Conformal Mapping for Optimal Network Slice Planning Based on Canonical Domains

Abdelquoddouss Laghrissi, Tarik Taleb, and Miloud Bagaa

***Abstract*— The evolution towards 5G consists of managing highly dynamic networks and making decisions related to the provisioning of networks in an as-a-service and cost-aware fashion. This is translated by 5G verticals that are dedicated to specific services, applications, or use cases fulfilling the constant demand of vertical industries. In this vein, to achieve the high-level goals defined by operators and service providers, and to answer to the elasticity and low-latency specifications of the upcoming 5G mobile system, the optimal placement of virtual network functions must cope with the non-uniform service demands and the irregular nature of network topologies. This paper addresses this issue by mapping the non-uniform distribution of signaling messages in the physical domain to a new uniform environment (i.e., canonical domain) whereby the placement of core functions is more feasible and efficient by means of Schwartz–Christoffel conformal mappings. The experimentation results, compared to some baseline approaches, have proven the efficiency of the conformal mapping based placement in allocating the virtual resources (i.e., virtual CPU and virtual storage) with regard to the optimal end-to-end delay, cost and activated virtual machines. Another interesting contribution is that all placement decisions are based on a realistic spatio-temporal user-centric model, which defines both the mobility of user equipments and the underlying service usage.**

***Index Terms*— NFV, cloud, network slice, 5G, mobile network, conformal mapping, and VNF placement.**

I. Introduction

OVER the years, network architectures have evolved with different generations of telecommunication technologies (i.e., 2G, 3G and 4G) to serve end users. The 5^{th} generation of mobile networks will support massive numbers of connections and promise the support of ultra-short latency services, connecting different entities (e.g., connected cars, mobile devices, and sensors) and impacting different sectors (e.g., automotive, e-learning, and e-health). 5G relies upon different technologies, with the Network Function Virtualization (NFV) and Software Defined Networking (SDN) being its keystones.

NFV stands for virtualizing the services and network functions that are currently provided by a dedicated hardware. NFV reduces the amount of resources and hardware needed to start and operate a network service. Different Virtual Network Functions (VNFs) can inter-work and be instantiated within the same data center (DC) or across multiple DCs, which results into a flexible and dynamic network that is rapidly deployable in the cloud [5], [6].

The Mobile Edge Cloud (MEC) consists of virtualized small-scale DC infrastructures deployed in a variety of locations, which enables the provisioning of computing, storage, and network resources in closer proximity to end users. The customization of services taking into account the mobility patterns, preferences and requirements of end-users has motivated many research works, involving new architectures which combine cloud computing and mobile networks [9], [20]. MEC provides processing capacity and storage within the range of User Equipments (UE) by facilitating the Mobile Cloud Computing (MCC). The latter is achieved by placing Edge Clouds (EC) jointly with the mobile network base stations, at the mobile network edge, which eliminates the need to move storage and computation of intensive and heavy services from a UE to a centralized cloud server [31].

This raises the need for network slicing, which ushers new ways to divide the traditional single network into multiple instances; each of them being dedicated to a specific service or a given requirement [32]. Network slicing is the key concept to a flexible network provisioning and support of different business verticals with diverse service and resource requirements [22]. In order to answer to the increasing service requirements of incoming requests, network slices must join the cloud resources (e.g., CPU and storage) with the network architecture (e.g., network functions) [30]. Within the envisioned 5G mobile system, network slices consisting of a set of VNFs can be dedicated to Internet of Things (IoT) services, as well as for specific verticals (e.g., virtual reality and autonomous driving).

With the explosion of data usage and the eager need for competitive usage plans, service and cloud providers are challenged to determine what resources are necessary to deploy and operate to satisfy the requirements of a particular service while still staying within reasonable cost ranges. Placing the resources in close proximity to the cloud users, with just the needed amount of virtual resources to deploy, will have an important impact on the Capital Expenditure (CAPEx) and the

Manuscript received October 4, 2017; revised February 5, 2018; accepted February 27, 2018. Date of publication March 12, 2018; date of current version May 21, 2018. This work was supported in part by the Academy of Finland Project CSN-under Grant 311654 and in part by the European Union's Horizon 2020 research and innovation program under the 5G!Pagoda project with under Grant 723172. This paper was presented at the IEEE Wireless Communications and Networking Conference [24]. *(Corresponding author: Abdelquoddouss Laghrissi.)*

A. Laghrissi and M. Bagaa are with the Department of Communications and Networking, School of Electrical Engineering, Aalto University, 02150 Espoo, Finland (e-mail: abdelquoddouss.laghrissi@aalto.fi; miloud.bagaa@aalto.fi).

T. Taleb is with the Department of Communications and Networking, School of Electrical Engineering, Aalto University, 02150 Espoo, Finland, and also with Sejong University, Seoul 05006, South Korea (e-mail: tarik.taleb@aalto.fi).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2018.2815436

Operational Expenditures (OPEX); associated with operating both the needed cloud resources during a given period of time.

The idea behind this paper considers a different angle to deal with the provisioning of such virtual resources. It goes beyond the VNFs barrier and considers dealing with network functions' VMs deployment, in an already set up EC environment, and leveraging the characteristic of "uniformity" offered by conformal maps (i.e., the preservation of angles). Even though the idea seems simple, it introduces many important challenges, mainly the conformal mapping spatial transformation and its corresponding inverse function. As elaborated in this paper, the theoretical and practical challenges that should be considered are numerous and the analysis in the canonical domain is not straight forward.

The remainder of this paper is structured as follows. Section II discusses some relevant research work. Section III defines the target problem. Our solution, dubbed "Canonical Domain Framework for Slice Planning (CDSP-F)", is introduced in Section IV. The experimentation setup and the obtained results are presented in Section V. Finally, the paper concludes with some insights and future research work in Section VI.

II. Related Work

Several research work have addressed the optimal placement of specific VNF types taking into account the requirements of different 5G verticals. For video streaming verticals, the ability to move transcoding resources could help in the optimization of the network's bandwidth and latency, ultimately offering better Quality of Experience (QoE) [23], [25]. In the same vein, dynamic live migration of streaming VNFs to different locations while the streaming service is taking place, is made possible thanks to coupling NFV and open cloud computing architecture [4], as well as using heuristic designs [15]. In [7], [10], [11], game-theoretic models are used to evaluate the needed joint caching for content providers, transit networks and access networks. Content caching in wireless networks can exploit the backhaul links for collaborative caching, offering better pricing choices and better QoS [12], [26]. Also, a Content Delivery Network as a Service (CDNaaS) platform was proposed to manage a high number of videos deployed on virtualized caches, transcoders, and streamers, using a Gurobi Optimization tool with the objectives of maximizing the QoE of the streaming service and reducing the overall system cost [33].

