REPEL: A Strategic Approach for Defending 5G Control Plane from DDoS Signalling Attacks

Renato S. Silva, Carlos C. Meixner, Rafael S. Guimarães, Thierno Diallo, Borja O. Garcia, Luís F. M. de Moraes and Magnos Martinello

***Abstract*—5G relies on its pervasive and convergent cloud-based architecture to accomplish its futuristic challenge of being the next-generation communication platform. However, the new perspectives opened by 5G networks do not go unnoticed. Regardless of their motivation or objectives, cyberattackers find in the new 5G ecosystem, including its tenancy-driven control plane, an attractive greenfield to create new types of denial of services attacks. In this paper, we leverage on the virtualised environment of 5G to propose REPEL – an intelligent resource scaling strategy to mitigate DDoS signalling attacks preserving legitimate traffic. Our prevention-based approach uses games theory to build up a defence front line, able to keep services availability and discourage the attacker. To demonstrate the effectiveness and feasibility of our approach, we feed a queuing model with parameters obtained from a testbed, where simulated subscribers connect to a virtualised evolved packet core prototype. The final results show a dramatic signalling losses reduction, which can ensure the appropriate control plane availability under a DDoS attack.**

***Index Terms*—5G, LTE, EPC, DDoS, NFV, control plane, signalling, cloud.**

I. Introduction

The increasing use of smartphones and many other types of network connected mobile devices applications represent an important gear to leverage Internet growth [1]. According to [2], nearly three-quarters of the world will use just their smartphones to access the Internet by 2025. Thanks to the technological development of devices, mobile applications are forecast to generate 164 EB (ExaBytes) data traffic per month in 2025 [3]. Indeed, these mobile devices have become crucial to people's lives, and their widespread adoption drives the development of several new groundbreaking applications every day [4]. However, this ever-growing diversity of bandwidth-greedy applications demands more and more from network to keep operating.

The 5G networks already emerge inheriting hard challenges such as the need for higher data rate and capacity, lower cost, and smaller end-to-end latency. Aiming to move from whiteboard towards the *de facto* next-generation communication platform, 5G depends on key groundbreaking technologies such as self-organising network (SON), multiple-input/multiple-output (MIMO) antennas technology, and small-cells hyper-densification [5], [6]. 5G control plane relies on its distributed and virtualised architecture to share network resources with different tenants [7] and make these new heterogeneous technologies work like an orchestra.

The Internet evolution and the development of new technologies have brought about a significant increase in the number of cyberattacks and crimes. Recent statistics presented in [8] show that the number of cybernetic attacks has increased in 50% since 2017. Regardless of their motivation or objectives, cybernetic attackers see the new 5G ecosystem, including powerful smartphones, high bandwidth, and massive software-based deployments as an attractive greenfield to create new types of attacks [9]. Even though it has improved several protection mechanisms from its predecessor 4G, the reinforced 5G security architecture still present open issues concerning signalling attacks against the 5G control plane [10].

The main pillars of the 5G security framework lie on the long term evolution (LTE) architecture, where the control flows from the core to the access network [11]. In other words, 5G architecture follows the same premise of 4G, in which user-equipment (UE) just replies control requests from the evolved packet core (EPC). Although it sounds like going back to the past, such a centralised control strategy still present many advantages comparing to the distributed approach. One benefit of the centralised approach lies at preserving the control plane integrity by standardising signalling domains and enabling control decisions related to mobility, scaling, and protection.

While the centralised approach of 5G architecture helps to prevent internal attacks against the control plane, the large number of signalling messages triggered to accomplish any procedure also enables amplification attacks [12]. As the border signalling gateway between the radio access network (RAN) and EPC, the mobility management entity (MME) plays a crucial role in filtering the signalling messages from the access network. However, despite being addressed in several papers [13]–[21], there are many open issues concerning the detection of signalling attacks, and how to mitigate them while also keeping the legitimate traffic preserved [22].

In this paper, we address the threatening scenario presented in [23], [24], in which an attacker remotely controls many smartphones already registered on the network as subscribers

The authors would also like to thank the editor and the anonymous reviewers, for their valuable comments and suggestions, which helped to improve the quality of the paper.

Renato S. Silva and Luís F. M. de Moraes are with the Federal University of Rio de Janeiro, RJ – Brazil. email: {renato, moraes}@ravel.ufrj.br.

Carlos C. Meixner, Thierno Diallo and Borja O. Garcia are with the University of Bristol, Bristol – UK, e-mail: {carlos.colmanmeixner, thierno.diallo, borja.garcia}@bristol.ac.uk.

Rafael S. Guimarães and Magnos Martinello are with the Federal University of Espiríto Santo, ES – Brazil. e-mail: rafaelg@ifes.edu.br and magnos@inf.ufes.br.

to start a signalling attack against 5G control plane. However, instead of applying hard blocking countermeasures, we propose a defence-in-depth strategy [25], [26], in which an insurance-based front-line mitigates the disruptive effects of signalling attacks without compromising legitimate traffic in the 5G vEPC. The main idea is to use extra allocated resources, which are not meant to be used under normal conditions, as an insurance plan to reduce risks of 5G control plane outages due to DDoS signalling attacks.

The key contributions of this paper can be summarised as:

- An insurance-based insight to prevent control-plane outages and potentially mitigating DDoS signalling attacks by scaling virtualised network resources in the cloud.
- The introduction of a game theory model that allows to find the equilibrium points (*i*) in order to determine the best time to scale-up the resources from a defence perspective, and (*ii*) to understand the attacker behaviour according to her/his incentive rate.
- The development of an experimental vEPC testbed, able to demonstrate how the LTE overload control performs to offload the vMME under a DDoS attack.
- The formulation of a queuing model to analytically reproduce a flooding scenario in the control plane, in order to extrapolate the testbed outcomes.

To evaluate the effectiveness and feasibility of our approach, we also bring forth a testbed composed by a virtualised evolved packet core (vEPC) prototype and some UE simulators. The results obtained from testbed show that managing the weighting factor of the virtualised mobile mobility entity (vMME) works as expected to lure the new incoming traffic from the radio access network (RAN). Further, balancing the signalling traffic by doubling the number of vMMEs reduces proportionally their memory usage without compromising legitimate traffic. The disruptive effect of a signalling attack is analytically evaluated by using a queuing model, in which the number of bots is extrapolated to project signalling losses.

The remainder of this paper is organised as follows. Section II provides basic information to make a comparative analysis among related works and to describe the problem addressed in this paper. Section III describes the proposed architecture and presents a modelling framework for performance evaluation. In Section V, we describe the experimental model and present a testbed to demonstrate its operational characteristics. Section VI leverages the modelling framework to extrapolate the numerical results from the testbed aiming to evaluate the performance of the approach. Finally, in Section VIII, we close the paper with a summary, emphasising its contributions and projecting some future works.

II. Background, Problem Description, and Related Previous Work

The 4G EPC is a flat architecture with a multi-access IP-based mobile core that provides the user with converged voice and real-time data on the long-term evolution (LTE) network. Figure 1 shows an EPC architecture composed by three separate entities: mobility management entity (MME), home subscriber sub-system (HSS), and serving packet data network gateway (SPGW)[1]. These entities interact with each other through standardised interfaces and protocols to accomplish signalling procedures. MME is in charge of controlling the signalling process between the RAN and the EPC, which involves mobility and security. HSS is a database that contains user-related and subscriber-related information. SPGW deals with both control and user planes to transport IP traffic between the UE and the external networks.

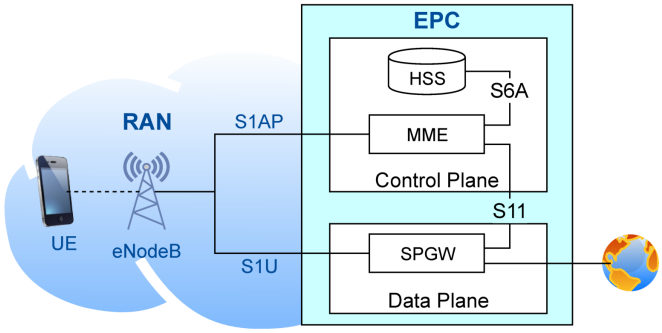


Fig. 1. 4G EPC typical architecture collapsing PGW and SGW into SPGW, according to [27]. The EPC entities, including radio access network (RAN), interact with each other through standardised interfaces.

Although the 5G next-generation core (NG-Core) has inherited from the 4G EPC the same set of stateful-based functions, its so-called service-based architecture (SBA) shown Figure 2 is organised into ten different service-oriented network functions. The goal of its denser architecture is to pave the way for network slicing by fully separating control plane from data plane. However, the need for the network functions to constantly interact with each other creates a critical environment, where any traffic congestion, for instance, due to a denial of service attack (DoS), can rapidly deteriorate or even disrupt control plane services [28], [29].

