Reliable Broadcasting in 5G NFV-Based Networks

Hernani D. Chantre and Nelson L. S. da Fonseca

This article introduces a scheme to determine the number of redundant VNF components necessary to satisfy the reliability requirements of broadcasting services in 5G networks. A series-parallel redundant model determines the optimum number of virtual components instantiated to achieve a required service reliability level.

Abstract

The emerging 5G technology will support high data rates with low latency and high levels of reliability. To satisfy these requirements, mobile operators have deployed enhancements to LTE networks in a move toward the development of future 5G networks. One of these enhancements is the LTE-broadcasting service. Another enhancement is the employment of network functions virtualization, which will provide elasticity of resources, scalability, and flexibility. In a virtualized LTE-broadcasting network, the main components of the LTE-broadcasting service will be implemented as VNFs. This article introduces a scheme to determine the number of redundant VNF components necessary to satisfy the reliability requirements of broadcasting services in 5G networks. A series-parallel redundant model determines the optimum number of virtual components instantiated to achieve a required service reliability level. The model is compared to a parallel-series one and shown to achieve greater effectiveness, achieving five nines reliability and promoting low end-to-end delay.

Introduction

Telcos have witnessed an astonishing increase in mobile data traffic, and, as a consequence, there has been an explosion in control signaling traffic in mobile networks. This has caused great stress on the infrastructure and can potentially compromise the provisioning of network services. The challenges involved are triggered not only by a rise in the number of mobile users to some 63 percent of the world's population [1] but also by the emergence of new patterns of communication, such as machine-to-machine communication and the Internet of Things (IoT). These emerging trends are accompanied by a boom in a variety of services and applications such as augmented reality, live streaming, online gaming, and mobile social networks. These services and applications involve strict levels of requirements such as low latency, high transmission rates, and high levels of reliability.

The evolution of the technology proposed by the Third Generation Partnership Project (3GPP), from second generation (2G) to 4G and now to the emerging 5G is an attempt to meet these challenges. 5G networks should provide intelligent mechanisms to deliver services close to the end user in densely populated areas. Such mechanisms should perform in real time and in a cost-effective way. Moreover, 5G will employ a heterogeneous architecture access network composed of small cells and macrocells connected to the core network through wired/wireless links [2].

Moreover, 5G networks must support high data rates, with very low latency (below 1 ms) and high reliability levels (the reliabilities of five nines; 99.999 percent). To satisfy such requirements, mobile operators have deployed enhancements to their Long Term Evolution (LTE) networks for the future deployment of 5G networks. One of these enhancements is the LTE-broadcasting service as reported in 3GPP Release-12 [3]. The evolution of broadcasting services has opened new opportunities for operators, and content providers to create innovative services and augmented broadcasting services. For instance, vehicle-to-anything applications will benefit from latency reduction as a consequence of the employment of several broadcasting areas. Another example is the efficient distribution of software and security updates to millions of connected end devices in IoT. Moreover, broadcasting service can support the coordination of remote radio heads and 5G points of attachment in 5G-crosshaul integrated fronthaul/backhaul transport networks.

A key technology for a 5G heterogeneous network is network virtualization. 5G networks will capitalize on both software-defined networking (SDN) and network functions virtualization (NFV) technologies. In a virtualized LTE-broadcasting network aiming at 5G, the main components of the LTE-broadcasting service will be implemented as virtualized network functions (VNFs), enhanced with both network-aware and content-aware adaptation capabilities, allowing a dynamic flow adaptation scheme driven by variation in channel conditions. The VNFs will run on commercial off-the-shelf (COTS) equipment in already existing data centers and network nodes and even at end-user premises. Moreover, the virtualized LTE-broadcasting network must be agile and scalable.

Mobile network operators (MNOs) need a reliable low-cost virtualized LTE-broadcasting service that can increase the coverage of radio access networks wherever and whenever needed. However, general-purpose hardware components can fail, and there is a strong need to rapidly replace a faulty component with minimal management effort.

This article shows a way to determine the

Digital Object Identifier: 10.1109/MCOM.2018.1700648

The authors are with the State University of Campinos; Hernani D. Chantre is also with the University of Cape Verde.

number of redundant VNF components necessary to satisfy the reliability requirements of broadcasting services in 5G networks. The model determines the optimum number of virtual components instantiated to achieve a required level of reliability. Optimum redundant VNF provisioning depends on several factors such as individual component reliability, capacity of hosting servers, and the number of hosted services.

