing optimized solutions for different IoT applications. Last but not least, the end nodes of the 5G infrastructure, such as devices and objects, will be virtualized enabling concurrent IoT functionalities and services to be simulated and executed in different 5G network slices.

The remaining of this paper is structured as follows. In Section II, similar approaches in the related field are presented among with the motivation for the proposed

architecture. In Section III, a reference is made on the key advantages of 5G networks, while Section IV introduces the proposed IoT-Oriented architecture among with the functionality of each component included. Finally, the paper concludes with the future work and potentials for the current research.

II. MOTIVATION AND RELATED WORK

The 5G technology will empower a wide range of future industries, from retail to education, transportation to entertainment and smart homes to healthcare. The IoT scenarios are characterized with ultra-dense interworking of billions and billions of devices through a variety of technologies for the delivery of smart personalized services and applications. The main business verticals of applications and services, enabled by the evolution of the IoT services, are several, some of which are Smart Home Connectivity, Connected Cars, Remote Surgery and Mission Critical Applications. Therefore, a need of fulfilling the plethora of controversy requirements is present, especially regarding the network recourses [11].

As stated in [12], edge computing is another important technology for IoT services [13]. Due to data transfer with limited network performance, cloud computing's central structure becomes ineffective for processing and analyzing large amounts of data collected by IoT devices [14]. As limb calculation eliminates computer operations from the cloud near the IoT devices, the data being transferred is normally reduced by processing procedures. The authors in [15] state that edge computing could scale from a single person to a smart home to even an entire city. Given that a city with 1 million people will produce about 180 petabytes of data per day by 2019, the benefits could be of paramount importance. However, to realize this vision, systems, network and application communities must work together in front of the many groups that could benefit from technology such as environmental and public health, law enforcement, fire protection and services of general interest.

Till now, many of these applications perform on a satisfying level with the low data rate solutions, along with the low costs, explaining the success of current cellular IoT solutions. Yet, while the next generation wireless networks (i.e. 5G and beyond) comes closer, new capabilities arise. Towards this end, the authors in [16], provides a comprehensive survey on the utilization of AI integrating machine learning, data analytics and natural language processing (NLP) techniques for enhancing the efficiency of wireless network operation. In the same context, to enhance service provisioning and satisfy the coming diversified requirements, especially from the tons of different IoT applications and services, it is necessary to revolutionize the cellular networks with cutting-edge technologies. The standardization of next-generation (5G) cellular networks is being expedited, which also implies more of the candidate technologies will be adopted [17].

The trend is to support a heterogenous 5G Core (5GC) network, in which all the network operators will be connected through one single core infrastructure, regardless of the corresponding access technologies, their type of application or the amount of recourses needed. The basic concept is based on network slicing which is an end-to-end

(E2E) topic, as stated in [18]. In the case of 5G, a single physical network will be sliced into multiple virtual networks that can support different Radio Access Networks (RANs), or different IoT application types (i.e. massive machine type communications, mission critical applications) with distinct requirements, running across a single RAN. It is envisaged that network slicing will primarily be used to partition the 5GC network. 3GPP has proposed a data model, which consists of a list of Network Slice Subnetworks instances (NSSI), which contain a set of network functions and the resources for these network functions which are arranged and configured to form a logical network [19].

According to [20], a logical architecture for network-slicing-based 5G systems is introduced, and a scheme for managing mobility between different access networks is presented. Also, the authors recommended a solution based on the LaGrangian dual decomposition method. More specifically, they decompose the objective function into a master problem and $K * NK * N$ sub-problems (for K small cells and N subchannels). The Karush-Kuhn-Tucker (KKT) conditions are used to get the optimal power allocation, and the sub-gradient method is exploited to update the LaGrangian multipliers to obtain the optimal subchannel allocation. Moreover, authors in [21], describe their model of consisting with four main elements: the service slice layer, the virtual network layer, the physical resources, and the admission control manager. The service slices present different services (e.g., car management, TV streaming and web browsing) which require resources to be served. The virtual network layer provides an abstraction of the physical network resources. The physical resources refer to the radio resources available in the virtual network, and the Admission Control Strategy where a heuristic-based prioritized admission control mechanism had been designed [22].