As key VNFs of the Evolved Packet Core (EPC) architecture [4], many solutions addressed the optimal placement of Packet Data Network Gateways (P-GW) and Serving Gateways (S-GW) [14], [16]. In [17], the respective authors aimed for the enhancement of paths between UEs and the respective P-GWs, and for the minimization of the SGW relocations and the delay overheads. In contrast to these works, our proposed solution takes into account the geographic distribution of variant ECs, as well as the spatio-temporal distribution of users' requests, which reflects where, when, for how long, and how much data were used by UEs.

From another side, the conformal mapping, precisely the Schwartz-Christoffel mapping, was proposed in [27] to cope with the non-uniform distribution of nodes in cellular networks, by taking the analysis from the physical domain to the canonical domain, with the objective to make the best placement choices of access points.

It was also used in [8] for routing in a mobile network and to avoid local minimum solutions by encoding the map of the network domain to virtual coordinates of the canonical domain. To the best knowledge of the authors, the work presented in this paper is the first to apply canonical domains in EC environments. The proposed solution copes with the non-uniform distribution of signaling messages; an issue highly overlooked in all previous relevant research work. Connecting the physical domain to the canonical domain shall help to deal with the irregularities in terms of the compatibility between capacity provisioning and service demand by defining a uniform distribution that makes it easier to dimension and accordingly plan the resource distribution, as the work in the canonical domain is easier and more efficient than the analysis of the irregular spatio-temporal model of users' requests.

III. Problem Formulation

In the emerging environment whereby mobile services are mostly hosted in the cloud and demands for these services are constantly varying, the evaluation of resource management policies in the cloud is very challenging, mainly when it involves other parameters, such as the different configurations of physical resources, the different software stacks, users' profiles and their constantly changing preferences, and QoS requirements [2], [29]. To efficiently provision and manage such cloud resources, simulating and modeling the users' consumption of mobile services must overcome the non-uniform barrier, yet mostly neglected. Our paper proposes an optimal placement of virtual resources, under the constraint of predefined EC positions, to serve the end users' demands for services, expressed by the generated signaling messages. As will be detailed later, this placement is essentially based on canonical domains obtained using the Schwartz-Christoffel conformal mapping.

A. Problem Description

The objective of this paper is to instantiate VMs in optimal positions for serving the variant users while the QoE is ensured and the cost is minimized. We assume that the network consists of a set of ECs, whereby each one is supposedly, positioned in an optimal place (i.e., at the network edge) in closer proximity to where data are generated (Edge computing paradigm). Let consider the usage of a vertical in the envisioned area translated by the generation of signaling messages from UEs. We can summarize the target problem through the following question: How many VMs and in which ECs shall one instantiate them in order to optimally cope with the generated signaling messages? Here, VMs shall be instantiated in a cost-efficient manner while ensuring QoS, and reducing both the end-to-end latency and the energy consumption. These requirements are difficult as the service deployment has a complicated non-uniform distribution [18], [19]. It is therefore difficult to directly decide where to position VMs. To simplify

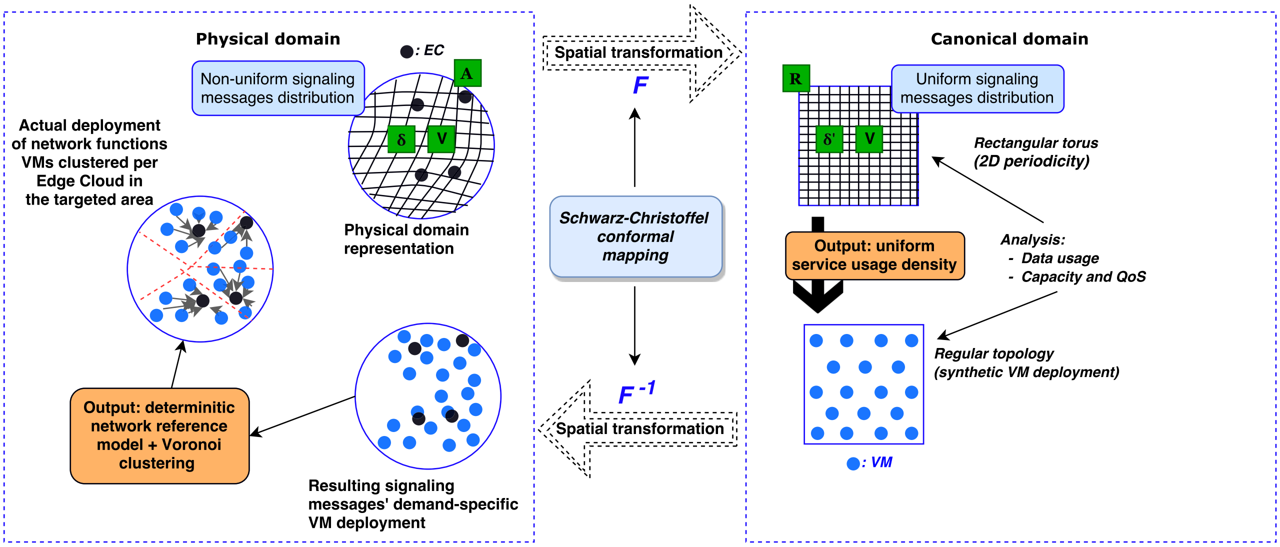


Fig. 1. The envisioned CDSP-F framework.

the problem, the conformal mapping technique can be used to obtain a uniform distribution of signaling messages. This will be detailed in Section IV.

B. Service Usage

The deployment of the services in our system is given as follows:

- Since sessions' lengths are highly variable, users are assumed to connect to each service following a Zipf distribution, which is of the form: $\beta e^{-\alpha}$ where β is the scaling factor and α denotes a constant exponent.
- The inter-arrival time of the i^{th} and $(i+1)^{th}$ sessions is given as time series $a(i)$ defined by $t(i+1) - t(i)$. $a(i)$ is fitted to a log-normal distribution.
- The service demand volume V_{SD} is modeled in terms of inter-arrival and session times. It denotes the average number of users in the system. V_{SD} is therefore expressed by:

$$V_{SD} = \frac{E(sessiontime)}{E(interarrival)} = \frac{\frac{\beta}{\beta-1} max(x_t, x_0)}{e^{\frac{\mu+\sigma^2}{2}}} \tag{1}$$

where

- μ and σ are the mean and standard deviation of the variable natural logarithm, respectively.
- The mean for the log normal distribution is:

$$E(X) = e^{\frac{\mu+\sigma^2}{2}} \tag{2}$$

- $x_0 = \alpha^{\frac{1}{\beta}}$. Also, it is worth noting that the density function f_X, for the Zipf distribution, of X is such that, for every x:

$$f_X(x) = \frac{\alpha\beta}{x^{\beta+1}} \tag{3}$$

And that:

$$E(X; X > x) = \int_x^{\infty} t f_x(t) dt \tag{4}$$

IV. Canonical Domain Framework Core

In this section, we will define the main functions used for our canonical domain framework, namely the mapping transformation functions (i.e., transformation from the physical domain to the canonical domain and vice-versa), the service density distribution functions δ and δ' in the physical domain and the canonical domain, respectively, and the Voronoi diagram algorithm. In addition, we define the algorithm of the CDSP-F core function.

