Redundant Placement of Virtualized Network Functions for LTE evolved Multimedia Broadcast Multicast Services

Hernani D. Chantre*, Nelson L. S. da Fonseca
State University of Campinas, Institute of Computing, Brazil
Email:hernani@lrc.ic.unicamp.br, nfonseca@ic.unicamp.br

***Abstract*—Mobile Operators have experienced a growth of demand of new services with strict requirements and features which imposes challenges to operate the mobile wireless network. To tackle these challenges, Telcos must provide flexible, dynamic network architecture, furnishing resiliency to ensure end-to-end (E2E) service continuity in case of resource failures. Virtualization of the core network elements of cellular Long-Term Evolution (LTE) and LTE evolved Multimedia Broadcast Multicast Service (eMBMS) technology has been proposed as a key solution to cope with these new demands. By employing point-to-multipoint services to any wireless device, the LTE eMBMS, allows Mobile Operators (MOs) to send a single stream of data to all mobile users in a specific area. The LTE eMBMS is a potential E2E service technology to leverage costumer experience as well as to optimize use of network resources. We propose to improve the reliability of LTE eMBMS E2E services via redundancy, virtualized network functions (VNF), and autonomic service placement. In this paper, we introduce a redundancy series-parallel model to improve reliability of LTE eMBMS services. The model is formulated as a VNF redundancy allocation problem (RAP). We employed the particle swarm optimization (PSO) technique to solve the RAP problem to evaluate the performance of the proposed model.**

keywords: LTE, Multimedia Broadcast Multicast Service (MBMS), Network Function Virtualizations (NFV), Redundancy Allocation Problem (RAP)

I. Introduction

The growth of the number of mobile devices and Internet of things (IoT) devices has been tremendous, and the number of devices is expected to be, respectively, 7.3 and 26 billion units by 2020 [1]. Such growth has generated an increasing demand of broadband network connectivity to transport big volumes of data in mobile operator networks. In addition, as the number of mobile subscribers increases, so does the demand of new type of applications, such as mobile video streaming. This has pushed Telecommunication Service Providers (TSPs) to deliver high quality content with consistent quality of service (QoS) in an efficient way. As a consequence, the growth in signalling traffic and mobile data traffic has increased up to 166% in 3G networks during busy hours [2], imposing considerable challenges to the TSP.

To face the above challenges, the LTE-Broadcast technology, also known as LTE evolved Multimedia Broadcast Multicast Services (eMBMS) standard in 3GPP R9 [3] introduces new opportunities to optimize network resources, as well as to boost network capacity by employing a point-to-multipoint content distribution service with consistent QoS to any wireless device. However, mobile operators still face challenges due the adoption of highly centralized network core, with network elements statically deployed, tightly coupled with the network topology and proprietary equipment. Moreover, the need to deliver dynamic services efficiently and in a timely manner has motivated TSP to employ Network Function Virtualization (NFV) [4] to address these challenges.

Actually, NFV decouples software from hardware enabling network functions to run on commercial off-the-shelf (COTS) equipment [5]. In NFV networks, functions such as Load Balancers, Performance Enhancement Proxies (PEP), Firewall, Network Address Translation (NAT) are dispatched as virtual network functions (VNFs) hosted on servers in datacenters, or even at user premises [4]. By employing NFV, it is expected that Operating Expenses (OPEX) and Capital Expenses (CAPEX) will significantly be reduced.

Virtualization techniques for the evolved packet core (EPC) network functions and Radio Access Network (RAN) of LTE-Advancement deployment have been proposed to mitigate the limitation of cellular mobile networks by employing dynamic resources slicing and spectrum sharing [6]. Meanwhile, virtualization empower TSP to deploy flexible and scalable networks. Moreover, mechanisms to cope with failure of network elements should be incorporated in mobile networks to assure continuity of service as well as to deal with signaling traffic increase due to the failure of network elements.

In this paper, we introduce a virtualized LTE eMBMS E2E service to allow operators to deal with the failure of network elements in order to improve customer experience. We introduce mechanisms to support operators to offer LTE eMBMS E2E service continuity and availability.