Motivated from the aforementioned challenges we propose an IoT-Oriented architecture which will be based in the existing 5G architecture, enabling also network slicing for better QoS provisioning and assurance for any type of IoT application. In addition, we introduce a Dynamic Slice Manager, which can be used in order to deal with the diverse IoT application network requirements and provide a global optimization of the resources allocated to application slices. Last but not least, through the optimized recourse allocation into the 5GC, it will be achieved and probably guaranteed the appropriate QoS for each application/service.

III. 5G NETWORK KEY ADVANTAGES

Software Defined and 5G Networks - a combination of Software-Defined Networking (SDN) and Network Function Virtualization (NFV) - is a significant area that tend to make IoT services and applications more flexible. 5G will allow upcoming devices to be smarter, faster and more reliable introducing key advantages, as depicted in an abstract way in Figure 1.

A. Network Function Virtualization (NFV)

Network Function Virtualization (NFV) decouples network functions from the underlying hardware and centralizing them at network servers making the network

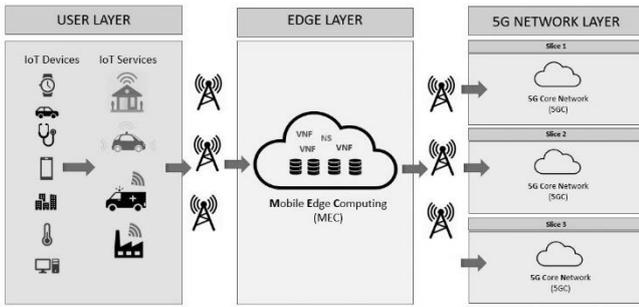


Fig.1 . High Level Architecture

architecture highly flexible from quick and adaptive reconfiguration [23]. Such network functions could be virtualized IoT services for any kind of diverse areas, IoT for smart cities, public transportation, even in healthcare domains. The diversity and the advanced needs of those services could be easily managed through the transformation to NFV and the deployment of the applications in a Software Defined environment. As stated in [25], there are many key features of NFV / SDN, virtualized technology and cloud computing in general: on demand for self-service, broad network access, pooling of resources, rapid elasticity and measurement of services. A set of attributes though, that have a significant impact on the OPEX equation include: automated elasticity, software and hardware independence, multi-tenancy and resource pooling, hardware uniformity, virtualized software infrastructure, operations process streamlining, etc. Furthermore, worth mentioned is the fact that with the transition to NFV/SDN, many service providers are expected to realign network operations to support an operating model where network services and network resources are efficiently managed.

B. Software Defined Networking

SDN is a network architecture approach that enables the network to be intelligently and centrally controlled, or ‘programmed,’ using software applications. It also aims at making the network as agile and flexible as the virtualized server and storage infrastructure of the modern data centers. The goal of SDN is to allow network engineers and administrators to respond quickly on changing business requirements. SDN brings ease of programmability in changing the characteristics of whole networks. This simplifies the management of the network, since it is decoupled from the data plane. Therefore, network operators can easily and quickly manage, configure and optimize network resources with dynamic, automated and proprietary-free programs.

Network Functions Virtualization as presented in the previous paragraphs, can be implemented without a SDN being required. Although, the combination of the aforementioned technologies, can enhance the performance, simplifying compatibility of different IoT services with conflicting requirements, and facilitate operation and maintenance procedures at the same time [24].

C. Network Slicing

Network slicing in 5G networks defines logical sub-networks, consisted by a mix of shared and dedicated resource instances, such as radio spectrum or network

equipment, and Virtual Network Functions [24]. This kind of virtualization allows the decoupling of network functions from the hardware and enables the creation of an abstract building block per IoT application type and provide the needed resources for each service. Thus, different QoS requirements can be tackled for different IoT use case needs (e.g. URLLC, eMBB, mMTC). As a result, optimized resource allocation is achieved while at the same time prevents the waste of network resources.

Although the slicing concept is presented in a way to enable different classes of communications, is not limited only to this, as it can also operate virtual networks on top of physical infrastructures, with virtual resource isolation and virtual network performance guarantees. Specifically, slicing technologies in the SDN IoT case, should coordinate different users’ controllers to isolate data plane traffic, and make users’ controllers have the illusion that there are no other users affecting the network.