As illustrated in Fig. 1, within the CDSP-F Framework, services are distributed non-uniformly and ECs are positioned in predefined positions. The spatial transformation function F helps to map every point of the physical domain to a point in the canonical domain. The service distribution in the canonical domain is uniform and such a regular distribution makes it easier to decide where to instantiate VMs and place VNFs. The inverse spatial transformation F^{-1} helps to map the created VM/VNF positions into the physical domain, which are finally assigned to the ECs in place, resulting in an optimal virtual resource deployment and an optimal VNF placement.

A. The Spatial Transformation Functions of CDSP-F

A conformal map is the transformation of a complex valued function from one coordinate system to another. This is accomplished by means of a transformation function that is applied to the original complex function. The original complex function defines the physical domain A (i.e., non-uniform distribution) and the obtained new function (using the given transformation) defines the canonical domain R (i.e., uniform distribution). The canonical domain which mimics this uniform distribution is depicted in Fig. 2. The idea behind this paper is to translate the non-uniform service

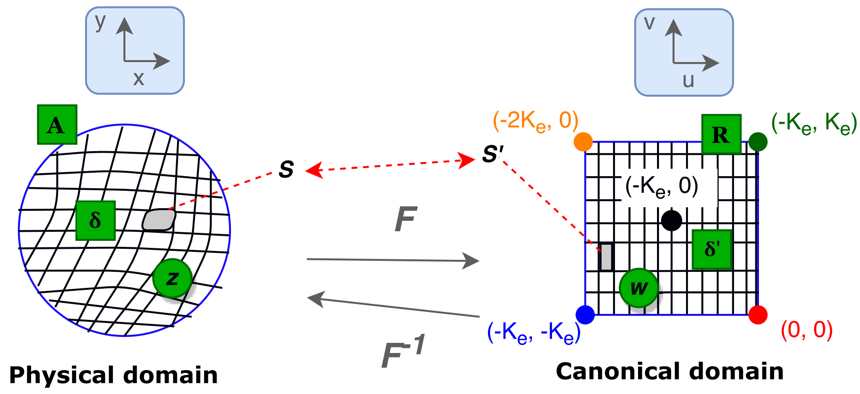


Fig. 2. The distribution of signaling messages in the physical and canonical domains.

usage/demand (physical domain) distribution using the transformation function of the Schwartz-christoffel conformal mapping into a canonical domain, whereby the service demand distribution is uniform. Let (x, y) and (u, v) denote the physical domain and the canonical domain, respectively. The idea, herein, is when we get the uniform service demand distribution, we uniformly deploy the variant VMs that satisfy these demands at the required QoE and in the most cost-efficient way. This deployment of variant VMs in a uniform fashion becomes a straight forward process. Then, using the inverse function of Schwartz-christoffel, we can get the real positions of those VMs (x, y) in the physical domain from their original positions in the canonical domain (u, v).

The signaling messages are generated within a circle which has a radius R while for the sake of efficiency as stated in [27], the targeted canonical domain is mapped as a rectangle divided into rectangular tiles. The mapping function that maps every point (u, v) in the circle to a point (x, y) in the square region and the inverse function must be found. This function must have derivative equations such that $(u, v) = F(x, y)$ and $(x, y) = F^{-1}(u, v)$. In other words, F will map each point (u, v) to point (x, y).

We model the service usage distribution in the simulation area as a rectangular grid inside the circle (the radius of the circle is equivalent to the radius of service deployment R (Fig. 2). The general form of the mapping function F is given as:

$$F(x + iy) = u(x, y) + iv(x, y) \tag{5}$$

The choice of the analytic components u and v reflects our spatial service demand distribution, which is graphically fitted to a rectangular grid within a circle (Fig. 1). x and y are obtained simply by applying the inverse function, reflected by the incomplete elliptic integral of the first kind E_f. Thus, using [21], we can write u, v, x and y as follows:

$$u = Re(\frac{1-i}{\sqrt{2}} cn(K_e \frac{1+i}{2}(x+iy) - K_e, \frac{1}{\sqrt{2}})) \tag{6}$$

$$v = Im(\frac{1-i}{\sqrt{2}} cn(K_e \frac{1+i}{2}(x+iy) - K_e, \frac{1}{\sqrt{2}})) \tag{7}$$

$$x = Re(\frac{1-i}{-K_e} E(cos^{-1}(\frac{1+i}{\sqrt{2}}(u+iv)), \frac{1}{\sqrt{2}})) + 1 \tag{8}$$

$$y = Im(\frac{1-i}{-K_e} E(cos^{-1}(\frac{1+i}{\sqrt{2}}(u+iv)), \frac{1}{\sqrt{2}})) - 1 \tag{9}$$

where

- K_e is a constant which depends on the radius R. Its exact value is the complete Legendre elliptic integral of the 1^{st} kind with modulus m. This constant arises from Schwarz and Christoffel's equations for the specific case when the desired polygonal shape is a square.
- cn is a Jacobi elliptic function.
- E_f is the incomplete elliptic integral of the first kind:

$$E_f(z|m) = \int_0^z \frac{1}{\sqrt{1 - msin^2(t)}} dt \tag{10}$$

The use of E_f is not only motivated by the fact that the incomplete elliptic integral of the first kind was extensively studied for the development of the theory of the double periodic functions, called elliptic functions, but most importantly: $i)$ $E_f(z|m)$ is an analytical function of z and m which are defined over $\mathbb{C}^2$ which allows its use for conformal mapping (i.e., it preserves angles throughout the whole mapping) and $ii)$ through the inverse Jacobi elliptic functions we have an interesting property:

$$E_f(cos^{-1}(z)|m) = cn^{-1}(z|m) \tag{11}$$

where $-1 < z < 1 \wedge m \in \mathbb{R}$.