Service resilience is an important key performance indicator (KPI) for mobile networks and it is well known by the requirement of five nines reliability. This paper proposes a dynamic and scalable redundancy system for the provisioning of reliable services in LTE eMBMS based on NFV. The proposed solution is based on a redundancy allocation problem (RAP) which includes a series-parallel and multiple component scheme and yet minimizes the delay of E2E services. The solution for

* on leave from University of Cape Verde

978-1-4673-8999-0/17/$31.00

the model is obtained by solving a nonlinear mixed-integer programming problem. Meta-heuristic particle swarm optimization (PSO) was used to solve the redundancy allocation problem. By extensive simulation, we analyzed the trade-off of reliability, delay and optimal number of components.

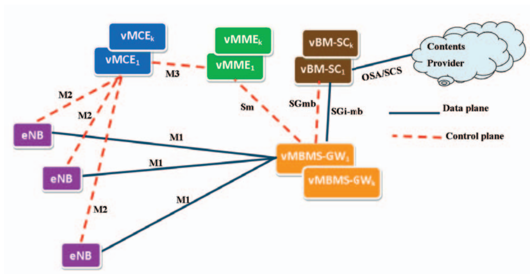

Fig. 1. Architecture for NFV LTE eMBMS

II. LTE eMBMS

This section briefly introduces the LTE eMBMS architecture (Fig. 1) which provides point-to-multipoint distribution of content to a very high number of users by efficiently using radio resources [3]. Its main components are Broadcast Multicast Service Center (BM-SC), Multimedia Broadcast Multicast Services Gateway (MBMS-GW), Mobility Management Entity (MME), and Multicell/Multicast Coordination Entity (MCE).

The Broadcast Multicast Service Center (BM-SC), located at the evolved packet core, serves as an entry point to content providers to inject multimedia content in the LTE network. It also schedules and delivers MBMS transmissions. In addition, it manages interfaces on control and data planes interfaces, the SGmb interface with MBMS-GW for control plane and the SGi-mb interface with MBMS-GW for MBMS data delivery. The Multimedia Broadcast Multicast Services Gateway (MBMS-GW) distributes IP multicast packets to all e-NBs that are part of the eMBMS service using the M1 interface. It performs MBMS session control signaling (session start/stop) to the E-UTRAN using the Sm interface to the MME. The Mobility Management Entity (MME) coordinates MBMS session control signaling (session start/update/stop), delivers MBMS information to the MCE including QoS and MBMS service area using the M3 interface. Multicell/Multicast Coordination Entity (MCE) manages the radio resources for MBMS to all radios that are part of the MBSFN service area. It coordinates the transmission of synchronized signals from different cells (e-NBs) using the M2 interface for control plane.

The Evolved Node B (eNB) gathers information (session start/stop) to be transmitted to mobile devices (UE). The Multimedia Broadcast Multicast Service Single Frequency Network (MBSFN) area consists of a group of synchronized radio cells to make MBSFN transmission, seen as a single transmission by a mobile device.

III. Related Work

A. Function virtualization in mobile networks

Network function virtualization (NFV) in mobiles network has gained significant attention from both industry and academia. In [7], a framework was proposed to optimize transmission over mobile network, using multicast. For this purpose, the MBMS-GW and BM-SC functionalities are integrated into the Service Function Chain in the SGi LAN (Operator's IP services) of a mobile network. Our contribution is related to this technique since it evaluates and compares more suitable reliability model to NFV LTE eMBMS while analyzing failure and restoration procedures for the control nodes involved in the transmission of control messages.

In [5], an NFV framework was proposed for gathering functions of a virtualized evolved packet core based on their interactions and workload with the aim of reducing signaling traffic as well as to enhance the performance of the mobile core network.

In [8], a proposal for 5G mobile network architecture enabled by service resilience-aware mechanisms reduces network overload by controlling signaling messages issued to restore an MME VNF entity. The main objective of the work in [8] was to estimate the number of active/idle User Equipments (UEs) when an MME VNF fails. The work proposed in this paper differs from that in [8] since it focuses on system reliability while minimizing the cost of deployment of LTE eMBMS over NFV.

In [9], a reliable multicast VN mapping problem for an IP network over orthogonal frequency division multiplexing (OFDM)-based elastic optical network (EONs) was study with the goal of maximizing the reliability of requests that has low reliability level. This paper differs from [9] since it minimizes the cost of deployment of NFV LTE eMBMS solutions.