Achieving the above goal, slicing implementations should exist beyond the general NFV Management and Orchestration Platform (MANO) platform [26]. Attaching a Dynamic Slicing Manager, on top of the MANO Framework, bridge the 3GPP Standardizations in 5G networks with the concurrent and diverse IoT world.

IV. IoT-ORIENTED ARCHITECTURE

In this paper, we propose an IoT-oriented architecture that aims dynamically provisioning of 5G slices, with respect to concurrent IoT applications and services, having diverse requirements. In Figure 2 the proposed architecture of the major components is introduced, which each one of them is going to be described in the following sections. Briefly, in the upper layer, the user and the IoT devices and applications are the key participants. This is the entry point of the 5G-enabled IoT application/s lifecycle, where the possible conflicting requirements are defined and set.

A. Aggregation Mechanism

The Aggregation Mechanism is responsible for a) gathering the requirements by the concurrent and different IoT applications/services, b) identify possible conflicting requirements c) resolve the conflicts and provide an optimal-aggregated result. To be more specific, the aggregation mechanism intends to resolve conflicts for heterogeneous requirements of different and parallel IoT applications/services, taking into consideration the multi-objective optimization techniques and requirements engineering.

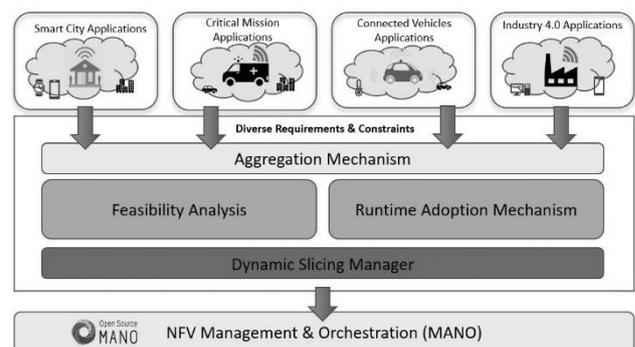


Fig.2 . Proposed IoT oriented Architecture

Specifically, the mechanism will provide a specific manner - a 5G specific language - to the application and service providers, in order to gather the necessary requirements aligned with ITU requirements and 5G PPP KPIs. In that way, the diverse IoT applications will be expressed in a common way - in terms of definition of recourse requirements.

B. Feasibility Analysis

During the Feasibility Analysis - a process triggered after the collection of application's requirements - an end-to-end optimal deployment planning for each application/service starts. Taking under consideration the available infrastructure recourses and the current situation of the network (i.e. active users, connected IoT devices etc.), the feasibility analysis decides whether the requested resource demands are achievable or not.

C. Dynamic Slicing Manager

The main concern of the Dynamic Slicing Manager is to support the dynamic provisioning of network slices using the aggregated and feasible requirements coming from the concurrent IoT applications/services. The Dynamic Slicing Manager will autonomously analyze the application requirements as well as the available resources, and then decide upon the most appropriate deployment of applications/services. The most appropriate deployment should achieve the best balance between system performance, QoS and cost. Moreover, a scale-in/out model for network slices will allow the reservations of only the necessary resources to the deployed slices.

Furthermore, unlike fixed network slices which can be scaled up by adding more hardware resources, radio access network (RAN) slicing quickly runs into a physical constraint of the limited availability of spectrum. For this reason, Dynamic Slicing Manager it will also apply network slicing from the mobile network operator's (MNO's) perspective. Therefore, the main objective is to properly manage the compensation between flexible and static resource assignments, considering the heterogenous requirements, coming from a diversity of IoT devices and applications/services.

D. Runtime Adoption Mechanism

Given that not only the data coming from different IoT devices, but also the applications and service themselves are data intensive, the Runtime Adoption Mechanism should perform all kind of operations according to data-driven decisions for the appropriate infrastructure management. Regarding infrastructure-related aspects (e.g. resource re-allocation, new deployment patterns compilation, etc.) and data-related aspects should also be concerned. The performance and optimization of the infrastructure recourses will be achieved by basing all infrastructure management decisions (e.g. resource allocation, orchestration, etc.), as well as the possible slicing re-allocation.