Now that the mapping function F is defined, we can obtain the positions of each signaling message in the canonical domain based on their positions in the physical domain and use its reverse to obtain the positions of VMs in the physical domain based on their positions in the canonical domain. The question is about how we can link the service distribution δ to the mapping function F (physical domain) in order to be able to deduce the uniform service distribution δ' obtained using F^{-1}.

B. Service Density Distribution

In this subsection, we determine the service density distribution δ' that will enable the positioning of VMs in the canonical domain. We define the service density distribution in the canonical domain as follows.

Considering the physical concept of mass density, an object is dense if it has an important amount of mass for it is relatively small. Similarly, we can say that a region is dense if it has a large amount of signaling messages for it is relatively small. We can start by writing that:

$$\ddot{\delta} = \frac{|S_m|}{\ddot{V}} \tag{12}$$

where $\ddot{\delta}$ is the density (i.e the ratio of the mean of signaling messages and spatial volume $\ddot{V}$) and $|S_m|$ is the mean of signaling messages. We can go further and suppose that the region is divided into smaller and smaller areas and consider the density of each area. Let us think of it as a continuum which is the case of the physical domain we are considering. We can take the process to the limit where we are finding the density at each point in the area. This will result in the following:

$$\ddot{\delta}(\vec{V}) = \lim_{\ddot{V} \to 0} (\frac{|S_m|}{\ddot{V}}) \tag{13}$$

where the limit reflects the shrinking of the spatial volume under consideration. The reverse process of calculating the mean of signaling messages of a given small area a is an integral over the spatial volume:

$$|S_m|_a = \iiint_V \ddot{\delta}(\vec{V}) d\ddot{V} \tag{14}$$

Using the probability density, we replace "mean of signaling messages" with "probability" and spatial volume $\ddot{V}$ with volume V in the parameter space considering that our signaling messages follow the log normal distribution. This will result in the following:

$$\begin{aligned} \delta &= \frac{d}{dx} Pr(X \leq x) \\ &= \frac{d}{dx} Pr(ln(X) \leq ln(x)) \\ &= \rho(\frac{ln(x) - S_m}{V}) \frac{d}{dx} (\frac{ln(x) - S_m}{V}) \\ &= \rho(\frac{ln(x) - S_m}{V}) \frac{1}{S_m x} \\ &= \frac{1}{x} \frac{1}{S_m \sqrt{2\pi}} e^{\frac{-(log(x)-V)^2}{2S_m^2}} \end{aligned} \tag{15}$$

Note that S_m and V depend on the inputs of the simulation (e.g., number of signaling messages, inter-arrival of sessions, and average duration of a session). The analysis in the canonical domain aims at finding the number of VMs that is required to satisfy the uniformly distributed service demand with volume V along with their locations. We chose the rectangular tiling for the sake of efficiency. As the service demand is uniformly distributed in R, the uniform density can be written as follows [27]:

$$\delta' = \frac{L}{H.W} \tag{16}$$

where W and H denote the width and height of the rectangle, respectively. L denotes the number of VM clusters. It can be obtained as follows:

$$L = \frac{2WH}{(\frac{W^2}{4} + \frac{H^2}{4})3\sqrt{3}} \tag{17}$$

To distribute the VMs uniformly within the rectangular tiles, taking into consideration the data usage generated in the canonical domain, we denote by L_{Argmin} the effective number of VMs needed as follows:

$$L_{Argmin}(x) = Argmin(\frac{x}{VMcapacity}) \tag{18}$$

where x is the variable data usage of given signaling messages.

To illustrate this with more clarity, we consider the following example. We assume that three services are used in three different places of the canonical domain; namely Position 1, Position 2, and Position 3, whereby Service 1 generates 600Mb of traffic, Service 2 generates 300Mb of traffic, and Service 3 generates 400Mb of traffic. We assume that each VM can process up to 100Mb. Thus, using Equation (18), $L_{Argmin}(1) = 6$, $L_{Argmin}(2) = 3$, and $L_{Argmin}(3) = 4$. This corresponds to 6 VMs to process Service 1, 3 VMs for Service 2 and 4 VMs for Service 3. These VMs are then positioned on the tiles which are the closest to service generation (i.e., Position 1, Position 2, and Position 3).

C. VM Clustering Using the Voronoi Diagram

For the finite set of VM positions $p = \{p_1, p_2, \ldots, p_n\}$ in the physical domain, where each position p_k is considered as a point with latitude and longitude coordinates. A corresponding Voronoi EC position ϖ_k consists of every point in the plane of which the distance to p_k is less than or equal to its distance to any other p_i. Each such EC position is obtained from the intersection of half-spaces. The line segments of the Voronoi diagram are all the points in the plane that are equidistant to the two nearest sites. The Voronoi vertices (nodes) are the points equidistant to three (or more) sites (Fig. 3). Formally, given n EC sites $\{\varpi_1, \varpi_2, \ldots, \varpi_n\}$ in a distance space (X, H), the Voronoi partition X into regions $\{vo_1, vo_2, \ldots, vo_n\}$ for VM positions $p = p_1, \ldots, p_n$ is carried out, based on the Haversine distance function H, such that:

$$\forall p_i \in p : vo_k = \{x \in X | H(x, \varpi_i) < H(x, \varpi_j), i \neq j\} \tag{19}$$

$$H(x, \varpi_i) = r \times c \tag{20}$$

$$a = sin^2(\frac{\Delta\phi}{2}) + cos\phi_1 \times cos\phi_2 \times sin^2(\frac{\Delta\alpha}{2}) \tag{21}$$

$$c = 2 \times atan2(\sqrt{a}, \sqrt{1-a}) \tag{22}$$

where

- r is the Earth's radius (mean radius = 6,371km).
- c is the angular distance in radians.
- a is the square of half the chord length between the points x and ϖ_i.

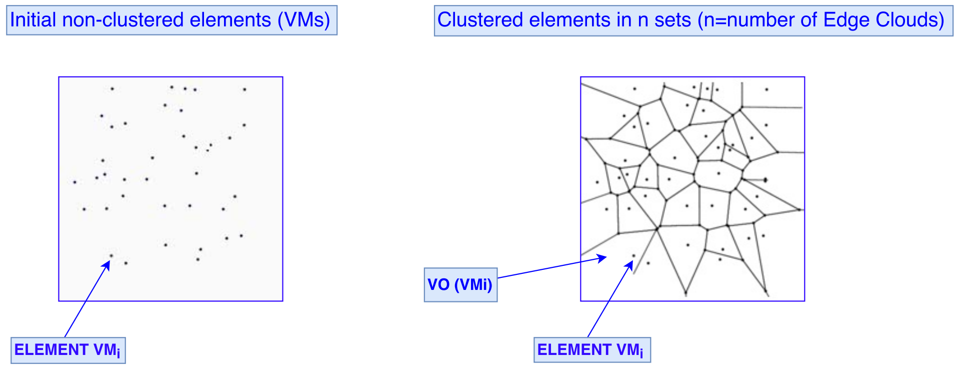


Fig. 3. Illustration of the clustering of VMs using the Voronoi diagram.