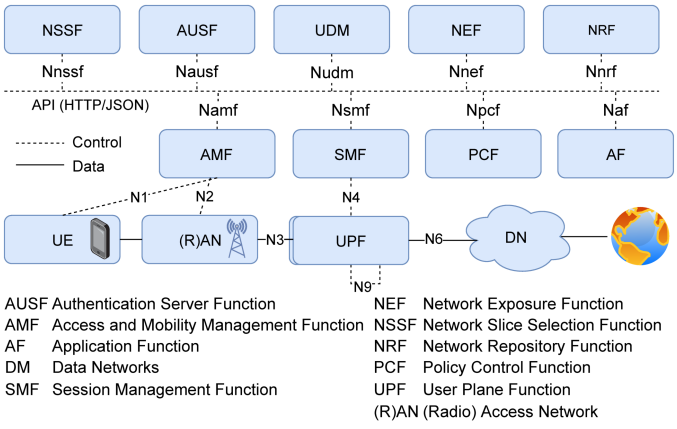


Fig. 2. Service-Based Architecture (SBA) of 5G Next Generation Core (NG-Core) [30].

The DoS attack threatens network availability. Unlike other kinds of attacks, the primary goal of a DoS attack is not to steal information but to slow or disrupt its target's services. Although DoS attacks are usually launched from a single

[1]Serving gateway (SGW) and packet gateway (PGW) collapsed into SPGW.

source, a distributed denial of service attack (DDoS) can be stealthily launched from a huge number of different sources at the same time. When it is carried out from internal sources, supposed to have subscriber rights, the effects of a DDoS attack tend to be more threatening since it can exploit further vulnerabilities to achieve disruptive goals.

In the same idea, a DDoS signalling attack aims to disrupt control plane services. Once the attacker has recruited enough number of bots, he just take advantage of any stateful process running in the control plane to amplify the attack. As practical examples of signalling attack vectors, one may consider the 3GPP *user attach* and user *handover* procedures [27], which trigger a large number of signalling messages in the control plane. If many compromised UEs, strategically distributed on multiple small-cells, are simultaneously controlled to repeatedly request for such procedures, the total signalling traffic can stealthily deplete the EPC performance as a whole [31].

Detecting a denial of service attack is very difficult, due to its mimetic user-based behaviour. Therefore, a defence approach that deliberately blocks signalling traffic may end up collaborating with the attacker. In this regard, although scaling-up up victim resources is not a novelty [32], [33], the flexibility and facility to manage abstracted resources brought by the virtualisation trend, provide new insights to improve defence strategies with new mitigation mechanisms.

Taking advantage of the virtualisation trend, SCALE [34] proposes a framework to flexibly and transparently scale-up and scale-down vMMEs to meet the performance requirements of IoT traffic. SCALE tackles the load balance and virtual machine (VM) footprint problems by splitting the MME architecture into two entities: MME processing (MMP) and MME load balance (MLB). Scaling-up MME as an alternative solution to manage signalling overload in 5G control plane is also the main motivation in [15]. In their approach, a comprehensive review of mobile network infrastructure is introduced to meet the QoS requirements of 5G by splitting the MME architecture into lightweight microservices, in which the UE states are held into a centralised data store.

The analysis presented in [13] tackles the problem of signalling vulnerabilities in 3G mobile network related to the significant amount of signalling traffic triggered by the radio resource control protocol (RRC) to negotiate and to establish logical channels between UE and EPC. The authors also compare three different defending solutions and suggest the target randomisation method as the most effective cost and overhead. Addressing a more specific issue, the authors in [14] assess a potential signalling-oriented denial of service attack against the EPC. The assumed scenario considers the high number of messages exchanged between the EPC and the UE to establish a dedicated bearer as an amplification attack vector. After establishing one or more dedicated bearers (eight at most), the attacker simply leaves them to be released by the EPC after twenty seconds (inactivity timer) and so the process is automatically restarted by the attacker. In the same paper, a simulation model is presented, whose traces are used to propose a detection mechanism, based on statistic parameters of dedicated bearers collected from UEs.

More recently, [35] have addressed the mutual impact of DDoS attacks against 5G slices. In their work, they consider inter- and intra-slice isolation to propose a mathematical model to optimise placing virtual network functions (VNFs) aiming to guarantee end-to-end delay for 5G core network slices. The graphical results show significant improvements when compared with the non-isolated scenario.

Differently from previous related works, REPEL proposes an insurance-based insight to mitigate the attack by anticipating scaling-up extra allocated cloud resources. To evaluate the performance of the proposed approach, we use analytic models that combine the behavioural analysis of players in a non-cooperative game, with an extrapolation of the effects of a DDoS signalling attack against the 5G control plane. The behavioural analysis captures the equilibrium points of the attacker and defender, from which they tend to change their strategies throughout the game. The disruptive analysis infers about the offloading level obtained by scaling-up resources in the control plane. The mitigation approach proposed in REPEL relies on both (*i*) providing the security team with some critical time to improve the defence strategy and (*ii*) potentially discouraging the aggressor to keep attacking. Additionally, instead of managing states synchronisation among the EPC entities or even disintegrating the MME architecture, REPEL proposes using its weighting factor to balance the vMMEs load. To demonstrate the weighting factor operation, we perform a testbed composed of three enodeB simulators connected to a vEPC prototype. Next section we describe the REPEL architecture and propose two analytic models to evaluate its mitigation performance. Table I summarises these related works according to their contributions and objectives.

III. Repel Architecture

The diversity of network services, some of them with stringent requirements like ultra-reliable low-latency communication (URLLC), demands 5G to be a fully convergent network, with a virtualised core, which must be compatible with different many types of access network technologies. This heterogeneous environment will demand 5G control plane to handle with a veritable avalanche of signalling messages, even greater than its predecessor LTE. [36]. However, at the same time such a flooding scenario represents a security threat for 5G control plane [37], its virtualisation pillar also paves the way to new defence insights for mitigating DDoS signalling attacks, without compromising the legitimate traffic.

The defence strategy proposed in this paper aims to set up a front-line of defence, able to ensure 5G control plane availability in case of DDoS signalling attacks. Our approach relies on four main enablers to (*i*) intelligently balance the signalling traffic to the vMMEs, and (*ii*) to potentially mitigate the attack by imposing payoff frustration to the attacker.

- An IDS able to detect signalling traffic anomalies.
- The virtualised control plane of 5G in the cloud that permits quickly instantiating network functions.
- The load-balance features of 5G control plane.
- The payoff frustration imposed on the attacker, once having the target immediately absorbed his attack.

Figure 3 depicts the architecture that supports REPEL operation. The RAN is composed by enodeBs serving both

TABLE I
COMPARATIVE TABLE AMONG DIFFERENT RELATED WORKS.

Strategy of Defence	Approach	Strategy of Defence	Test Methodology	Pros	Cons
SCALE [34]: MME virtualisation	Meeting performance requirements for control plane signalling traffic	Decoupling standard 3GPP interfaces and eNodeB based devices assignment from the MME processing	Prototype and probabilistic models	vMME load balance	Change the vMME architecture
CNS-MME [15]: Cloud-based MME auto-scaling	Meeting performance requirements for 5G signalling traffic	Splitting MME functionalities into container-based micro-services	NFV-LTE-EPC simulator	Lightweight services	Change the vMME architecture
Protection mechanisms for 3G [13]: Comparative analysis	Detection and mitigating DoS attacks against the 3G control plane	Randomisation, CUSUM, and IP packet examination	N/A	N/A	N/A
LTE signalling attack detection [14]: Dedicated bearer attack	Detecting a kind of signalling attack that can disrupt LTE control plane services	Semantic based detection	OPNET simulation model	Does not change vMME architecture	It is applied just to detect dedicated bearer attacks
Secure slicing [35]: Distributed Denial-of-Service attacks in 5G	Guarantee the inter- and intra-slice isolation in case of DDoS signalling attacks	Slice isolation	NS3 simulator and Optimisation model	Does not change vMME architecture	Resource-intensive VNFs in the same host might reach their maximum load at similar times
REPEL: Intelligent control plane resource scaling	Guarantee of control plane services availability	Intelligent vMME scaling to discourage attacker to keep attacking	OpenAirInterface prototype, game theory and queue models	Does not change vMME architecture and does not affect legitimate traffic	The attacker may not stop attacking, even being frustrated

compromised and legitimate UEs. The access network is connected to vEPC through S1 interfaces (S1U and S1AP). All entities that makeup vEPC, namely vMME, vHSS and vSPGW, must be virtualised network functions running on abstracted hardware platforms.

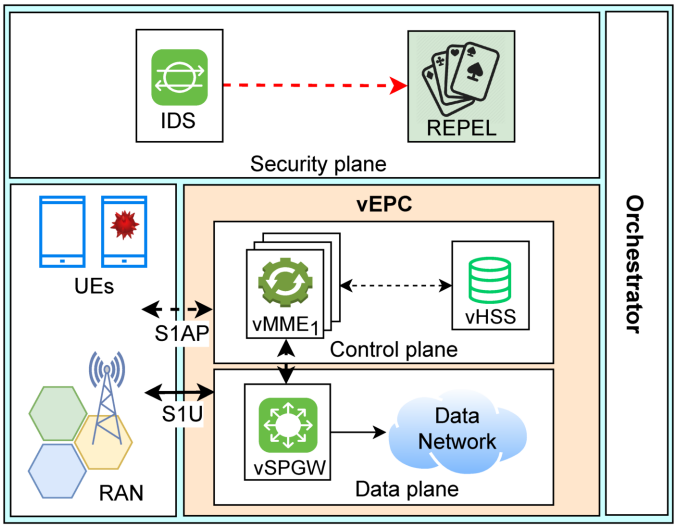


Fig. 3. REPEL Architectural Components.