B. Reliability Systems Overview

Several studies have been devoted to tackle software reliability engineering (SRE) problems [10] [11],[12]. Indeed "*SER is defined as the quantitative study of the operational behavior of software-based systems with respect to user requirements concerning reliability*" [11]. Several algorithms have been proposed to solve reliability problems involving redundancy. In [10], it was presented a general classification of optimization techniques to provide reliability by employing redundancy of elements. In [13], the authors presented an interval arithmetic optimization technique to minimize the cost of system redundancy. In [14], an optimal algorithm is proposed to produce reliability based on redundancy of parallel-series systems, while minimizing system costs. In [15], an algebraic method, which solves nonlinear integer programming problem with a linear objective function was proposed to obtain the exact minimum cost of a series-parallel system. With the same goal, the work in [16] transforms the problem into a multiple choice knapsack problem using branch and cut algorithm. The work in [12] used a parametric approach to solve optimization of system reliability with linear constraints. In this paper, we propose to evaluate the NFV LTE eMBMS redundancy allocation problem (RAP) by employing a series-parallel model and a meta-heuristic PSO formulation to solve it.

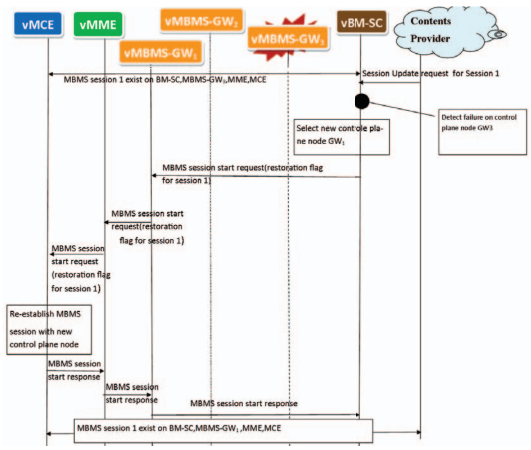

Fig. 2. Failure and recover procedure of a vMBMS-GW [17]

IV. Recovery from Failure of VNF Control Plane Node

A. Failure of control plane node

In the envisioned architecture as depicted in Figure 1, VNF control plane nodes are represented as software images hosted on VMs. A VM can host a single VNF or host a pool of VNFs. Thus, VMs are provisioned based on virtual resources of an underlying hypervisor. Thus, the probability of VNF failure becomes remarkably high, since VNF failures can occur for several reasons such as failure of hardware, hypervisor or software component of a network node. Despite the deployment of VNFs, it is crucial to provide VNF failure detection mechanisms and restoration mechanism for the provisioning of reliable end-to-end mobile services [17]. Hereinafter, we consider the VNF nodes involved in delivering eMBMS services to mobile subscribers since the failure of these nodes can seriously affect the service offered to the mobile subscribers.

In [17], 3GPP provides series of procedures and a restoration principle to limit the failure effects of the nodes on eMBMS services.

The control nodes that can fail in eMBMS are the following: BM-SC, MBMS-GW, MME and MCE. For the sake of simplicity, we focus on the MBMS-GW control node failure, highlighting the main aspect of an MBMS-GW failure process described in [17]. Note that upon a failure of a control node, all its MBMS Bearer context affected by the failure become invalid and should be deleted. The following actions should be taken:

1) Whenever an MBMS-GW fails, the affected MBMS Bearer contexts becomes invalid and should be deleted.
2) Once BM-SC detects an MBMS-GW failure, it initiates a procedure to start a session with the same content as in the original MBMS Session towards alternative MBMS-GW node.
3) The MME detects a restart of an MBMS GW by observing a restart counter in its GPRS Tunnelling Protocol Control Plane (GTP-C). It then deactivates all MBMS Bearer contexts associated with the failed MBMS-GW node.
4) The MME instantiate either a reset procedure or a session stop procedure towards the downstream node MCE to deactivate the associated MBMS Bearer contexts.
5) The MCE should accept the MBMS session start request from the MME. In case the request contains a restoration flag, the MCE should release the M3 resources involved in that session.
6) The eNBs that maintains the former MBMS session should release the M1 IP multicast address and join the new M1 IP multicast address. It can also use the common tunnel end point identifier (C-TEID) in case of different IP multicast source address.

The above-described procedure adopts a reactive approaches. It is expected that the MBMS session restoration to be performed before bringing down the sessions, in order to minimize E2E service disruption.