- ϕ_1, α_1 are the latitude and longitude values, in radians, of x, respectively.
- ϕ_2, α_2 are the latitude and longitude values, in radians, of ϖ_i, respectively.

Note that Equations (20), (21), and (22) concern the calculations of the Haversine method, which is an approach to calculate the great-circle distance between two points on the basis of a spherical Earth (ignoring the ellipsoidal effects as the Earth is very slightly ellipsoidal) [1]. This spherical model gives errors, typically up to 0.3% which we can consider negligible, and hence the calculations can be assumed accurate.

D. Algorithm of the Core Function of CDSP-F

As defined in Algorithm 1, the first step consists of collecting the positions where the data were generated in the current snapshot of time. In the second step, using mapping functions (6) and (7), which map every point of the physical domain to its corresponding one in the canonical domain, we obtain the positions of signaling messages (i.e., data) in the canonical domain. In the third step, due to the uniform nature of the canonical domain, we divide the canonical domain in rectangle tiles using Equations (8) and (9), which map every point of the canonical domain to its corresponding one in the physical domain, and using the argmin function (18), which determines how much resources are needed in the given rectangle tile. This results in positioning the VMs that could fulfill the requests generated in each tile of the canonical domain. In the fourth step, we apply equations (8) and (9) to obtain the positions of generated VMs in the physical domain. In the last step, using the Voronoi diagram (See Equation (19) which determines how the VM clusters for ECs are defined, and Equations (20), (21), and (22) which measure the Haversine distance between VM positions in the physical domain for the Voronoi diagram), we allocate the needed VMs to their corresponding ECs in place.

Algorithm 1 CDSP-F Core Function

Require:
Γ: Signaling messages.
X: A set of x positions in the physical domain of signaling messages.
Y: A set of y positions in the physical domain of signaling messages.
X_e: A set of x positions in the physical domain of edge clouds.
Y_e: A set of y positions in the physical domain of edge clouds.

Ensure:
U: The set of signaling messages' u positions in the canonical domain obtained by F
V: The set of signaling messages' v positions in the canonical domain obtained by F
X_{VM}: The set of VMs' x positions in the physical domain obtained by F^{-1}
Y_{VM}: The set of VMs' y positions in the physical domain obtained by F^{-1}

```
for all Γ_i ∈ Γ do
    (U_i, V_i) ← F(X_i, Y_i); {Calculate position(u,v) in the canonical domain}
    ComputeNeededResources(Γ_i);
    if NeededResources_{Γ_i} > VMCapacity(U_i, V_i) then
        CreateNewVM(NeededResources_{Γ_i}, U_i, V_i) ;
    else
        UpdateCurrentVM(NeededResources_{Γ_i}) ;
    end if
end for
(X_i, Y_i) ← F^{-1}(U_i, V_i) {Calculate position(x,y) in the physical domain}
Cluster VMs of the Physical Domain using Voronoi diagram
𝒜ℒs ← VO(X_e, Y_e, X, Y);
return 𝒜ℒs;
```

V. Experimentation and Results

In this section, we evaluate the performance of the proposed CDSP-F solution and compare it against the performance of existing base-line approaches, which are i) an adaptation of the Best Fit Algorithm (BFA) [3], [13] and ii) the Random Fit Algorithm (RFA) [28]. The variant solutions are evaluated in terms of the following metrics:

- End-to-End delay: This metric is defined as the end-to-end delay between variant UEs and the VMs handling their requests. Formally, the end-to-end delay is computed as the maximum Euclidean distance between the UEs and their respective VMs;
- Cost: This metric is defined as the cost for handling the variant traffic generated from variant UEs. In this paper, we considered the cost of a solution as the total number of VMs instantiated in the cloud by each solution;
- Execution time: This metric is defined as the time needed to execute each Algorithm. For each solution, we practically subtract its finishing time from its starting time.

The variant algorithms are evaluated by using our Java-based simulator that defines a spatio-temporal model of mobile service usage over a particular geographical area.[1] It enables to simulate the behavior of a group of mobile users, in terms of mobility patterns and mobile service consumption. The output of the tool is the number of handoff operations, tracking area updates, and service requests issued over a specific geographical area during a specific time window. Knowing these values and based on the performance of VNFs of Mobile Networks (e.g., Mobility Management Entity and Serving Gateway) when running over specific virtual resources (e.g., CPU and memory), one can optimally decide where to instantiate the VMs over the edge cloud and how much virtual resources to use.

All algorithms are executed in the same environment using Ubuntu 16.04 on an Intel Core i5 2 core CPU and 8 GB of RAM. The algorithms are evaluated by varying the radius of deployment area and the number of signaling messages. We conducted two sets of experiments: i) First, we fixed the radius of the target area to 20 km; ii) Second, we fixed the radius area to 40 km. In both experiments, we varied the number of signaling messages from 0 to 2600.

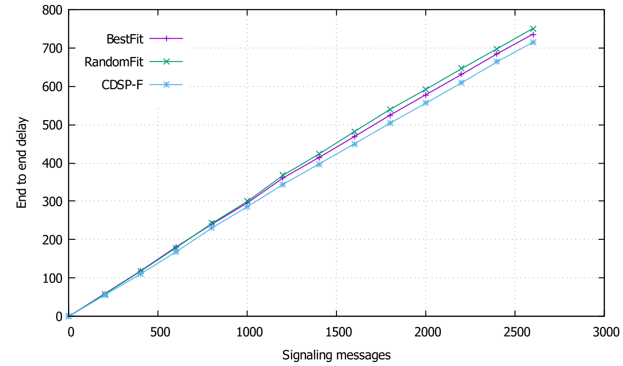
A. End-to-End Delay

In this subsection, we show the impact of signaling messages on the end-to-end delay. Figs. 4 show the performance of variant solutions in terms of delay. Fig. 4(a) shows the end-to-end delay for the network area of 20 km radius, whereas Fig. 4(b) shows the end-to-end delay for the network area of 40 km radius. The first observation that we can draw from these figures is that the number of signaling messages has a negative impact on the end-to-end delay. This can be explained as follows: The higher the number of signaling messages in the network is, the higher the possibility for UEs to be far away from their VMs is, thus the longer the end-to-end delay gets. From these figures, we also observe that CDSP-F exhibits better performance in terms of end-to-end delay.