As described in 3GPP specification [27], all the control traffic from RAN to vEPC control plane is received at vMME. In the control plane, the vHSS is in charge of replying the vMME queries with the subscriber's information. The same vMME also asks vSPGW to set up a session with the UE. After requesting, the UE connects to the vSPGW in the data plane through interface S1U to forward the user traffic.

A not compromised UE is expected to have a usual signalling behaviour. For instance, in the regular operation, a fair UE is supposed to ask for *handover* procedure from time to time or to attach and detach from the network a few times a day. However, in case of a signalling attack, the compromised UE is controlled to create a flooding scenario in the control plane by synthetically and repeatedly imitating signalling procedures like to *attach/detach* or *handover*. [38].

After being ordered by the attacker, we assume each compromised UE coordinately starts an individual DoS signalling attack against the control plane. Due to its small volume of signalling traffic and its implicit stealthy behaviour, it is not possible for any IDS placed at each enodeB to detect that malicious transaction flow as an anomaly[2]. However, the IDS in the security plane can combine evidences from the layers below to detect the misbehaving transaction traffic [39]. Intrusion detection systems for detecting signalling anomalies have been addressed in the literature [40]–[42].

Once detecting the anomalous traffic, the IDS gauges its intensity and triggers the REPEL module to start the mitigation game. REPEL starts the game by evaluating how many vMMEs are currently running in the control plane and the traffic intensity. In case of having just one vMME, regardless of the attack intensity, REPEL triggers the orchestrator to immediately double it. This first move has two main reasons:

[2]No detailed information is provided about the IDS, since it is out of the scope of this work.

to frustrate the attacker by reducing its profit and to automatically create a backup vMME in order to prevent a service outage. If there already are two or more vMMEs running in the control plane, the defender can wait for the upcoming attacker's move before playing again. If the suspicious traffic either does not change or decreases after a time parameter, then REPEL can stop one of the current vMMEs and wait for the next information from the IDS. Otherwise, REPEL triggers the orchestrator to instantiate a new vMME. The flow chart that illustrates REPEL operation is shown in Figure 4.

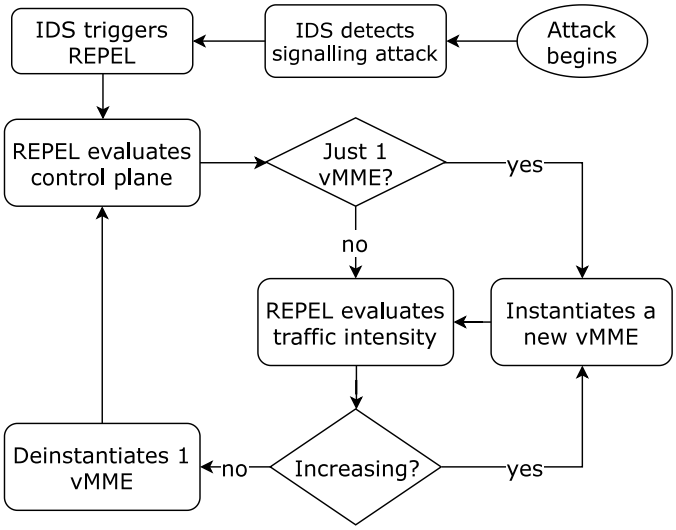


Fig. 4. REPEL logic to scale control plane resources.

The orchestrator vertically interfaces all the layers, including the control plane. It participates on the mitigation game by either adding or removing vMMEs from the control plane, according to Repel instructions. The freshest vMME receives a higher weighting factor to attract to new *attach* requests from both compromised and legitimate users, preventing control plane from eminent service disruption due to signalling flooding. Moving a legitimate UE to another vMME does not affect its traffic because, excepting their relative capacity, both vMMEs have the same setup.

The enodeB interconnection to multiple MMEs is supported by the LTE architecture, according to [43]. In this case, the enodeBs network setup should be beforehand configured to reach the vMMEs IP addresses pool, in which the new vMME will be placed. The weighting factor of a vMME is represented by its "relative-MME-capacity" parameter, that is related with the relative processing capacity of the vMME, regarding the other vMMEs in the same pool. The relative capacity can be sent by vMME to enodeB via S1-AP to load-balance the RAN traffic within the same pool, as defined in [43].

IV. Analytical Framework for Performance Evaluation

Modelling DDoS signalling attacks is a quite challenging task due to its stealthy nature and the scale it needs to reproduce their stress scenarios, among other issues. However, some of these attack situations can be outlined using analytic models based on results obtained from practical tests. In this section, we use queue and game theories to study the service behaviour in the vEPC and the most likely moments for the attacker to either start or stop acting.

Table II shows a list of symbols used in the analytic models.

TABLE II
List of symbols

Sym	Description	Sym	Description
n	Total number of UEs	N	Number of legitimate UEs
M	Number of bots	λ_l	Legit. inc. transact. rate
λ_a	Attack inc. transact. rate	μ	Server processing rate
λ	Signalling transact. rate	q	Server buffer size
K	Storage size	Q	Queue size
m	Number of vMMEs	S	Max. number of vMMEs
U_d	Defender payoff function	U_a	Attacker payoff function
c_d	Defender cost function	c_a	Attacker cost function
S_d	Defender strategy funct.	S_a	Attacker strategy function

In order to make the model tractable without losing generality, we consider the following assumptions and constraints:

- A single attacker controls all the bots sending the same signalling transaction rate λ_a to the vEPC.
- Although the defender can notice abnormal transaction rate growth, he has no prior knowledge of whether the transaction is coming from an attacker or legitimate users.
- There is an infinitely high bandwidth available on the network that connects the vEPC entities.
- The attack rate of each triggered bot λ_a is small and cannot be changed by the attacker nor individually detected by the intrusion detection system.
- Both the attacker and the defender manage finite resources. Thereby, as decision-makers, they are constantly concerned with the cost-benefit ratio of their strategies.
- The attacker's cost c_a, related to his effort and time to recruit and launch bots, increases along with the number of bots already running.
- The defender's cost c_d, related to the extra resources he provides to mitigate the attack, increases along with the number of vMMEs already running.
- The attacker somehow monitors the impact of his attack. A possible way to do that is trying to use attacked services. Otherwise, the attacker can follow online news about the target service outages [44].

A. Modelling the Effects of a DDoS Attack

To evaluate the disruptive effects of a DDoS signalling attack upon the vEPC, we propose an asymptotic analysis based on a queue model, in which the incoming customers represent signalling transactions generated by compromised (λ_a) and non compromised UEs (λ_l). These arriving customers are processed by identical servers with service rate μ, representing the vMMEs running in the vEPC. For each server, we consider a buffer (q) to store incoming customers, while the servers are busy. The load-balancer plays a role of the vEPC weight factor, selecting the vMME with the highest relative capacity to process each incoming transaction. Figure 5 shows the proposed model.

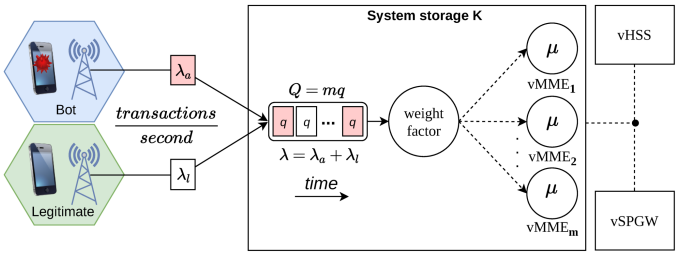


Fig. 5. Queue model representing the vMMEs pool processing incoming transactions from the RAN.

Although the disruptive nature of any denial of service attack weakens the ergodic premise for the system, we assume an eventual steady-state approximation to propose a $M/M/m/K/M$ queue model with m servers, K-size storage, and finite population M, where $M \geq K \geq m$.

The customers (signalling transactions from compromised and non compromised UEs) arrive at the system, represented by the right big box in Figure 5, with rate $\lambda = \lambda_a + \lambda_l$. Service times – the time a server takes to attend a customer – are exponentially distributed with mean μ^{-1}, independent of each other and the arrival process.

The system storage $K = m(q+1)$ accommodates the customers being served and those waiting in the queue Q for an available server. The system queue $Q = mq$ equals to the number of servers running in the system times the buffer size q of a server. In other words, for each new server that starts running in the system, a new buffer q is added in the queue Q. Any arriving customer that finds the system storage K full returns immediately to the initial state and have no further effect upon it. This queue model is fully described in [45] Equations 3.51 and 3.52, in which the probability of having k customers in the system p_k considering two different state regions is given below.

$$p_k = \begin{cases} p_0 \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k} & ; \text{for } 0 \leq k \leq (m-1) \\ p_0 \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k} \frac{k!}{m!} m^{m-k} & ; \text{for } m \leq k \leq K \end{cases} \tag{1}$$

where

$$p_0 = \left[\sum_{k=0}^{m-1} \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k} + \sum_{k=m}^{K} \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k} \frac{k!}{m!} m^{m-k} \right]^{-1} \tag{2}$$

B. Modelling Behavioural Trends During a DDoS Attack

Game theory is a set of analytic tools used to model and understand the interaction between decision-makers (players) under certain assumptions or constraints. The game is a full description of strategic interaction between players, which compete with each other to obtain their own maximum payoff. In particular, game-theoretic approach has been widely employed in network security to explore the interaction between an attacker and a defender during DDoS attacks for suggesting probable actions according to predicted outcomes [46]–[52]. Due to its capacity to model "what-if" behaviours, using game theory to model signalling DDoS attacks can be considered as an interesting approach, since the attacker does not attempt directly against its target. Instead, he takes advantage of previous knowledge about the system to let it exhaust by itself.