B. Restoration procedures

Figure 2 depicts a scenario of an MBMS-GW failure and restoration process proposed by the WIPO PCT patent [18] to overcome the disadvantages mentioned above. As shown in Figure 2, all the network control nodes responsible to establish MBMS bearer context are virtualized. Each VM can contain one or more VNF.

In this example, upon the reception of a session update request from the content provider, the vBM-SC detects a vMBMS-GW3 failure and selects an alternative vMBMS-GW1 for the start of an MBMS session, employing a restoration flag, which informs whether it is the first node in the restoration path. If it is, it can decide to send a message delete to the MBMS session of the old node. This process is then propagated towards the downstream nodes. A restoration flag with a timer and a counter is used by the node which initiated the restoration process, to select an alternative downstream control node. Moreover, the existence of the same MBMS session in more than one control node should be avoided, since an MBMS session stored in an old control node (failed node) can conflict with the same MBMS session initiated at a downstream node. The procedure proposed in [18] allows subsequent update and/or MBMS session interruption along the downstream nodes, letting the re-establishment of a session to be performed before the MBMS sessions goes down, minimizing the failure impact on the mobile subscribers. In the next section, we introduce the architecture model to support the restoration process, and to improve the reliability of an E2E service.

V. System Model

The entities involved in the eMBMS service provides control plane (CP) functions with strict requirements in terms of processing latency and computation, thus making them more suitable to be virtualized. We consider a $1:N$ mapping model [19] (Figure 3). Each VNF is decomposed into three types of logical components, forming a virtual component pool. They are:

- The *front end* (FE) which is responsible for communication interfaces with other entities of the network;
- A *service logic* (SL) which implements the logic of a specific VNF entity. The SL is a stateless component since its session information are stored in the SDB;
- The *state database* (SDB) which is a central point containing user session state information handled by the SL, making the SDB a stateful component.

This architecture provides high availability and resiliency to the eMBMS service, as it splits VNFs into independent logical components, allowing dynamic scaling of resources and on-demand resource provisioning, without affecting ongoing session.

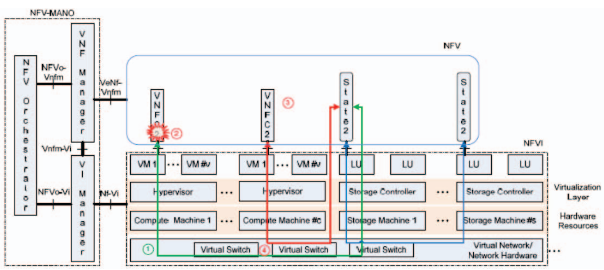

Fig. 3. Architecture model for 1:N mapping stateful Resiliency [20]

VI. Proposed Model

In this section, we introduce the proposed system redundancy model, a series-parallel system to evaluate the reliability model to NFV LTE eMBMS. We focus on guaranteeing the required reliability while minimizing the impact on the end-user services (processing delay) under different kind of failure over the VNF nodes (MCE,MME,MBMS-GW,BMSC). Service provided by service function chain (SFC) is decomposed in user plane to carrier MBMS user services(eNB:MBMS-GW:BM-SC) and control plane (eNB:MCE:MME:MBMS-GW:BMSC) to carry MBMS control services. Our goal is to evaluate the trade off between the processing delay (cost) and level of service redundancy in a series-parallel system design.

A. Series-Parallel System

A series-parallel system is a design model to assure a certain level of reliability to the series model which operation depends on the proper operation of all system components. This design model consists of $i \in (T_1, T_2, T_3, T_4)$ subsystem (VNF classes) organized in series, each subsystem containing k_i redundant VNFs in parallel, with independent probability of failures. The system becomes unresponsive if all VNFs in a subsystem fails. Figure 4 depicts an example of a series-parallel system, focusing on components (VNFs) redundancy. By employing a series-parallel system model [10], our goal is to minimize the design cost of NFV LTE eMBMS with multiple VNF component choices whilts guaranteeing reliability and redundancy for eMBMS service session.