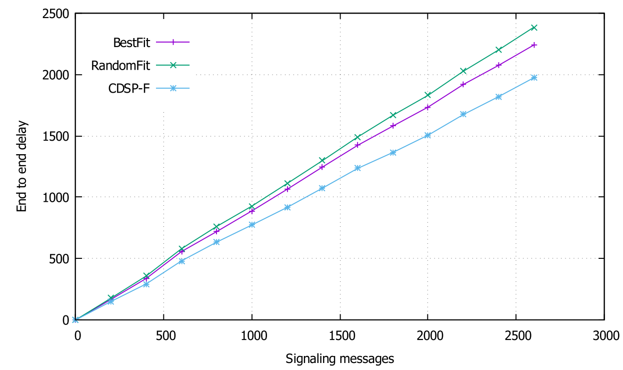
From Fig. 4(b), the end-to-end delay of CDSP-F does not exceed 2000, while the other base-line solutions exceed 2300.

When the number of signaling messages is low (inferior to 300) the gap in terms of end-to-end delay is relatively low for the three algorithms. As the number of signaling messages increases, the analysis in the canonical domain gives better

[1]Network Slice Planner: http://mosaic-lab.org/implementations.aspx



(a) 20 km radius



(b) 40 km radius

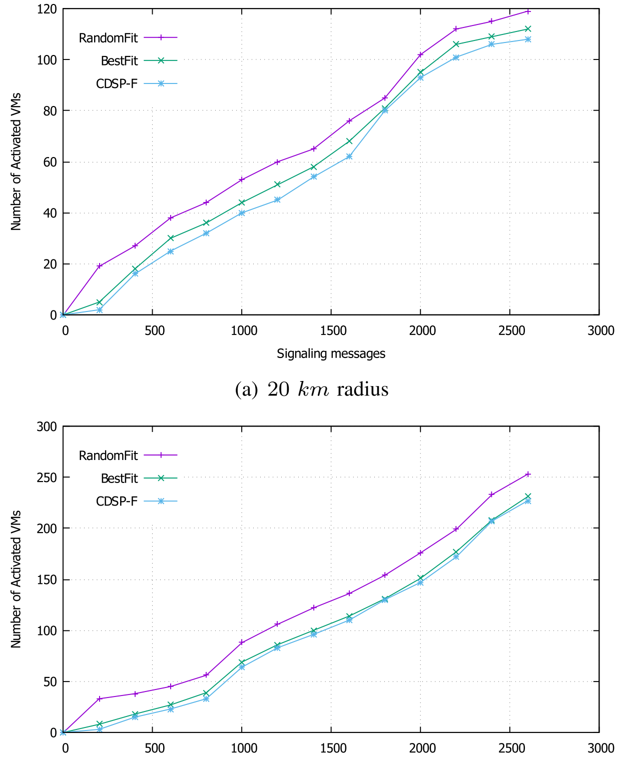
Fig. 4. The performance evaluation of variant solutions in terms of end-to-end delay.

positions of VMs in terms of distance to where the signaling messages are triggered, which encourages their usage by the neighboring UEs and impacts positively the total number of created resources without an impinge on the delay. Moreover, as the CDSP-F scheme continues to give satisfying results, the gap compared to BFA and RFA becomes more important, when the number of signaling messages is high and the deployment area is bigger (i.e., from a 20 km radius area to a 40 km radius area).

B. Number of Activated VMs

In this subsection, we evaluate the variant solutions in terms of cost, defined as the total number of VMs which should be instantiated across the available ECs. Fig. 5 shows the performance of the different solutions in terms of cost (number of activated VMs). Fig. 5(a) shows the cost for the network area of 20 km radius, whereas Fig. 5(b) shows the cost for the network area of 40 km radius.

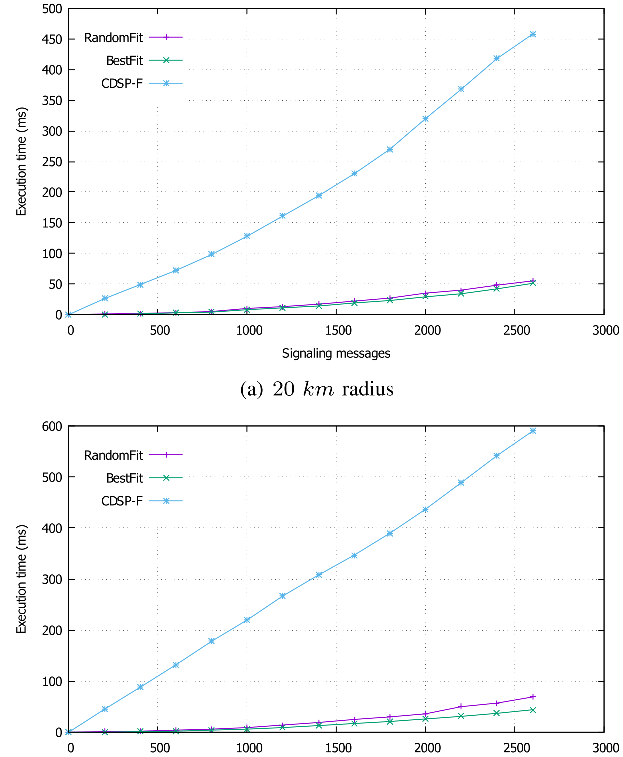
The first observation that we can draw from these figures is that the number of signaling messages has a negative impact on the cost, due to the fact that more virtual resources become needed when the end-users generate higher number of signaling messages. The performance gap related to the number of activated VMs is very important when the radius of service deployment area increases; this can be explained by the need to create resources in the proximity of end-users that generate these signaling messages. These resources are randomly generated by the RFA scheme, which is translated



(a) 20 km radius

(b) 40 km radius

Fig. 5. The performance evaluation of variant solutions in terms of total number of activated VMs.