The proposed reliability scheme is based on the solution of a redundancy allocation problem [4, 5] and employs a series-parallel redundancy model to determine the optimum number of VNFs to be instantiated. Moreover, the trade-off between deploying a series-parallel redundancy model and a parallel-series redundancy one in 5G networks is evaluated. It is shown that the proposed series-parallel scheme is capable of providing the requested reliability level with fewer redundant instantiated VNFs than a parallel-series redundancy scheme, while simultaneously promoting low end-to-end service delays. The series-parallel model provides MNOs with the opportunity to design a reliable network by dynamically scaling up or down network elements as a function of end-to-end broadcasting service outages.

Our approach differs from those in [6–9] since it employs a series-parallel redundancy allocation. Unlike existing approaches, we take into account the resilience of an LTE-broadcasting service as a key performance metric in 5G networks, as well as considering the end-to-end delay of the service.

The rest of this article is organized as follows. The following section presents 5G broadcasting service based on LTE-broadcasting. Then we discuss the challenges of virtualization and resilience mechanisms in NFV networks. Following that, we introduce the proposed series-parallel model and present an alternative parallel-series model. Then we assess the effectiveness of the proposed model and compare it with that of the alternative model. The final section presents concluding remarks.

5G Broadcasting Based on LTE-Broadcasting Service

The LTE-broadcasting service (Fig. 1), also known as evolved multimedia broadcast multicast service (eMBMS), was designed to broadcast multimedia content to multiple devices in an LTE network. The LTE-broadcasting service allows mobile operators to send a single stream of data to a large number of end users with both fixed and mobile devices in a specific area.

The LTE-broadcasting service can be activated for small or large areas, whether a city center or an entire region. It provides enhanced performance of high throughput and excellent coverage [10].

This service introduces three main components into the LTE architecture: the broadcast multicast service center (BM-SC), the multimedia broadcast multicast services gateway (MBMS-GW) components, and the multi-cell/multicast coordination entity (MCE).

The BM- SC, located at the evolved packet core (EPC), serves as an entry point to content providers to inject multimedia content into the LTE network. It also schedules and delivers MBMS transmissions. The MBMS-GW distributes IP multicast packets to all evolved Node Bs (eNBs) that are part of the eMBMS service. It performs MBMS session control signaling (session start/stop) to the evolved Universal Mobile Telecommunications System radio access network (E-UTRAN) using the Sm interface with the mobility management entity (MME), which coordinates signaling in MBMS sessions (session start/update/stop), delivers MBMS information to the MCE, including information on quality of service (QoS), using the M3 interface. The MCE component manages the radio resources for MBMS to all radios that are part of the MBMS single-frequency network MBSFN service area. It coordinates the transmission of synchronized signals from different cells (eNBs) using the M2 interface. The eNBs gather information about starting and stopping of a session to be transmitted to mobile devices (user equipment, UE). The MBSFN area consists of a group of synchronized radio cells, seen as a single transmission by a mobile device.

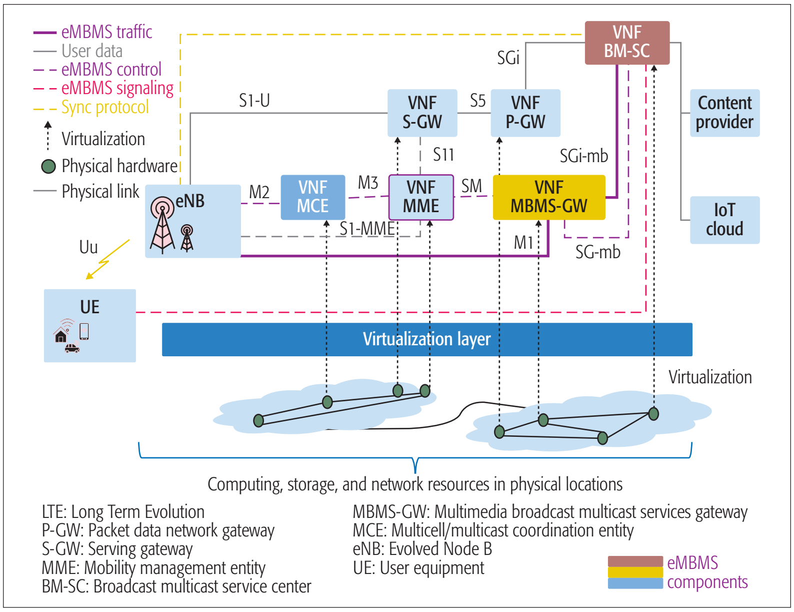

Figure 1. LTE-broadcast NFV-based network.