TABLE I
TABLE OF NOTATIONS

Symbol	Description
i	Index of VNF component type
j	Index of parallel redundant VNF component
k_i	Num. of parallel VNF component of type i
C_s	Overall service cost
R_s	Overall service reliability
R	Minimum required service reliability
p_{ij}	Reliability of j-th redundant VNF component of type i
x_{ij}	Num. of j-th redundant VNF component of type i
c_{ij}	Cost of j-th redundant VNF component of type i
u_{ij}	Upper bound number of j-th redundant VNF component of type i

TABLE I summarizes the notations used in the proposed model. The problem can be modelled as a nonlinear integer programming, its formulation is:

$$\text{Minimize } C_s = \sum_{i \in (T_1,T_2,T_3,T_4)} \sum_{j=1}^{k_i} c_{ij} x_{ij} \quad (1)$$

$$\text{subject to, } R_s = \prod_{i \in (T_1,T_2,T_3,T_4)} [1 - \prod_{j=1}^{k_i} (1-p_{ij})^{x_{ij}}] \geq R$$

$$\sum_{j=1}^{k_i} x_{ij} \geq 1 \ ; \ 0 \leq x_{ij} \leq u_{i_j} \ ; \forall_{ij}, x_{i_j} \in \mathbb{Z}^+ \quad (2)$$

Equation (1) is the objective function, representing the total cost of the system to be minimized, as the sum of the cost c_{ij} of all the VNFs components x_{ij} in each subsystem, $i \in (T_1, T_2, T_3, T_4)$. Equation (2) represents the system constraints defining the overall system reliability. The constraint $\sum_{j=1}^{k_i} x_{ij} \geq 1$ establishes that each subsystem must have at least one VNF component. The problem is a VNF placement problem which solution determines the optimal number of redundant VNF component instantiated, x_{ij}, that minimizes the overall cost of the system given a required reliability level R. This is an *NP*-hard problem usually formulated as a nonlinear mixed-integer programming problem [14].

We used PSO technique to find the solution of the RAP [14] as just described.

VII. PSO Algorithm

We propose a meta-heuristic algorithm, based on an Particle swarm optimization (PSO) nonlinear formulation [21] to solve the RAP problem. PSO is a population-based meta-heuristic global search algorithm based on the social interaction of individuals living together such as fish schooling and bird flocking. The goal of PSO algorithm is to find the optimal solution of an objective function [22]. Particles interacts with

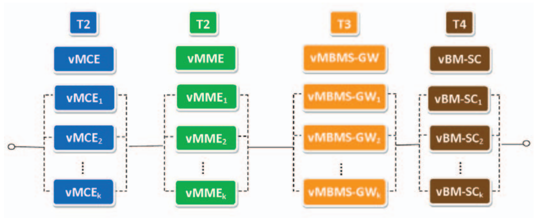

Fig. 4. A Series-Parallel Reliable System for NFV LTE eMBMS

every member of the swarm sharing information about the best positions. Based on the global and local information, each particle updates its own position towards a global optimum. The main elements of the PSO algorithm are:
$X_i = [x_{11}, ..., x_{1k_1}, ..., x_{n1}, ..., x_{nk_n}]^T$, the vector particle, represents the position of the i th particle of the swarm.
$V_i = [v_{11}, ..., v_{1k_1}, ..., v_{n1}, ..., v_{nk_n}]^T$, the vector velocity, representing the velocity of the i th particle.
$L_i = [l_{11}, ..., l_{1k_1}, ..., l_{n1}, ..., l_{nk_n}]^T$, the vector local best, represents the best position the i th particle has visited so far.
$G(t)$, the global best, represents the best position achieved among all the particles in the swarm.
The following equations describe the particles movements:

$$V_i(t+1) = wV_i(t) + c_1\rho_1[L_i(t) - X_i(t)] + c_2\rho_2[G(t) - X_i(t)] \tag{3}$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{4}$$

The PSO parameters are the following: constants c_1 and c_2 also called acceleration coefficients represents cognitive learning and social learning respectively, and ρ_1 and ρ_2 are uniformly-distributed random numbers in [0,1]. The factor w is the inertia used to prevent the unbounded growth of the particle velocities.

Algorithm 1 shows the pseudo code of the PSO algorithm, which implements the equations describing particles movement. In the first step, particles positions and velocities are randomly initialized, then particles start to "fly" throughout the search space according to Equations (3) and (4). During the search process, particles compare its current position with the local best (individual's best position) and global best (group's best position). If the current position is better than the local and global best than it sets the local and global best to the current value. The movement equations are iterated, until some fixed termination criterion is reached, usually a maximum iteration or a sufficiently good fitness which is the best solution met so far. Figure 5 represents the particle of the RAP described in section VI, whereas Equations (1) and (2) represent the fitness function to be minimized and the constraints used to bound the search space for feasible solutions, respectively. We used the penalty-based functions which penalizes unfeasible solutions to handle the constraints related to the problem. This approach convert a constrained problem to unconstrained one by modifying the search space.