(a) 20 km radius

(b) 40 km radius

Fig. 6. The performance evaluation of variant solutions in terms of runtime execution.

into the continuously high number of activated VMs, while the BFA and CDSP-F schemes have similar numbers of activated VMs, with better results in favor of the CDSP-F scheme not exceeding 210 VMs in Fig. $5(b)$. On the one hand, the BFA scheme is known for seeking for VMs with the best available resources, which encourages the reutilization of existing resources. On the other hand, the analysis in the canonical domain has a positive impact on positioning VMs closer to UEs, without an impinge on the reutilization of available resources. The most interesting observation is that in comparison to BFA and RFA, the clustering of VMs based on their deployment in the canonical domain helps reducing the number of VMs without negatively impacting the end-to-end delay, especially in case a high number of signaling messages is generated across a wider area.

C. Execution Time

Fig. 6 shows the impact of the number of signaling messages on the execution time of each individual solution. While Fig. $6(a)$ shows the performance of the algorithms in the area with radius 20 km, Fig. $6(a)$ shows their performance in the area with radius 40 km. The first observation that we can draw from these figures is that the number of signaling messages has a negative impact on the execution time of each solution. From the obtained results, we observe that the execution time of the proposed CDSP-F solution requires more time than the ones of the base-line approaches. This is due to the complex computation needed by CDSP-F. Moreover, the gap in execution time between CDSP-F and the base-line approaches sharply increases along with growth in the number of signaling messages. For example, when the number of signaling messages reaches 2600, the execution time of the base-line approaches does not exceed 100 ms, while it reaches 600 ms in case of CDSP-F. From Figs. 4, 5 and 6, we observe that CDSP-F outperforms the base-line approaches in terms of end-to-end delay and the total number of activated VMs. However, this comes with an inevitable cost, which is in terms of the execution time. Based on the observation that these solutions would be executed off-line and only once, and the new generation of servers have high computational capacities, the execution time factor can be neglected.

VI. Conclusion

Due to the dynamic and constantly changing characteristics of mobile edge clouds, we proposed in this paper the analysis of service usage and VM deployment in the canonical domain which is obtained by means of Schwartz-Christoffel conformal mappings. The transformation functions of such a mapping helps obtaining a uniform distribution of service consumption over a specific geographical area, thus ushering efficient ways to quantify and define the needed virtual resources in MECs to meet the requirements of services consumed by UEs. Our initial modeling gave satisfying results and showed the interesting features of the usage of conformal mapping, as it enhanced the end-to-end delay and reduced the total number of activate VMs, and that is in comparison to the best fit and the random fit algorithms. In an environment characterized by redundant patterns, we intend to propose an improved version of CDSP-F, which iterates the steps mentioned in Section IV-D on the data generated in relevant snapshots of time (i.e., peak hours, weekends, holidays) and extracts patterns that will allow the prediction of optimal VM positions, composing a knowledge-based prediction database. The latter will be used to answer to the following question: Based on previous placement decisions and in scenarios with similar patterns, how to predict near-optimal VNF positions? In addition to that, we intend to go further and investigate how to personalize and optimize the modeling of canonical domains for specific VNFs (e.g., mobility anchoring gateways such as S-GW in EPC and data anchoring gateways such as P-GW in EPC).

REFERENCES

[1] C. C. Robusto, "The cosine-haversine formula," *Amer. Math. Monthly*, vol. 64, no. 1, pp. 38–40, 1957.

[2] R. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, "CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," *Softw., Pract. Exper.*, vol. 41, no. 1, pp. 23–50, 2011.

[3] M. Guazzone, C. Anglano, and M. Canonico, "Exploiting VM migration for the automated power and performance management of green cloud computing systems," in *Proc. 1st Int. Workshop Energy Efficient Data Centers*, 2012, pp. 81–92.

[4] A. Basta, W. Kellerer, M. Hoffmann, K. Hoffmann, and E.-D. Schmidt, "A virtual SDN-enabled LTE EPC architecture: A case study for S-/P-gateways functions," in *Proc. IEEE SDN Future Netw. Services (SDN4FNS)*, Trento, Italy, Nov. 2013, pp. 1–7.

[5] K. Kirkpatrick, "Software-defined networking," *Commun. ACM*, vol. 56, no. 9, Sep. 2013, pp. 16–19.

[6] H. Kim and N. Feamster, "Improving network management with software defined networking," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 114–119, Feb. 2013.

[7] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless video content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 1107–1115.

[8] S. Li, W. Zeng, D. Zhou, D. X. Gu, and J. Gao, "Compact conformal map for greedy routing in wireless mobile sensor networks," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 1409–1417.

[9] T. Taleb, "Toward carrier cloud: Potential, challenges, and solutions," *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 80–91, Jun. 2014.

[10] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[11] H. Ahlehagh and S. Dey, "Video-aware scheduling and caching in the radio access network," *IEEE/ACM Trans. Netw.*, vol. 22, no. 5, pp. 1444–1462, Oct. 2014.

[12] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[13] L. Tomás and J. Tordsson, "An autonomic approach to risk-aware data center overbooking," *IEEE Trans. Cloud Comput.*, vol. 2, no. 3, pp. 292–305, Jul. 2014.

[14] M. Bagaa, T. Taleb, and A. Ksentini, "Service-aware network function placement for efficient traffic handling in carrier cloud," in *Proc. IEEE WCNC*, Istanbul, Turkey, Apr. 2014, pp. 2402–2407.

[15] P. Farrow, M. Reed, M. Glowiak, and J. Mambretti, "Transcoder migration for real time video streaming systems," *CoRR*, vol. abs/1509.08091, pp. 1–13, Sep. 2015.

[16] F. Z. Yousaf, P. Loreiro, F. Zdarsky, T. Taleb, and M. Leibsch, "Cost analysis of initial deployment strategies of a virtual network infrastructure in a datacenter," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 60–66, Dec. 2015.

[17] T. Taleb, M. Bagaa, and A. Ksentini, "User mobility-aware virtual network function placement for virtual 5G network infrastructure," in *Proc. IEEE ICC*, London, U.K., Jun. 2015, pp. 3879–3884.

[18] J. Cichon, M. Gebala, and M. Klonowski, "Ad-hoc, mobile, and wireless networks," in *Proc. ADHOC-NOW*, Wroclaw, Poland, Jul. 2013, pp. 172–184.

[19] A. H. Labordere, *Virtual Roaming Data Services and Seamless Technology Change: GSM, LTE, WiFi, Satellite, CDMA* (River Publishers Series in Communications). Gistrup, Denmark: River Publishers, Apr. 2014.

[20] T. Taleb *et al.*, "EASE: EPC as a service to ease mobile core network deployment over cloud," *IEEE Netw. Mag.*, vol. 29, no. 2, pp. 78–88, Mar. 2015.

[21] C. Fong, "Analytical methods for squaring the disc," *Seoul ICM*, pp. 1–40, Sep. 2015.