The deployment of LTE-broadcasting services in the evolution toward 5G-broadcasting demands different protocols and transmission mechanisms to support dynamic adaptation service continuity and 5G requirements. Next, we explain some of these requirements:

- 5G broadcasting must support MBMS operation on demand so that the eMBMS can be switched on and off in congested areas.
- Multiple routes to the BM-SC should also be available to allow alternative paths from the content server to the users.
- Network nodes and enhanced switches at local VLAN aggregation points need to support multicast IP address ranges.
- Multicast listener discovery capability should be available to reduce signaling load and to avoid unnecessary multicast duplication.
- A variety of modulation and coding schemes have to be supported. Based on cell statistics and individual channel state information reports, a 5G broadcasting service should

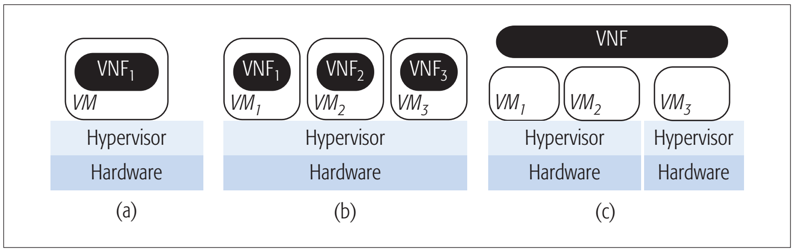

Figure 2. Deployment options for VNFs: a) VNF installed on a dedicated VM provided by the hypervisor; b) multiple VNFs hosted on the same physical host; c) multiple VNF components deployed in separate VMs on the same physical host or on a different physical host allocated in the cloud [12].

apply the most suitable modulation and coding scheme for each end-user device.

- Adaptation of media flows can be supported by employing dynamic adaptive streaming over HTTP and scalable video coding features. The latter may require the instantiation or reallocation of redundant VNFs. The dynamic adaptation of media flows is fundamental to achieve low latency and highly reliable services.
- Application layer forward error correction must be supported to improve the reliability of reception of stream data, as this will allow MBMS end-user devices to recover lost packets [3].

The NFV technology enhances 5G mobile networks with flexibility and agility by allowing dynamic scaling of resources using virtualization of network functions. In 5G broadcasting NFV-based networks, VNFs are instantiated on demand at various network points of presence. Virtualized eMBMS components (e.g., vBM-SC, vMBMS-GW, vMME, vMCE) can be located at small cell gateways to function as aggregation points to support a larger number of applications.

The virtualized components should support session continuity. Virtualized MBMS-GWs should be deployed in small cells connected to the EPC, with the aim of avoiding inter-MME handovers so that MMEs will not be overloaded with signaling traffic, even in a dense network. The VNFs need to be empowered with traffic offloading capability, for example, the 3GPP standard Selected Internet IP Traffic Offload [11], which allows the transport of IP packets in GTP-U tunnels, bypassing the EPC, thus reducing the end-to-end delay.

A 5G-broadcasting NFV-based network facilitates innovation by allowing the creation of new types of services. It allows leveraging machine-to-machine services, as well as on-demand firmware update of IoT devices, and priority and mission-critical push-to-talk and push-to-video wireless services. LTE-broadcasting service should move seamlessly to a 5G broadcasting network, and for that purpose, the employment of NFV will play a key role in deploying new broadcasting services. These emerging services must comply with requirements such as "five nine" reliability and low latency.

Virtualized radio access networks (RANs) introduce potential points of failure, and this can impact the reliability of end-to-end services. This virtualization calls for the adoption of protective schemes to increase the reliability of broadcasting service components.

Resilience Mechanisms in NFV

This section summarizes the main points of resilience provisioning in NFV networks. Resilience is the capacity of the network to provide an acceptable quality of service in the face of problems that might compromise its normal operations. Resilience includes both reliability and security. Security refers to the freedom of a network from unauthorized access or change, destruction, and loss. Reliability refers to the ability to perform according to the specifications. A probability value, called the reliability level, is associated with the chances of maintaining this capability. Although security is one of the main aspects of resilience, this article addresses only the aspects of reliability.