Algorithm 1: Classical PSO

Input : Objective function, Min $f : \mathbb{S} \rightarrow \mathbb{R}$, w, c_1, ρ_1, c_2, ρ_2
Output: $G \in \mathbb{R}^D$

/* Initialization */
for $i = 1 \rightarrow N$ **do**
  Randomly Initialize $X_i \in \mathbb{R}^D$;
  Randomly Initialize $V_i \in \mathbb{R}^D$;
  $L_i \leftarrow X_i$;
end
$G \leftarrow argmin_(L_1, ..., L_n)f$;
/* Particles moves and update its local and global position */
repeat
  for $i = 1 \rightarrow N$ **do**
    for $d = 1 \rightarrow D$ **do**
      $V_{ik}(t+1) := wV_{nk}(t) + c_1\rho_1[L_{nk}(t) - X_{nk}(t)] + c_2\rho_2[G(t) - X_{nk}(t)]$;
      $X_{ik}(t+1) := X_{ik}(t) + V_{ik}(t+1)$;
      if $f(X_i) \leq f(L_i)$ **then** $L_i := X_i$;
      if $f(X_i) \leq f(G)$ **then** $G := X_i$;
    end
  end
until *Reached a sufficiently good fitness or a maximum iteration;*;
return G

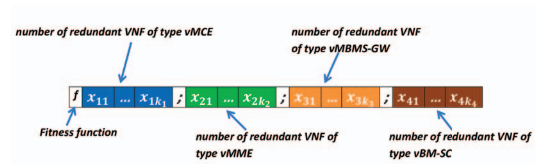

Fig. 5. Representation of the particle in PSO

VIII. Performance Evaluation

This section evaluates the deployment of a reliability model for NFV LTE eMBMS. For MCE operations, the processing delay for RRC procedures should be within [15, 20]ms for initial security activation and RRC connection re-configuration (RB establishment)[23]. Thus, there is a margin of 5ms for MCE processing time. As defined in [24], MME operations should be within 15ms, including 10ms for UE context retrieval also giving a margin of 5ms for MME processing delay. Similarly to [25], in our simulation, we consider that the processing delay of all LTE eMBMS entities are uniformly distributed in the interval [0.8, 3]ms characterizing a non overload scenario. Our approach classifies the delay cost of each entities based on their functionalities, and control signaling traffic generated during the session establishment. Therefore, MCE and MME were set as the most costly entities in comparison to the MBMS-GW and BM-SC. Reliability of an E2E service is the probability that this service has not failed after a certain period of time. As stated in [20], the call drop rate (CDR) reflects service reliability and it is used as a metric to specify the reliability of VNF components. We analyzed different level of reliability for the VNF components.

The parameters used in the simulations were set as follows: c_1 =c_2=2.05, w=0.4, to enhance the global search and convergence speed. The inertia was updated by 5% in each iteration step. Here are four different classes (subsystems), each with k_i parallel component in the range from 4 to

24, with a required reliability of tree nine R=0.999, with u_{ij}=4, and $c_{ij} \in [0.8, 3](ms)$. For the reliability of the VNF components, we used the following reliability level {99.8%,99%,98.7%,97.5%}, characterizing high reliable components. The proposed framework was coded in Java and run on a Debian GNU/Linux Squeeze, with two Intel Xeon (2.13GHz) with 4 cores each, and 78GB RAM. The PSO results were obtained from independent replication method with 95% confidence intervals.

Figure 6 shows the impact of the reliability of the components on the end-to-end processing delay as the number of k_i parallel component grows. Low reliability level causes long delays, since a higher number of components were instantiated to achieve the required reliability. Figure 7 shows that the higher the number of instantiated VNF components, the lower is its reliability level. Therefore, we can state that for low reliability levels, the framework will instantiate more VNF component to achieve the required reliability, increasing processing delay.