[22] *Study on Architecture for Next Generation System*, document 3GPP TR 23.799, Dec. 2016.

[23] S. Dutta, T. Taleb, P. A. Frangoudis, and A. Ksentini, "On-the-fly QoE-aware transcoding in the mobile edge," in *Proc. IEEE Globecom*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[24] A. Laghrissi, T. Taleb, and M. Bagaa, "Canonical domains for optimal network slice planning," in *Proc. IEEE (WCNC)*, Barcelona, Spain, Apr. 2018.

[25] S. Dutta, T. Taleb, and A. Ksentini, "QoE-aware elasticity support in cloud-native 5G systems," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[26] A. Gharaibeh, A. Khreishah, B. Ji, and M. Ayyash, "A provably efficient online collaborative caching algorithm for multicell-coordinated systems," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 1863–1876, Aug. 2016.

[27] D. G. González and J. Hämäläinen, "Looking at cellular networks through canonical domains and conformal mapping," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3703–3717, May 2016.

[28] A. Choudhary, S. Rana, and K. Matahai, "A critical analysis of energy efficient virtual machine placement techniques and its optimization in a cloud computing environment," *Procedia Comput. Sci.*, vol. 78, pp. 132–138, Jan. 2016.

[29] F. Z. Yousaf and T. Taleb, "Fine-grained resource-aware virtual network function management for 5G carrier cloud," *IEEE Netw. Mag.*, vol. 30, no. 2, pp. 110–115, Mar. 2016.

[30] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657–1681, 3rd Quart., 2017.

[31] Y. Jararweh *et al.*, "Software-defined system support for enabling ubiquitous mobile edge computing," *Comput. J.*, vol. 60, no. 10, pp. 1443–1457, Oct. 2017.

[32] T. Taleb, B. Mada, M. Corici, A. Nakao, and H. Flinck, "PERMIT: Network slicing for personalized 5G mobile telecommunications," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 88–93, May 2017.

[33] S. Retal, M. Bagaa, T. Taleb, and H. Flinck, "Content delivery network slicing: QoE and cost awareness," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.

Abdelquoddouss Laghrissi received the bachelor's degree in mathematics and computer science and the master's degree in applied computer science with a dissertation on empathy in vehicular ad hoc networks from the School of Sciences, Mohamed V. University, in 2012 and 2014, respectively. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering, Aalto University, Finland. From 2014 to 2015, he was a Volunteer Member of the cloud computing working group within a Euro-Mediterranean project MOSAIC on Cooperation with Mediterranean Partners to build Opportunities around ICT and Societal and Industrial Challenges of H2020, and is currently involved in a European project 5G!Pagoda on network slicing and 5G in the context of H2020, during which he published articles and contributed to several deliverables. He is a member of the MOSA!C Lab. His research interests include network softwarization and slicing mechanisms, mobile cloud computing, network function virtualization, software defined networking, and vehicular ad hoc networks.

Tarik Taleb received the B.E. degree (Hons.) in information engineering and the M.Sc. and Ph.D. degrees in information sciences from Tohoku University, in 2001, 2003, and 2005, respectively. Prior to his current academic position, he was a Senior Researcher and a 3GPP Standards Expert with NEC Europe Ltd., Heidelberg, Germany. Before joining NEC and in 2009, he was an Assistant Professor with the Graduate School of Information Sciences, Tohoku University, Japan, in a laboratory fully funded by KDDI. From 2005 to 2006, he was a Research Fellow with the Intelligent Cosmos Research Institute, Sendai, Japan. He is currently a Professor with the School of Electrical Engineering, Aalto University, Finland. He is also the Founder and the Director of the MOSA!C Lab.

His research interests lie in the field of architectural enhancements to mobile core networks, network softwarization and slicing, mobile cloud networking, network function virtualization, software defined networking, mobile multimedia streaming, inter-vehicular communications, and social media networking. He has been also directly engaged in the development and standardization of the Evolved Packet System as a member of 3GPP's System Architecture working group. He is a member of the IEEE Communications Society Standardization Program Development Board. As an attempt to bridge the gap between academia and industry, he founded the IEEE Workshop on Telecommunications Standards: from Research to Standards, a successful event that got awarded best workshop award by the IEEE Communication Society (ComSoC). Based on the success of this workshop, he has also founded and has been the Steering Committee Chair of the IEEE Conference on Standards for Communications and Networking.

Dr. Taleb is the General Chair of the 2019 edition of the IEEE Wireless Communications and Networking Conference to be held in Marrakech, Morocco. In 2016, he served as a Chair of the Wireless Communications Technical Committee, the largest in IEEE ComSoC. He also served as a Vice Chair of the Satellite and Space Communications Technical Committee of the IEEE ComSoc from 2006 to 2010. He has been on the technical program committee of different IEEE conferences, including Globecom, ICC, and WCNC, and chaired some of their symposia. He is the Guest Editor-in-Chief of the IEEE JSAC Series on Network Softwarization and Enablers. He is/was on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE *Wireless Communications Magazine*, the IEEE JOURNAL ON INTERNET OF THINGS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and a number of Wiley journals.

Dr. Taleb was a recipient of the 2017 IEEE ComSoc Communications Software Technical Achievement Award in 2017 for his outstanding contributions to network softwarization. He is also the co-recipient of the 2017 IEEE Communications Society Fred W. Ellersick Prize in 2017, the 2009 IEEE ComSoc Asia-Pacific Best Young Researcher award in 2009, the 2008 TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2008, the 2007 Funai Foundation Science Promotion Award in 2007, the 2006 IEEE Computer Society Japan Chapter Young Author Award in 2006, the Niwa Yasujirou Memorial Award in 2005, and the Young Researcher's Encouragement Award from the Japan chapter of the IEEE Vehicular Technology Society in 2003. Some of his research work have been also awarded best paper awards at prestigious IEEE-flagged conferences.

Miloud Bagaa received the Engineer's, master's, and Ph.D. degrees from the University of Science and Technology Houari Boumediene, Algiers, Algeria, in 2005, 2008, and 2014, respectively. From 2009 to 2015, he was a Researcher with the Research Center on Scientific and Technical Information, Algiers, where he was a member of the Wireless Sensor Networks team, DTISI Division. From 2015 to 2016, he was granted a Postdoctoral Fellowship from the European Research Consortium for Informatics and Mathematics, and was with the Norwegian University of Science and Technology, Trondheim, Norway. He is currently a Senior Researcher with the Communications and Networking Department, Aalto University. His research interests include wireless sensor networks, the Internet of Things, 5G wireless communication, security, and networking modeling.