Thus, real-time adaptation across all levels of the architecture is a key towards enhanced dynamicity. To this end, the Runtime Adoption Mechanism should collect information for the three main building blocks: a) the aggregation mechanism, b) the feasibility analysis mechanism and c) the MANO Framework. This information

is going to trigger a new lifecycle as the new requirements and data are going to be aggregated through the Aggregation Mechanism, interlinked and evaluated through the Feasibility Analysis Mechanism. The whole process is inside an internal loop until, providing the best QoS the infrastructure can guarantee at that point. It is worth pointing out that the decisions will be adaptable during runtime in order to ensure the quality of the data-intensive of IoT application.

V. REMOTE MEDICAL EXAMINATION USE CASE

Remote medical examination is a key use-case that aims to break the obstacle of geographical boundaries in providing high quality healthcare in the most complex medical situations. To this end, IoT and wireless sensor networks based on mobile communication such as 5G, can provide remote monitoring for health parameters. The ultra-dense device-to-device communication networks and co-operative multi-node/cell networks as described in the proposed architecture, can support reliable and low-latency uplink connectivity for very large numbers of connected devices, including wearable sensor for health monitoring and haptic devices. In the scenario of the remote examination, each patient will be equipped with a wearable fitness tracker, and a mobile smartphone app. These sensors will make it possible to monitor and visualize blood levels, heart rate and stress levels, as well as geographical location (via GPS) in real time. Considering that each monitored person could wear a number of different sensors, the amount of IoT devices increases to very high levels. Each health sensor device could communicate directly with the edge layer, minimizing the communication path to the health care provider, reducing the latency and ensuring continuity and seamless mobility, when the monitored person is moving. The proposed IoT oriented architecture is capable of creating and managing virtual instances of access networks. The network slicing will keep communication interference between the different networks, ensuring extremely high throughput and ultra-low latency. In this case, the bandwidth allocated to each IoT wearable for the health monitoring the management of the traffic including delay, loss, active bearers, etc., are synchronized by the NFV controller located in the MEC.

VI. CONCLUSIONS

5G is one of the most sophisticated wireless technologies we have ever developed so far. It will revolutionize the entire area where wireless network can be used for efficient communication. Even though the specifications of 5G remain unstable, the next generation of mobile technology is predicted to greatly benefit IoT innovation. 5G networks among with dynamic network slicing are expected to provide faster speeds, lowered latency as well as network support for massive increases in data traffic coming from different and numerous IoT devices. With network slicing, operators can allocate the appropriate amount of network resources to a specific slice, suitable for diverse IoT requirements.

In this paper, we have presented an IoT-oriented architecture at the top of MANO Framework, that enables the facilitation of concurrent IoT applications and services with diverse requirements based on dynamic 5G slices allocation. The proposed enabling components allow

heterogenous IoT applications to be integrated at any time and become immediately available in an ultra-reliable 5G network infrastructure, providing the necessary amount of recourses during runtime.

The next step in our work, include the competition of design and implementation of all the components of the proposed architecture, while we aim to streamline and enhance the interfaces between the different components, as well as the interaction with the MANO. What is more, we envision the need for a deeper investigation in real-time adoption based on Quality of Experience (QoE) parameters extracted from the users.

ACKNOWLEDGMENT

This work has been partially supported by the 5GTANGO project, funded by the European Commission under Grant number H2020ICT-2016-2 761493 through the Horizon 2020 and 5G-PPP programs (<http://5gtango.eu>).