Figure 8 analyzes the impact of VNF classes on the number of VFN instantiated. VNFs with low processing delay (vBM-SC, vMBMS-GW) were instantiated more than those with high processing delay (vMCE,vMME). This is due the fact that our goal is to minimize E2E service delay. The objective function in Equation (1) gives preferences to the VNFs with low processing delay. Figure 9 shows that the number of instantiated VNF grows with the number of parallel components in each subsystem. VNF classes with high reliability level where less instantiated in comparison with classes with low reliability level. These results lead us to conclude that the higher the level of component reliability, the lower is the number of resources instantiated, as a result of the reduction of the cost of the end-to-end service. The framework was able to place redundant VNF based on their reliability level and their processing delay while minimizing the E2E service delay.

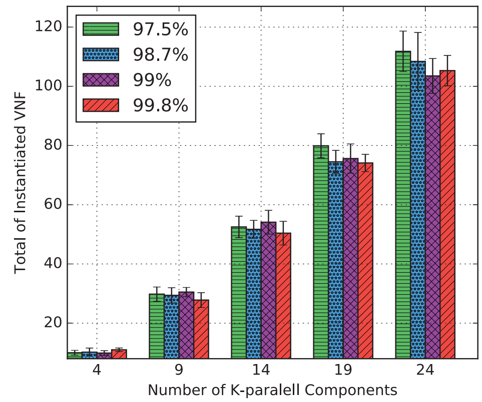

Fig. 7. Overall number of redundant VNF components instantiated

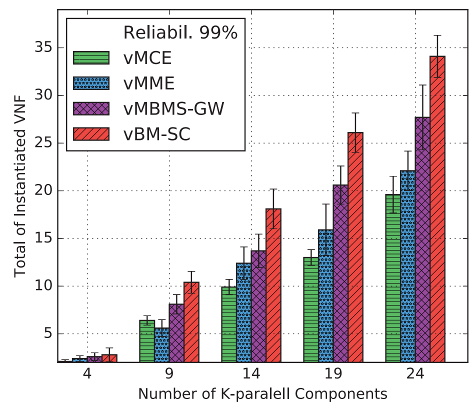

Fig. 8. Overall number of redundant VNF class based instantiated

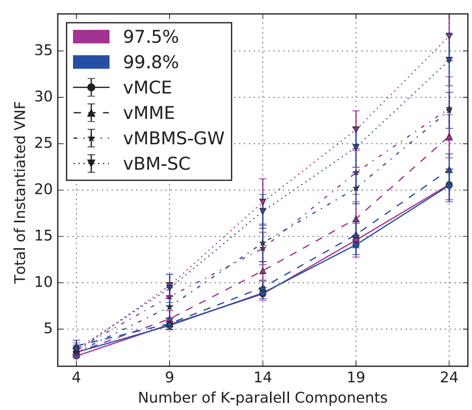

Fig. 9. Overall number of redundant VNF class based instantiated

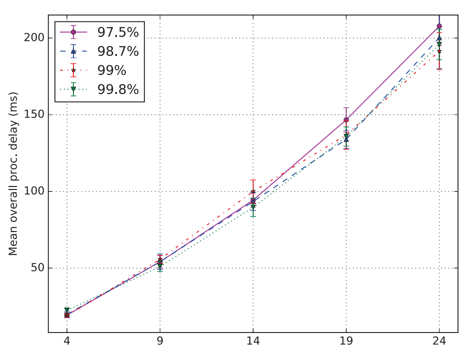

Fig. 6. Mean overall processing delay

IX. Conclusion

This paper proposes a VNF redundancy placement framework to leverage LTE eMBMS in an NFV environment. The algorithm was formulated as a nonlinear integer program with nonlinear constraint, using the design of a series-parallel reliability system. To solve the VFN placement problem, the PSO heuristic was employed. We considered a 1:N mapping architecture for the NFV LTE eMBMS in which a failed stateless component would not impact the session continuity. Using the proposed model, we evaluated the trade-off between minimizing the processing delay of services and the required number of redundant VNF components. Results shows that the proposed model leads to E2E service flexibility, resilience and dynamic scaling. Moreover, the proposed scheme can be adopted in a cognitive management of NFV deployment.

Acknowledgment

The authors would like to thank PEC-PG-CAPES/CNPq Brazil for the financial support.