Our model relies on a two-players non-cooperative game G, where the attacker and the defender compete with each other for their own payoff (U_a and U_d), related to the disruptive impact of a DDoS signalling attack on the 5G network subscribers. Both the attacker and the defender have just two strategies: attack or no-attack (S_a), and defence or no-defence (S_d), respectively. The attacker goal is to launch the most effective attack to harm signalling services to as many UEs as possible from the smallest possible number of bots. On the other hand, the defender's challenge is to determine the best time to scale-up the best number of vMMEs, preventing service disruption and getting some precious time to enforce his defence, with the lesser cost.

$$Game = G(S_a, S_d, U_a, U_d)$$

Based on the mathematical approach proposed in [50], we consider a unique shooting game, in which each player chooses his strategy, and they make both their own game decisions at the same time.

Firstly, we define $\mathcal{N}(\gamma,\sigma^2)$ representing a Gaussian (Normal) random variable with mean γ and variance σ^2. Let $N \leq n$ be legitimate UEs generating the same mean signalling transactions[3] rate λ_l to the vEPC.

We can particularly model the probability distribution of the legitimate signalling transactions rate by picking $N \leq n$ samples from a Normal distribution $X_i = \mathcal{N}(\lambda_l, \sigma_l^2)$, $i = (1,2,\cdots,N)$. Further, let $T_l = (X_1 + X_2 + \cdots + X_N) \sim \mathcal{N}(N\lambda_l, N\sigma_l^2)$ the probability distribution of the total legitimate signalling transactions rate arriving at vMMEs.

In the case of a coordinated attack involving $M \leq n$ compromised UEs, we consider $T_a \sim \mathcal{N}(M\lambda_a, M\sigma_a^2)$ the probability distribution of the total signalling transactions rate with mean rate $M\lambda_a$ and variance $M\sigma_a^2$. Thus, the total signalling transactions rate, considering both legitimate and attack transactions rate $T = T_l + T_a$.

Lets define $\alpha = \mu/T$ the fraction of transactions that will be dropped by the overloaded vMME when $T > \mu$, and β as the minimum signalling transaction rate of a legitimate UE.

$$\alpha = \frac{\mu}{(N\lambda_l + M\lambda_a)} \tag{3}$$

The number of legitimate UEs whose signalling transactions will be processed by the overloaded vMME can be written as:

$$n_p = N \times \Pr\left[X_i > \frac{\beta}{\alpha}\right] \tag{4}$$

The average rate of UEs whose signalling transactions will be dropped by the overloaded vMME can be written as:

$$D = \frac{N - n_p}{N} \tag{5}$$

[3]We define transactions as a set of messages exchanged among the vEPC entities to accomplish the same signalling primitive.

Combining Equations 3 and 4, into Equation 5, we have:

$$D=\Pr\left[X_i<\frac{\beta\times(N\lambda_l+M\lambda_a)}{m\mu}\right] \quad (6)$$

In Equation 6, $m\leq S$ refers to the number of vMMEs currently processing signalling transactions, and S is the maximum vMME number the defender can instantiate.

The attacker payoff function U_a is related to the signalling losses for the bigger possible number of UEs $u_a=N\times D$, and the cost he pays to control his bots $c_a=M^2/N$. In this case, we assume the attacker cost increases exponentially with the number of bots due to the work of recruiting and controlling them [53]. Therefore, the payoff U_a the attacker tries to maximise by selecting its best strategy can be evaluated by Equation 7.

$$U_a \text{ (UEs/sec.)}=(N\times D)-\left(\frac{M^2}{N}\right) \quad (7)$$

In the same way, it is possible to evaluate the defender payoff U_d as a function of the number of UE attempts successfully processed $u_d=N(1-D)$ by the vEPC, discounting the cost he pays to run its vMMEs $c_d=\mu\times(m/S)$. Note that, the defender cost also increases with the number of vMMEs running in the vEPC m and their individual processing capacity μ in transactions/sec.

$$U_d \text{ (UEs/sec.)}=[N\times(1-D)]-\mu\times\left(\frac{m}{S}\right) \quad (8)$$

The attacker and defender payoff functions defined in the Equations 7 and 8 are respectively placed in a non-cooperative static game table, as shown in Table III.

TABLE III
GAME STRATEGY/PAYOFF ATTACKER X DEFENDER

		defender	
		no defend	defend
attacker	no attack	(0,0)	$(c_d,-c_d)$
	attack	$(U_a,-U_a)$	(c_d-c_a,U_d+c_a)

The game equilibrium can be defined as the point where the players are supposed to be at steady-state. In other words, once achieving their equilibrium points, neither the attacker nor the defender has the incentive to go further or to retreat. In a mixed strategies game, the player chooses his move according to a probability distribution. Thus, we propose a different approach, in which the player is more likely to change his strategy regarding his previous equilibrium point.

Next section we propose a testbed, whose numerical results derive input parameters in the analytic models presented in the Sections IV-A and IV-A aiming to assess the effectiveness and the viability of our approach.

V. EXPERIMENTAL MODEL

In this section, we propose an experiment to reproduce the scenario where a supposed attacker exploits the UE *attach* procedure as the signalling attack vector. In other words, the attacker controls 6 bots, which have been previously recruited by him, to attach to the vEPC at the same time. The group of bots is strategically distributed on 3 different small-cells so that the attack attempt can not be detected by any individual IDS. Note that, although we have chosen the *attach* procedure as our attack vector, we could opt for any other signalling procedure that triggers a large number of messages in the control plane. In this case, besides being simpler to reproduce, it has been previously addressed in [19], [54].

Reproducing a vEPC bottleneck due to a message flooding by a supposed *attach*-procedure-based attack is out of the scope of this testbed. Instead, our experimental goals are to evaluate the vMME service load for processing *attach* attempts and to demonstrate how the LTE overload control performs to offload the vMME under a DDoS attack.

Such a balancing mechanism relies on manipulating the vMME weight factor by changing its relative capacity parameter, which is related to the processing capacity of each vMME regarding the others in the same group. The idea is to balance the signalling traffic from the RAN to the vMMEs according to their processing capacity. In this case, the new vMME is provisioned with a higher relative capacity and tends to lure all the next *attach* attempts, offloading the former vMME.

To demonstrate the offloading mechanism, we built a virtualised testbed composed by one vEPC and one evolved universal terrestrial radio access network platform (E-UTRAN). The vEPC setup uses the 0.5.0 version of Openairinterface [55] running on four different virtual machines hosted by an OpenStack cluster. The VMs 1 and 2 emulate the vMME_1 and vMME_2 respectively, while VMs 3 and 4 emulate vHSS and vSPGW. The E-UTRAN setup distributes itself in three different VMs hosted by the same Openstack cluster. Each VM runs the Openairinterface OAISIM to simulate a pair enodeB + UE together [56], where the enodeB_1 works with one attached UE, the enodeB_2 hosts 2 UEs, and enodeB_3 hosts 3 UEs. Figure 6 depicts the proposed testbed setup.

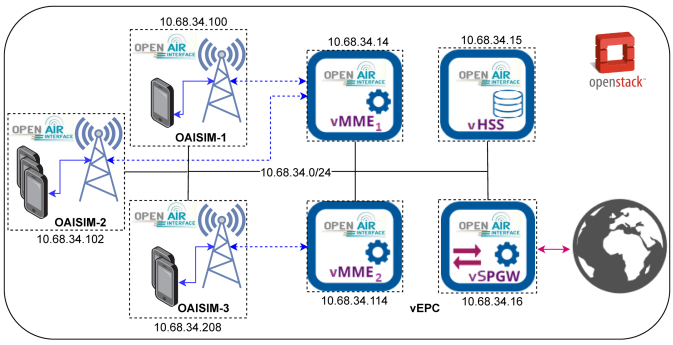


Fig. 6. Testbed setup used to test vMME to enodeBs bindings, according to the vMME relative capacity. The same OpenStack cloud hosts the virtual machines that compose both the vEPC and E-UTRAN platforms.

The virtual machines hosting OAISIM simulators (OAISIM_1 and OAISIM_2) are equipped with two virtual CPUs and 30GB of the disk to run Ubuntu 14.4 low latency kernel. The virtual machines hosting the vEPC entities (vMME_1, vMME_2, vHSS, and vSPGW) use the same physical setup to run Ubuntu 16.5 with kernel version 4.7. The embedded enodeBs of OAISIM_1, OAISIM_2, and

$OAISIM_3$ are simultaneously connected with both vMMEs IP addresses (10.68.34.14 and 10.68.34.114) to permit them choosing the proper vMME during the UE attachment phase, according to their relative capacity. Although each vMME has its vMME ID, they share the same group ID.

In the first phase, our experimental monitors the memory usage of the $vMME_1$ serving as the signalling gateway to all the OAISIM simulators, including the 6 UEs embedded in the simulators. After that, we establish a signalling connection between each E-UTRAN simulator and vEPC through the $vMME_1$ with relative capacity 10, whose memory usage is monitored every 0.5 second. We have chosen the memory usage as our offloading parameter due to the state-full-based behaviour of the vMME functions in the virtual machines.