REFERENCES

- [1] "A vision of the 5G core: flexibility for new business opportunities", Online:<https://www.ericsson.com/en/ericsson-technology-review/archive/2016/a-vision-of-the-5g-core-flexibility-for-new-business-opportunities>
- [2] "What is the IoT? Everything you need to know about the Internet of Things right now", Online: <https://www.zdnet.com/article/what-is-the-internet-of-things-everything-you-need-to-know-about-the-iot-right-now/>
- [3] "Gartner", Online: <https://www.gartner.com/newsroom/id/3598917>
- [4] "5G in 2020 – what should we expect?", Online:<https://blog.gemalto.com/mobile/2017/06/09/5g-2020-expect/>
- [5] "IDC Reveals Worldwide Digital Transformation Predictions", Online: <https://www.idc.com/getdoc.jsp?containerId=prUS43188017>
- [6] J. Ni, X. Lin and X. S. Shen, "Efficient and secure service-oriented authentication supporting network slicing for 5G-enabled IoT," in *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 3, pp. 644-657, March 2018.
- [7] X. Shen, "Device-to-device communication in 5G cellular networks," *IEEE Netw.*, vol. 29, no. 2, pp. 2–3, Mar./Apr. 2015.
- [8] A. Osseiran et al., "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," in *IEEE Commun. Mag.*, vol. 52, no. 5, , pp. 26–35, May 2014.
- [9] M. Peng et al., "Fronthaul-constrained cloud radio access networks: Insights and challenges," in *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–60, Apr. 2015.
- [10] "GSMA, The Mobile Economy 2018", Online: <https://www.gsma.com/mobileeconomy/>
- [11] T. Baker, M. Asim, H. Tawfik, B. Aldawsari, R. Buyya, "An energy-aware service composition algorithm for multiple cloud-based IoT applications", in *Journal of Network and Computer Applications*, vol. 89, pp. 96-108, 2017
- [12] H. Li, K. Ota and M. Dong, "Learning IoT in edge: deep learning for the internet of things with edge computing," in *IEEE Network*, vol. 32, no. 1, pp. 96-101, Jan.-Feb. 2018.
- [13] G. Rodrigues et al., "Hybrid method for minimizing service delay in edge cloud computing through VM migration and transmission power control," in *IEEE Trans. Computers* vol. 66, no. 5, pp. 810–19, May 2017.
- [14] J. Liu et al., "Energy consumption minimization for FiWi enhanced Lte-A hetnets with UE connection constraint," *IEEE Commun. Magazine* vol. 54, no. 11, pp. 56–62, , Nov. 2016
- [15] W. Shi and S. Dustdar, "The promise of edge computing," in *Computer*, vol. 49, no. 5, pp. 78-81, May 2016.
- [16] T. E. Bogale, X. Wang, L. B. Le, "Machine intelligence techniques for next-generation context-aware wireless networks", Cornell University Library, 2018
- [17] R. Li *et al.*, "Intelligent 5G: when cellular networks meet artificial intelligence," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 175-183, Oct. 2017.
- [18] "ETSI's new zero touch network and service management group starts work", Online: <https://www.etsi.org/news-events/news/1254-2018-01-news-etsi-s-new-zero-touch-network-and-service-management-group-starts-work>
- [19] 3GPP. "A study on management and orchestration of network slicing for next generation network", 2017
- [20] H. Zhang, N. Liu, X. Chu, K. Long, A. H. Aghvami and V. C. M. Leung, "Network slicing based 5G and future mobile networks: mobility, resource management, and challenges," in *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138-145, 2017
- [21] M. Jiang, M. Condoluci and T. Mahmoodi, "Network slicing management & prioritization in 5G mobile systems," in 22th European Wireless Conference, pp. 1-6, 2016
- [22] S. Swaminathan and N. R. Raajan, "Allocating resource on multiple user basis in OFDMA systems by using regenerative algorithm," in *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, pp. 1710-1716, 2017
- [23] I.F. Akyildiz, SC Lin, P. Wang, "Wireless software-defined networks (W-SDNs) and network function virtualization (NFV) for 5G cellular systems: An overview and qualitative evaluation", *Computer Networks*, vol. 93, pp. 66-79, 2015
- [24] T. Kuo, B. Liou, K. C. Lin and M. Tsai, "Deploying chains of virtual network functions: on the relation between link and server usage," in *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1562-1576, Aug. 2018.
- [25] K. Samdanis et al, "5G network slicing Part I: concept, principles and architectures", in *IEEE Communications Magazine*, 2017
- [26] "What is NFV MANO?", Online: <https://www.sdxcentral.com/nfv/definitions/nfv-mano/>