References

[1] Gartner, "Forecast: The internet of things, worldwide, 2013," http://www.gartner.com/newsroom/id/2636073, 2013.
[2] Alcatel-Lucent, "Reduce core network signaling with a field-proven mme," https://resources.alcatel-lucent.com/asset/184386, 2015.
[3] 3GPP, "Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description," TS 23.246, Jun. 2010.
[4] R. Mijumbi, J. Serrat, J. L. Gorricho, N. Bouten, F. D. Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 236–262, Firstquarter 2016.
[5] H. Hawilo, A. Shami, M. Mirahmadi, and R. Asal, "Nfv: state of the art, challenges, and implementation in next generation mobile networks (vepc)," *IEEE Network*, vol. 28, no. 6, pp. 18–26, Nov 2014.
[6] G. Tseliou, F. Adelantado, and C. Verikoukis, "Scalable ran virtualization in multitenant lte-a heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6651–6664, Aug 2016.
[7] T. Taleb, A. Ksentini, M. Chen, and R. Jantti, "Coping with emerging mobile social media applications through dynamic service function chaining," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2859–2871, April 2016.
[8] T. Taleb, A. Ksentini, and B. Sericola, "On service resilience in cloud-native 5g mobile systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 483–496, March 2016.
[9] X. Gao, Z. Ye, J. Fan, W. Zhong, Y. Zhao, X. Cao, H. Yu, and C. Qiao, "Virtual network mapping for multicast services with max-min fairness of reliability," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 9, pp. 942–951, Sept 2015.
[10] F. A. Tillman, C. L. Hwang, and W. Kuo, "Optimization techniques for system reliability with redundancy 2014;a review," *IEEE Transactions on Reliability*, vol. R-26, no. 3, pp. 148–155, Aug 1977.
[11] M. R. Lyu, "Software reliability engineering: A roadmap," in *Future of Software Engineering, 2007. FOSE '07*, May 2007, pp. 153–170.
[12] S. K. Banerjee and K. Rajamani, "Optimization of system reliability using a parametric approach," *IEEE Transactions on Reliability*, vol. R-22, no. 1, pp. 35–39, April 1973.
[13] H. Munoz and E. Pierre, "Interval arithmetic optimization technique for system reliability with redundancy," in *Probabilistic Methods Applied to Power Systems, 2004 International Conference on*, Sept 2004, pp. 227–231.
[14] A. O. C. Elegbede, C. Chu, K. H. Adjallah, and F. Yalaoui, "Reliability allocation through cost minimization," *IEEE Transactions on Reliability*, vol. 52, no. 1, pp. 106–111, March 2003.
[15] F. Castro, J. Gago, I. Hartillo, J. Puerto, and J. Ucha, "Exact cost minimization of a series-parallel system," *arXiv preprint arXiv:1203.3307*, 2012.
[16] M. Caserta and S. Voß, "An exact algorithm for the reliability redundancy allocation problem," *European Journal of Operational Research*, vol. 244, no. 1, pp. 110–116, 2015.
[17] 3GPP, "Restoration procedures," TS 23.007, 2016.
[18] Q. XIA, D. ZHU, and Y. Yang, "Mbms session restoration in eps for path failure," Oct. 23 2014, wO Patent App. PCT/EP2014/057,722. [Online]. Available: http://www.google.ch/patents/WO2014170369A2?cl=pt-BR
[19] T. Taleb, M. Corici, C. Parada, A. Jamakovic, S. Ruffino, G. Karagiannis, and T. Magedanz, "Ease: Epc as a service to ease mobile core network deployment over cloud," *IEEE Network*, vol. 29, no. 2, pp. 78–88, March 2015.
[20] ETSI, "Network Functions Virtualisation (NFV); Resiliency Requirements," GS NFV-REL 001 V1.1.1, Jan. 2015.
[21] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, Nov 1995, pp. 1942–1948 vol.4.
[22] B. I. Schmitt, *Convergence Analysis for Particle Swarm Optimization*. FAU University Press, 2015.
[23] 3GPP, "LTE;Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification," TS 36.331 R13, Jan. 2016.
[24] ETSI, "LTE;Feasibility study for Further Advancements for E-UTRA(LTEAdvanced)," TR 136.912 v9.3.0 R9, Jun. 2010.
[25] C. C. Marquezan, X. An, Z. Despotovic, R. Khalili, and A. Hecker, "Identifying latency factors in sdn-based mobile core networks," in *2016 IEEE Symposium on Computers and Communication (ISCC)*, June 2016, pp. 484–491.