The next phase of our experimental script aims to evaluate the $vMME_2$ performance to offload the $vMME_1$. Once established the signalling connection between the $OAISIM_1$ and $OAISIM_2$ with the $vMME_1$, we start running the $vMM2_2$ with relative capacity 20. The $OAISIM_3$ is then started running supposed to connect with $vMME_2$. The memory usage of both vMMEs keeps being monitored every 0.5 second.

A. Testbed Results

Our results have shown that the weighting factor mechanism of vMMEs works as expected, luring new *attach* requests to the vMME with the highest relative capacity. According to our testbed script, the $OAISIM_3$ launching had its UEs attached at the $vMME_2$, instead of having it attached to the $vMME_1$, that was running before. Another point to emphasise here is the further possibility to increase the $vMME_1$ offloading level by sending an "overload start" command to the $OAISIM_1$ and $OAISIM_2$. Since the enodeBs have both vMMEs properly addressed, all the UEs will be detached from the $vMME_1$ to attach again to the $vMME_2$.

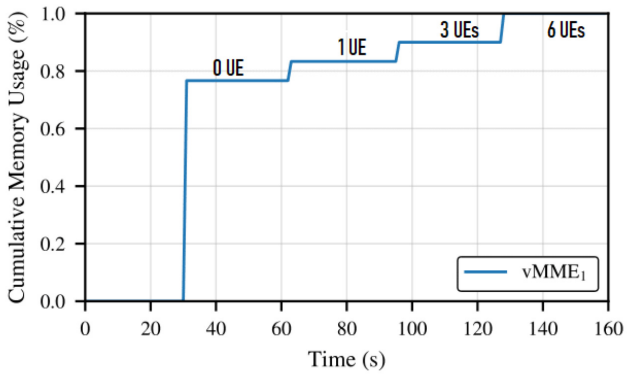
Figure 7a shows the memory usage behaviour in the first testbed setup, in which just $vMME_1$ is in charge of processing all the attach requests. Figure 7b depicts the memory usage progress in the second setup, running with two vMMEs. In this case, $vMME_2$ starts running to process the $OAISIM_3$ attach requests after 95 seconds.

Comparing the memory usage behaviour in each vMME setup, (a) initially running just one vMME and then (b) running two vMMEs simultaneously, it is possible to notice a memory usage offload near 20%. Therefore, as can be seen, the proposed approach confirms its effectiveness in protecting the control plane availability by offloading the first vMME.

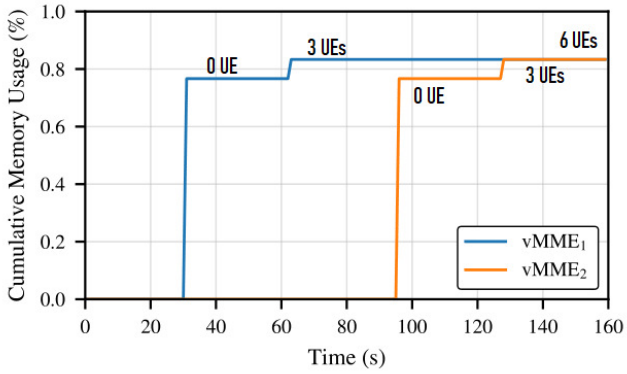
The numerical results plotted in Figures 7a and 7b can be used in further analysis over the disruptive effects of the proposed *attach*-based DDoS attack, and the behaviour of both players in the mitigation game.

VI. Performance Evaluation

In this section, we feed the analytic models proposed in Sections IV-A and IV-B with the numerical results obtained in Section V-A. Our goal is to extrapolate them in order to synthetically reproduce a flooding scenario in the control plane and to demonstrate the offloading trend by adding vMMEs.



(a) Cumulative Memory usage using only $vMME_1$ to attach all the OAISIM UEs. 100% of cumulative memory usage corresponds to 3% in absolute number.



(b) Cumulative Memory usage adding the $vMME_2$ to attach $OAISIM_3$.

Fig. 7. vMMEs cumulative memory usage monitored in two different situations: (a) just $vMME_1$ processing all UEs attachments, and (b) adding the $vMME_2$ to attach $OAISIM_3$.

We also analyse the behavioural trends of both attacker and defender based on the Nash equilibrium points, wherefrom they do not intend to move by changing their game strategies.

A. Evaluating the Offloading Performance

The analytic model presented in Section IV-A takes into account the time-stringent requirements of the signalling traffic, particularly the *attach* procedure traffic, to propose the queue system depicted in Figure 5 as an approximation to allow asymptotic analysis of the system.

In the attack scenario proposed in Section I, compromised UEs are remotely controlled by an aggressor to attempt against the vEPC availability. The attacker takes advantage of the high number of messages triggered by *attach* procedure to flood vEPC and exhaust control plane resources. Each compromised UE that looses its *attach* transaction immediately generates a new one, according to the attack script. This process keeps looped while the attack remains active. The defender, on the other hand, detects the signalling traffic anomaly and acts according to the traffic behaviour to assimilate the attack by anticipating offloading countermeasures.

Based on Figure 7b, we estimate the mean arriving transaction rate of each compromised UE as $\lambda_a \approx 6/120 = 0.05$

transactions per second. In this case, the vMME received 6 UE *attach* requests in approximately 120 seconds. After that, we consider λ_a=0.05 and the absolute memory usage of the vMME to process all the 6 UE *attach requests* equals 0.03 to evaluate the service rate of each vMME μ=0.05/0.03≈2 messages per second. Note that, the service rate (μ) is a fixed parameter, related to the capacity of the server (vMME) to process the incoming signalling traffic. Therefore, in the proposed model, the service times depends only upon the random size of these signalling messages.

To evaluate the buffer size q of each server (vMME), we conservatively assume another approximation considering a lossless scenario, in which the vMMEs pool process all the incoming transactions eventually. Once assuming a lossless scenario, we can use the Little's law to calculate the queue size q=$T\lambda_a$, where T is the time an *attach* transaction can wait until receiving back its *attach* accept/rejected response. According to [57], this time is parameterised on the timer $T3410$ and generally equals 10s. Thus, we have q=10×0.05≈1, as the smaller integer greater than 0.5.

Using the parameters estimated above in the Equation 1 for k=K, we evaluate the blocking probability P_B, when the vEPC loses signalling transactions due to the increasing numbers of compromised UEs attacking the control plane. Figure 8 shows the blocking probability considering 1 to 5 vMMEs running in the same vEPC at the same time.

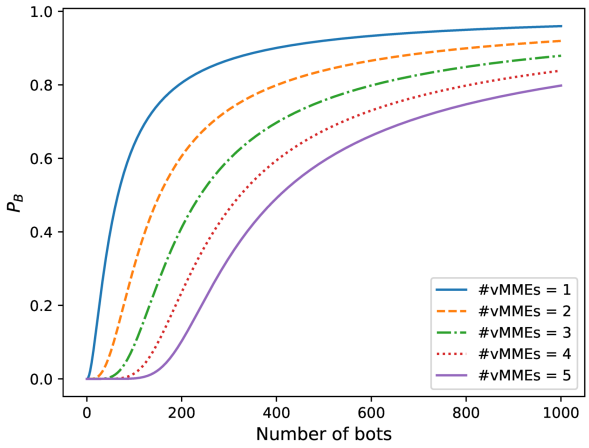


Fig. 8. P_B refers to the probability of blocking an arriving customer after it finding the system storage K full - Equation 1 for k=K.

As can be seen in Figure 8, the blocking probability P_B increases very fast up to 200 bots attacking the vEPC. Control plane services are seriously compromised loosing 80% of signalling transactions due to an attack from M=100 bots.

According to the proposed approach, adding another vMME is enough to absorb this first signalling attack wave and thereby, preventing control plane overload. Holding the same attack scenario (M=100) in Figure 8, adding the $vMME_2$ reduces signalling losses in about 20% (80% - 60%). Besides, intelligently proceeding with vMME instantiations till $vMME_5$ practically eliminates signalling losses, which might potentially frustrate new attacker attempts.

It is expensive for the defender to keep adding new vMMEs indefinitely. Instead, since he has gotten some time by mitigating the immediate attack impacts, the defender should keep enhancing his defence script with more elaborated security countermeasures to block novel attack attempts.

B. Evaluating the Mitigation Performance

The mitigation approach is based on the behavioural trends of the players (attacker and defender) in a non-cooperative game, in which they compete for maximising their own payoff.

The behaviour analysis of the players starts by searching the Nash equilibrium point for each M x m combination scenario. That is, while the number of bots M ranges from 0 to 100, the number of vMMEs m ranges from 1 to 5. The intervals we have used to simulate the player's payoff at each scenario come from the model proposed in [58], considering the UE inactivity time equals 10 seconds.

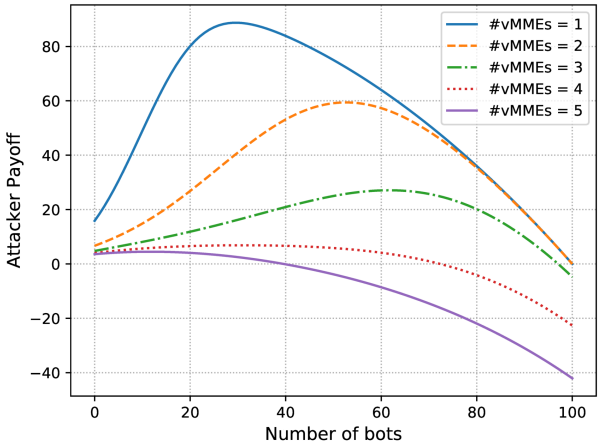
Figures 9a and 9b show the attacker and defender's payoff behaviours, considering different numbers of vMMEs. The attacker payoff peak (U_a=89.7) occurs with 30 bots and just 1 vMME (30,1). Since the attacker payoff is a function of the anomalous signalling rate, assumed as λ_a=10×λ_l, and the number of bots M, U_a=89.7 it means the attack potentially causes signalling disruptions for 90 UEs, which is a reasonable number according to [59], [60]. On the other hand, the defender's payoff peak (U_d=64.13) occurs at (0,1) point that is, 0 bot and just 1 vMME running. However, while it is more profitable for the defender to keep only one vMME running, there is a risk on that, because, for the attacker, it is also a good situation to start attacking.

Figure 9b also shows a particular behaviour considering #vMMEs = 1 or #vMMEs = 2. After M=40 and M=80, respectively, the defender's payoff remains flat because of the depletion of the processing capacity of the vMMEs. Moreover, already having 3 vMMEs running, it is no longer worth for the defender to keep adding new vMMEs after M=60 bots, once his payoff does not significantly change.

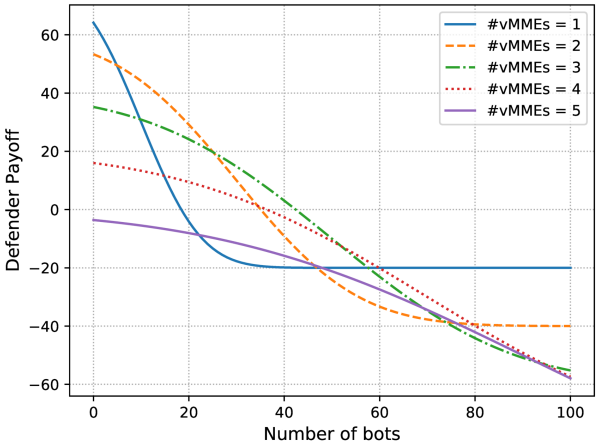
Figure 9c shows the Nash equilibrium points for just one vMME where attacker and defender tend to change their game strategy according to their payoff U_a and U_d respectively. In the attacker curve, A(11,52.77) marks the point where the attacker is more prone to attack. That is to say, before launching any attack, attacker recruits at least 11 to harm around 54 UEs/sec. Points B(71,50) and C(71,−20) indicate the next strategy changing for the attacker and the defender respectively. At C(71,−20) the attacker is likely to stop attacking once the cost to keep attacking is too high to him. In contrast, at C(71,−20) the defender tends to stop defending.

The last picture in Figure 9d shows the Nash equilibrium points for the players' payoff when the defender runs 2 vMMEs in his vEPC. In this case, if the attacker starts either with less than 37 bots or more then 69 bots (point B), he is more likely to stop attacking, once his reward is no longer profitable. Defender, in turn, tends to change his first strategy after 70 bots (point C), when he is likely to stop defending.

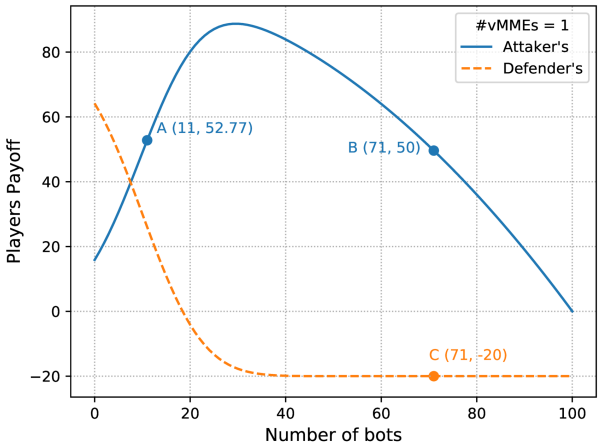
Comparing Figures 9c and 9d, the interval size of recruited bots, where the attacker is more prone to attack, reduces with



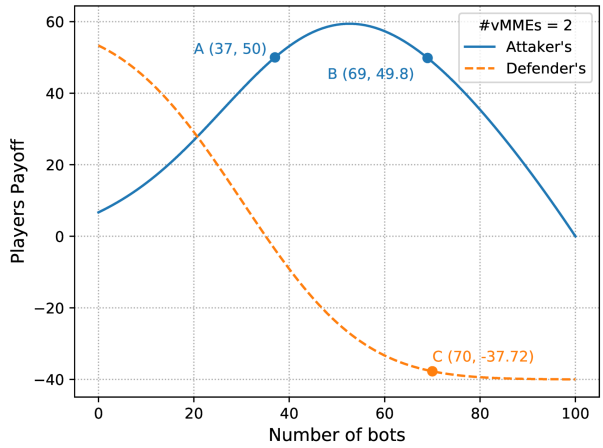
(a) 2D attacker's payoff scanning by each number of vMMEs.



(b) 2D defender's payoff scanning by each number of vMMEs.



(c) Players' payoff holding 1 vMMEs.



(d) Players' payoff holding 2 vMMEs.

Fig. 9. Player's payoff behaviour according to Equations 7 and 8 with the following parameters: N=100 UEs, λ_l=1 +/- σ_l=1 transaction/sec., λ_a=10 transactions/sec., β=0.5 transactions/sec., and μ=100 transactions/sec [58].

increasing the number of vMMEs. From the defender point of view, the equilibrium points are very close to each other at 71 and 70 bots, respectively. The closeness between these two points means the defender's decision is more related to the cost the attacker has to pay to keep attacking.

Although the defender rarely knows when the attacker will launch the attack or even how it will come, we demonstrate that scaling-up vMMEs prevents overloading the control plane and preserves service availability. Further, as shown in the prior paragraph, regardless of the number of vMMEs already running in the vEPC, immediately scaling-up vMMEs might mitigate the attack by frustrating the attacker.

VII. Limitations of REPEL

The core idea inside the REPEL approach is to use extra-allocated resources, which are not meant to be used in normal operation, for scaling-up the vEPC capacity, aiming to assimilate the flooding traffic due to a DDoS signalling attack. The scaling-up approach relies on the cloud elasticity to be accomplished intelligently, move by move as a game, according to the attacker behaviour. The main goals of the approach are to preserve vEPC services availability and to potentially mitigate the attack by imposing outcomes frustration to the aggressor. However, although the elastic environment of 5G vEPC in the cloud enables the scaling approach, it still manages finite resources. Therefore, an intense, long-term attack that persists, regardless of the results of the attack, can be successful in depleting cloud resources. In this case, the uptime provided by REPEL before the cloud depletion should be used to enhance the defence lines. For example, inspecting a sample of the incoming traffic can provide the security team with information to track and disable the attack sources.

Another point that can be considered as a limitation is related to the wasting of the extra allocated resources in the cloud, which would tend to remain idle in the typical operation. The scaling-up process takes into account that the

cloud can quickly provide vEPC with extra virtual machines any time to assimilate the traffic flooding and to prevent control plane outages. In this regard, the cloud should have enough available computing resources during the mitigation game. One possible way to overcome this limitation is to provide the cloud with a prioritised rearranging scheme, in which low priority applications share computing resources with vEPC.

VIII. Conclusions and Future Research

In this paper, we have addressed the implications of denial of services attacks against the control plane in 5G networks, taking into account services availability as one of the most valuable assets to protect. Based on the virtualised core of 5G networks, we propose a security strategy to prevent control plane disruptions by intelligently scaling-up the number of vMMEs to balance the total (legitimate plus non-legitimate) incoming signalling traffic and to frustrate the aggressor, potentially inducing him to give up. Also, we argue that freshest vMMEs do not need to be identical to the ones already running. Instead, once ensuring the control plane management, freshest vMMEs can embed further security policies based on the attack features observed in the first round to effectively and safely block the attack vectors. Repel proposes an alternative insurance-based insight aiming to protect the 5G control plane from attacks and prevent service outages. In fact, according to the statistical law of large numbers, the adoption of this strategy becomes even more advantageous to large providers, once they are usually the most common targets of DDoS attackers. In other words, the larger the provider is, the bigger should also be the reward obtained by its investment, therefore, decreasing the total cost/benefit ratio. The main benefit of anticipating scaling-up resources is to provide the security team with precious time to track the attack and to build enhanced defence lines to block it effectively while preserving the legitimate traffic. Another point we emphasise refers to inducing the aggressor, who is assumed to be aware of the offensive's outcomes, to stop attacking once realising the cost/benefit ratio of the attack is not worth. Indeed, the results obtained from analytic models, show that the aggressor is strongly discouraged to keep attacking once the defender doubles the number of vMMEs. Nonetheless, after instantiating the third vMME, the results also show a trade-off, wherefrom is no longer worth for the defender to keep adding vMMEs.

Moreover, using the weighting factor as load balance feature to offload the vMMEs dismisses using further complex systems to manage the exchanging states inside the EPC. The numerical results obtained from the testbed presented in Section V demonstrate that doubling the number of vMMEs reduces 20% on the memory usage of the attack target.

In general, security systems are quite attractive targets for cyber-attackers aiming to compromise their operation, or as attack vectors, paving the way for internal attacks. REPEL is no exception to this rule. Hence, as well as most of the security systems, REPEL architecture must operate under a particular security framework. One way to protect security systems from direct attacks is to hide them in the network by obfuscating their traffic [61].

For future works, we plan to build a more realistic testbed (or deploying the approach in experimental testbeds [62]) to investigate the practical effects of a signalling attack upon the EPC. The results obtained from the new testbed will help us to improve the payoff functions to get the analytic model closer to the attacker behaviour. We also plan to develop specific semantic protections as part of the new vMME instances.

Acknowledgments

This research received funding from the H2020 5GCity Project under the grant agreement no 761508 and FUTEBOL under grant agreement no. 688941. In Brazil it has also received funding from FAPERJ – the official funding agency for supporting science and technology research in Rio de Janeiro (Brazil) and RedeRio – the state's academic network.

References

[1] M. Sarwar and T. Soomro, "Impact of smartphone's on society," *European Journal of Scientific Research*, vol. 98, 02 2013.

[2] L. Handley. (2019) Nearly three quarters of the world will use just their smartphones to access the internet by 2025. [Online]. Available: smartphones-72percent-of-people-will-use-only-mobile-for-internet

[3] ©Ericsson. (2019) In 2025, 5g networks will carry nearly half of the world's mobile data traffic. [Online]. Available: www.ericsson.com/en/mobility-report/reports/june-2020/mobile-data-traffic-outlook

[4] Q. Han and D. Cho, "Characterizing the technological evolution of smartphones: insights from performance benchmarks," in *Proceedings of the 18th Annual International Conference on Electronic Commerce: e-Commerce in Smart connected World*. ACM, 2016, p. 32.

[5] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5g wireless networks: a comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1617–1655, thirdquarter 2016.

[6] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. D. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, June 2017.

[7] C. Colman-Meixner, P. Diogo, M. S. Siddiqui, A. Albanese, H. Khalili, A. Mavromatis, L. Luca, A. Ulisses, J. Colom, R. Nejabati, and D. Simeonidou, "5g city: A novel 5g-enabled architecture for ultra-high definition and immersive media on city infrastructure," in *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, June 2018, pp. 1–5.

[8] Hackmageddon. (2018) Information Security Timelines and Statistics. [Online]. Available: https://www.hackmageddon.com/2018-master-table/

[9] C. Xiang, F. Binxing, Y. Lihua, L. Xiaoyi, and Z. Tianning, "Andbot: Towards advanced mobile botnets," in *Proceedings of the 4th USENIX Conference on Large-scale Exploits and Emergent Threats*, ser. LEET'11. Berkeley, CA, USA: USENIX Association, 2011, pp. 11–11.

[10] P. Schneider and G. Horn, "Towards 5g security," in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 1, Aug 2015, pp. 1165–1170.

[11] D. Forsberg, G. Horn, W.-D. Moeller, and V. Niemi, *LTE Security*, 2nd ed. Wiley Publishing, 2012.

[12] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A Survey on Security and Privacy of 5G Technologies: Potential Solutions, Recent Advancements and Future Directions," *IEEE Communications Surveys & Tutorials*, pp. 1–52, July 2019.

[13] R. Ettiane, R. Elkouch, and A. Chaoub, "Protection mechanisms for signaling dos attacks on 3g mobile networks: Comparative study and future perspectives," in *Information Science and Technology (CiSt), 2016 4th IEEE International Colloquium on*. IEEE, 2016, pp. 860–866.

[14] R. Bassil, A. Chehab, I. Elhajj, and A. Kayssi, "Signaling oriented denial of service on lte networks," in *Proceedings of the 10th ACM International Symposium on Mobility Management and Wireless Access*, ser. MobiWac '12, New York, NY, USA, 2012, pp. 153–158.

[15] P. C. Amogh, G. Veeramachaneni, A. K. Rangisetti, B. R. Tamma, and A. A. Franklin, "A cloud native solution for dynamic auto scaling of mme in lte," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–7.

[16] P. P. C. Lee, T. Bu, and T. Woo, "On the detection of signaling dos attacks on 3g wireless networks," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007, pp. 1289–1297.

[17] E. P. Rathgeb, C. Hohendorf, and M. Nordhoff, "On the robustness of sctp against dos attacks," in *2008 Third International Conference on Convergence and Hybrid Information Technology*, vol. 2, Nov 2008, pp. 1144–1149.

[18] J. M. Vidal, A. L. S. Orozco, and L. J. G. Villalba, "Poster: Mitigation of ddos attacks in 5g networks: a bio-inspired approach," *Proc. 2nd IEEE European Symposium on Security and Privacy (EuroS&P)*, 04 2017.

[19] J. Henrydoss and T. Boult, "Critical security review and study of ddos attacks on lte mobile network," in *2014 IEEE Asia Pacific Conference on Wireless and Mobile*, Aug 2014, pp. 194–200.

[20] W. Jang, S. K. Kim, J. H. Oh, and C. T. Im, "Session-based detection of signaling dos on lte mobile networks," *Journal of Advances in Computer Networks*, vol. 2, no. 3, 2014.

[21] R. P. Jover, "Security attacks against the availability of lte mobility networks: Overview and research directions," in *2013 16th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, June 2013, pp. 1–9.

[22] D. Fang, Y. Qian, and R. Q. Hu, "Security for 5g mobile wireless networks," *IEEE Access*, vol. 6, pp. 4850–4874, 2018.

[23] A. Shaik, R. Borgaonkar, N. Asokan, V. Niemi, and J.-P. Seifert, "Practical attacks against privacy and availability in 4g/lte mobile communication systems," *ArXiv*, vol. abs/1510.07563, 2015.

[24] D. Ernsberger, K. George, and S. Arumugam, "Security study and monitoring of lte networks," *Journal of ICT Standardization*, vol. 7, pp. 43–60, 01 2019.

[25] D. Kuipers and M. Fabro, "Control systems cyber security: Defense in depth strategies," Idaho National Laboratory (INL), Tech. Rep., 2006.

[26] C. L. Smith, "Understanding concepts in the defence in depth strategy," in *IEEE 37th Annual 2003 International Carnahan Conference onSecurity Technology, 2003. Proceedings.*, 2003, pp. 8–16.

[27] ©3GPP Organizational Partners, "3rd generation partnership project; technical specification group core network and terminals; 3gpp system architecture evolution; ct wg1 aspects," 3GPP, Relase 8, 12 2008.

[28] R. P. Jover, L. Bloomberg, and N. York, "Some key challenges in securing 5g wireless networks," *Electronic Comment Filing System, Jan*, 2017.

[29] I. Alawe, A. Ksentini, Y. Hadjadj-Aoul, and P. Bertin, "Improving traffic forecasting for 5G core network scalability: A Machine Learning approach," *IEEE Network Magazine*, pp. 1–10, Sep. 2018. [Online]. Available: https://hal.inria.fr/hal-01933966

[30] ©3GPP, "5G; System Architecture for the 5G System," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.501, 06 2018, version 15.2.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.02.00_60/ts_123501v150200p.pdf

[31] J. Henrydoss and T. Boult, "Critical security review and study of ddos attacks on lte mobile network," in *2014 IEEE Asia Pacific Conference on Wireless and Mobile*, Aug 2014, pp. 194–200.

[32] C. Douligeris and A. Mitrokotsa, "Ddos attacks and defense mechanisms: classification and state-of-the-art," *Computer Networks*, vol. 44, no. 5, pp. 643–666, 2004.

[33] J. Mirkovic and P. Reiher, "A taxonomy of ddos attack and ddos defense mechanisms," *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 2, pp. 39–53, Apr. 2004.

[34] A. Banerjee, R. Mahindra, K. Sundaresan, S. Kasera, K. Van der Merwe, and S. Rangarajan, "Scaling the lte control-plane for future mobile access," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT '15. New York, NY, USA: ACM, 2015, pp. 19:1–19:13.

[35] D. Sattar and A. Matrawy, "Towards secure slicing: Using slice isolation to mitigate ddos attacks on 5g core network slices," *CoRR*, vol. abs/1901.01443, 2019. [Online]. Available: http://arxiv.org/abs/1901.01443

[36] H. Einsiedler, A. Gavras, P. Sellstedt, R. Aguiar, R. Trivisonno, and D. Lavaux, "System design for 5g converged networks," in *2015 European Conference on Networks and Communications (EuCNC)*, June 2015, pp. 391–396.

[37] P. Marsch, O. Bulakci, O. Queseth, and M. Boldi, *5G System Design: Architectural and Functional Considerations and Long Term Research*, 1st ed. Wiley Publishing, 2018.

[38] S. Mavoungou, G. Kaddoum, M. Taha, and G. Matar, "Survey on threats and attacks on mobile networks," *IEEE Access*, vol. 4, pp. 4543–4572, 2016.

[39] R. S. Silva and L. F. de Moraes, "A cooperative approach with improved performance for a global intrusion detection systems for internet service providers," *Annals of Telecommunications*, vol. 74, no. 3-4, pp. 167–173, 2019.

[40] J.-h. Bang, Y.-J. Cho, and K. Kang, "Anomaly detection of network-initiated lte signaling traffic in wireless sensor and actuator networks based on a hidden semi-markov model," *Comput. Secur.*, vol. 65, no. C, pp. 108–120, Mar. 2017.

[41] G. Escudero-Andreu, K. Kyriakopoulos, J. A. Flint, and S. Lambotharan, "Detecting signalling dos attacks on lte networks," in *Industrial Networks and Intelligent Systems*, T. Q. Duong, N.-S. Vo, L. K. Nguyen, Q.-T. Vien, and V.-D. Nguyen, Eds. Cham: Springer International Publishing, 2019, pp. 283–301.

[42] A. Gupta, R. K. Jha, and S. Jain, "Attack modeling and intrusion detection system for 5g wireless communication network," *International Journal of Communication Systems*, vol. 30, no. 10, p. e3237, 2017, e3237 IJCS-16-0396.R1.

[43] ©3GPP, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP)," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.401, 09 2018, version 15.3.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2446

[44] L. Ookla®. (2012) Downdetector. [Online]. Available: https://https://downdetector.com/archive/

[45] L. Kleinrock, *Theory, Volume 1, Queueing Systems*. New York, NY, USA: Wiley-Interscience, 1975.

[46] T. Spyridopoulos, G. Karanikas, T. Tryfonas, and G. Oikonomou, "A game theoretic defence framework against dos/ddos cyber attacks," *Computers & Security*, vol. 38, pp. 39 – 50, 2013.

[47] H. S. Bedi, S. Roy, and S. Shiva, "Game theory-based defense mechanisms against ddos attacks on tcp/tcp-friendly flows," in *2011 IEEE Symposium on Computational Intelligence in Cyber Security (CICS)*, April 2011, pp. 129–136.

[48] L. Huang, D. Feng, Y. Lian, Y. Zhang, and Y. Liu, "A game theory based approach to the generation of optimal ddos defending strategy," in *Proc. Int. Conf. Comput. Security Digit. Invest.(ComSec)*, 2014, pp. 14–20.

[49] A. Attiah, M. Chatterjee, and C. C. Zou, "A game theoretic approach to model cyber attack and defense strategies," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–7.

[50] Q. Wu, S. Shiva, S. Roy, C. Ellis, and V. Datla, "On modeling and simulation of game theory-based defense mechanisms against dos and ddos attacks," in *Proceedings of the 2010 Spring Simulation Multiconference*, ser. SpringSim '10. San Diego, CA, USA: Society for Computer Simulation International, 2010, pp. 159:1–159:8.

[51] Y. Wang, J. Ma, L. Zhang, W. Ji, D. Lu, and X. Hei, "Dynamic game model of botnet ddos attack and defense," *Security and Communication Networks*, vol. 9, no. 16, pp. 3127–3140, 2016.

[52] B. Kumar and B. Bhuyan, "Using game theory to model dos attack and defence," *Sādhanā*, vol. 44, no. 245, 2019.

[53] C. G. J. Putman, Abhishta, and L. J. M. Nieuwenhuis, "Business model of a botnet," *CoRR*, vol. abs/1804.10848, 2018. [Online]. Available: http://arxiv.org/abs/1804.10848

[54] M. Dano. The android im app that brought t-mobile's network to its knees. [Online]. Available: https://www.fiercewireless.com/wireless/android-im-app-brought-t-mobile-s-network-to-its-knees

[55] N. Nikaein, M. K. Marina, S. Manickam, A. Dawson, R. Knopp, and C. Bonnet, "Openairinterface: A flexible platform for 5g research," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 5, pp. 33–38, Oct. 2014.

[56] C. Nahum, J. Soares, P. Batista, and A. Klautau, "Emulation of 4g/5g network using openairinterfaces," *text in Portuguese - XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, p. 987, set 2017.

[57] ©3GPP, "Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS)," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 124.301, 09 2009, version 8.3.0.3. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/124300_124399/124301/08.03.00_60/ts_124301v080300p.pdf

[58] J. Prados-Garzon, J. J. Ramos-Munoz, P. Ameigeiras, P. Andres-Maldonado, and J. M. Lopez-Soler, "Modeling and dimensioning of a virtualized mme for 5g mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4383–4395, May 2017.

[59] R. Bassil, I. H. Elhajj, A. Chehab, and A. Kayssi, "Effects of signaling attacks on lte networks," in *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, March 2013, pp. 499–504.

[60] J. Jermyn, G. Salles-Loustau, and S. Zonouz, "An analysis of dos attack strategies against the lte ran," *Journal of Cyber Security and Mobility*, vol. 3, no. 2, pp. 159–180, 2014.
[61] J. Khan, "A study in protocol obfuscation techniques and their effectiveness," Ph.D. dissertation, University of Oxford, 08 2017.
[62] C. Both, R. Guimaraes, F. Slyne, J. Wickboldt, M. Martinello, C. Dominicini, R. Martins, Y. Zhang, D. Cardoso, R. Villaca, I. Ceravolo, R. Nejabati, J. Marquez-Barja, M. Ruffini, and L. DaSilva, "Futebol control framework: Enabling experimentation in convergent optical, wireless, and cloud infrastructures," *IEEE Communications Magazine*, vol. 57, no. 10, pp. 56–62, 2019.

Renato S. Silva received the BS degree in Electrical Engineering and Telecommunications from the Federal University of Espirito Santo, Vitoria, Brazil, in 1993, the M.Sc. degree in Systems Engineering and Computing from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 2015. Currently, he is Ph.D. student in System Engineering and Computing at the Federal University of Rio de Janeiro (UFRJ). In his job, he was in charge to deploy new data communication services and develop security policies to guarantee the integrity of the services. His research interests include network security and software-defined networking. Also, he interests in self-organising networks models applied in network security.

Carlos Colman-Meixner received the BS degree in systems analysis from the National University of the East, Ciudad del Este, Paraguay, in 2000, the M.Sc. degree in systems engineering from the National University of Asuncion, San Lorenzo, Paraguay, in 2008, the M.Sc. degree in electrical and computer engineering from the University of California, Davis (UC Davis), USA, in 2014. He has Ph.D. in computer engineering at the UC Davis. He worked as a Professor in computer engineering in the University of Cone Sur of the Americas (UCSA) and in the Autonomous University of Asuncion, Paraguay, from 2001 to 2011. His research interests include resiliency in virtualised networks and machines, in optical networks, in access networks, and in cloud computing.

Rafael S. Guimarães received the B.Sc. degree in Information of Systems from the Faculty of Vila Velha, Vila Velha, Brazil, in 2005, the M.Sc. degree in Computing from the Federal University of Espírito Santo, Vitória, Brazil, in 2015. He is Ph.D. student in computer science at the Federal University of Espírito Santo. He is actually an associate professor at Federal Institute of Espírito Santo and worked for 13 years with operation and advanced support using open source solutions. His research interests include software-defined networks and cross-layer orchestration. Also, his interest in self-organising networks models applied in a context of cross-layer orchestration.

Thierno Diallo received is Msc in high frequency communication systems delivered by a university collaboration (University of Paris Est Marne La vallé, ESIEE, CNAM, Telecom Paris tech and telecom Paris Sud) in 2012. he obtained his PhD in 2016 in Orange Labs collaborating with the University of Limoges where he had dealt with the problematic of optical fronthaul and the convergence of fixed and mobile networks. His research interest is SDN and the virtualization in the 5G networks.

Borja O. Garcia received a BS degree in Telecommunications from Universidad Politécnica de Madrid, Spain in 2007. He is an M.Sc. student in Electrical and Electronic Engineering at the University of Bristol. He worked 10+ years in Operation, Maintenance and Deployment of 2G, 3G and 4G networks for tier 1 operators in Spain and the UK. His research interests include Neutral Host networks and 5G networks enabling technologies such as Software Defined Networks, Multi-Access Edge Computing and Network Function Virtualisation.

Luís F. M. de Moraes received a BS and MSc degree in Electrical Engineering and Telecommunications from Pontifical Catholic University of Rio de Janeiro (PUC-Rio) Brazil, in 1973 and 1976, respectively. He received PhD degree in Electrical Engineering and System Science from University of California at Los Angeles (UCLA) in 1981. In 1992, he participated of the coordination efforts to implement REDERIO – the academic and research backbone network of the State of Rio de Janeiro, a pioneering Internet project in Brazil, supported by FAPERJ (the official funding agency for science and technology research of Rio de Janeiro) since its inauguration. He currently plays a role as technical-scientific adviser of REDERIO/FAPERJ, and is a professor of electrical, computer and systems engineering at the Federal University of Rio de Janeiro (UFRJ), where he also coordinates the high speed networks laboratory (Ravel).

Magnos Martinello received a BS degree in Informatics from Federal University of Paraná (UFPR) and MSc degree in Computer and Systems Engineering from Federal University of Rio de Janeiro (UFRJ) Brazil, in 1998 and 2000, respectively. He received PhD degree in Computer Science from the Institut National Polytechnique de Toulouse (INPT) in 2005. In 2016-2017, he worked as a visiting researcher at University of Bristol (UniBris). He currently holds an associate professor position in the Department of Informatics (DI) at the Federal University of Espírito Santo (UFES), Brazil. Also, he had a Research Productivity Fellowship granted by CNPq (2012-2015) and played a role as the general and technical-scientific coordinator of the PoP-ES project - backbone of the academic and research network of the State of Espírito Santo. His main interests are in Computer Networks, in particular Software Defined Networks (SDN) and Performance Analysis.