In a virtualized LTE-broadcasting network, a VNF implements a broadcasting component function (MME, MBMS-GW, BM-SC, MCE) on virtual machines (VMs). The employment of NFV enhances resilience in mobile networks as a consequence of the elasticity of this deployment upon the occurrence of a failure. The probability of a VNF failure depends on the implementation, that is, failures may occur at the hypervisor level, where service functions are installed (Fig. 2a). When one or more VNFs share the same physical host, a misconfiguration of resource isolation can negatively impact the performance of one VNF in relation to another (Fig. 2b). Moreover, when VNF components are instantiated on multiple physical or virtual resources (Fig. 2c), a VNF failure can occur as a consequence of a failure either of the underlying hardware or along the communication path [8].

Reliability calls for a set of mechanisms to counteract the failure of VNFs and to minimize the impact of service outages. These mechanisms include network function migration, failure detection and reporting, active and proactive actions upon failures, and session/state management control [13]. Such mechanisms can avoid traffic overload and the creation of network bottlenecks, which can lead to network and service unavailability. The automated deployment of mechanisms to recover from a failure of components should consider service continuity, service function chaining, and topology transparency.

Service continuity means that the services offered should not be impacted by VNF scaling or reallocation. Service continuity in NFV is possible due to the portability of the VNFs, which enables VNF migration triggered by failure events. Monitoring network services are thus essential in real time to automatically guarantee service continuity.

A state database component that maintains VNF state information should be hosted on separate VNFs so that when a VNF component fails, the session's state information is maintained and mapped to an available redundant VNF. The splitting of VNFs into independent logical components contributes to dynamic scaling (Fig. 3a). Proactive VNF failure restoration can also be ensured by using heartbeat messages to monitor the activity of the VNF components [14].

Service function chaining (SFC) defines that network flows and packets should traverse an ordered set of service functions, with ordering constraints forming a logical chain of functions, called service graphs. The design of resilient SFC should be enabled by on-demand policy changes and path selection. To deal with a compromised VNF in a service graph, a different chain can be chosen to host the function without compromising the support of QoS requirements.

Figure 3b illustrates a scenario of a vMBMS-GW failure upon the reception of a session update request from the content provider. The vBM-SC detects a vMBMS-GW failure and selects an alternative vMBMS-GW for the start of an MBMS session. This scenario illustrates how resilience can be provided in SFCs in order to prevent long outages after a service function is compromised. An SFC controller that implements rules for steering flows is used to monitor the health of each service function forwarder (SFF), which determines the destination service function. Upon the compromise of a component, the SFC controller issues a command to change the path by introducing another component.

The network topology should be abstracted to the application for the sake of transparency. MNOs can have an NFV infrastructure distributed over multiple geographical locations, but organized as a logically connected topology enabled with functions capable of managing the entire network, as well as the allocation of resources without any restriction on the underlying network topology (Fig. 3c). Redundancy should be supported across different data centers, enabling multi-site VNF replication. The platform should be able to place VNFs in a topologically independent multi-cloud environment while observing the requirements of the applications. These constraints should be described using policy-driven redundancy or a VNF descriptor, which provides information related to a VNF component.

Series-Parallel Redundancy Model

This section introduces a redundancy model for broadcasting service in NFV networks. A system (e.g., a network) is a set of components arranged in a specific design in order to achieve certain functionality meeting target performance and reliability requirements. The types of components and their quantity and quality as well as the manner in which they are arranged within the system have a direct effect on the system's reliability. In a series configuration, components are arranged one after the other, and a failure of any one of them leads to the failure of the system. A parallel configuration provides more than one path for a sequence of components. In such a parallel configuration, only one of the units must succeed for the system to function. Redundancy is achieved by providing alternative paths in the case of the failure of a component. Adding redundancy is one of the most common methods for improving system reliability. A redundancy allocation problem attempts to achieve a specified reliability level for a system by introducing parallel components or arrangements of a component.

In an NFV network, each component of an end-to-end service represents a potential point of failure. MNOs must consider both the physical components (storage, network, computing) and virtual ones (VNF, virtual links) involved in the delivery of an end-to-end service chain. VNFs are linked to each other by virtual links, thus forming an overlay (logical) network placed on the NFV infrastructure. This consists of an abstract virtualization layer hosted on hardware resources in different locations. Each available VNF component has a different level of reliability and processing delay. A reliable architecture design needs to address the dependencies among all of these components.

Figure 4 illustrates the procedure used to decide on a redundancy scheme. First, the requirements of broadcasting service are analyzed. These requirements can include timeliness and minimum bandwidth, among others. For instance, the time required to broadcast video streaming is considerably less than that required to update a set of mirrored servers. Next, an estimation is made of the number of redundant com-

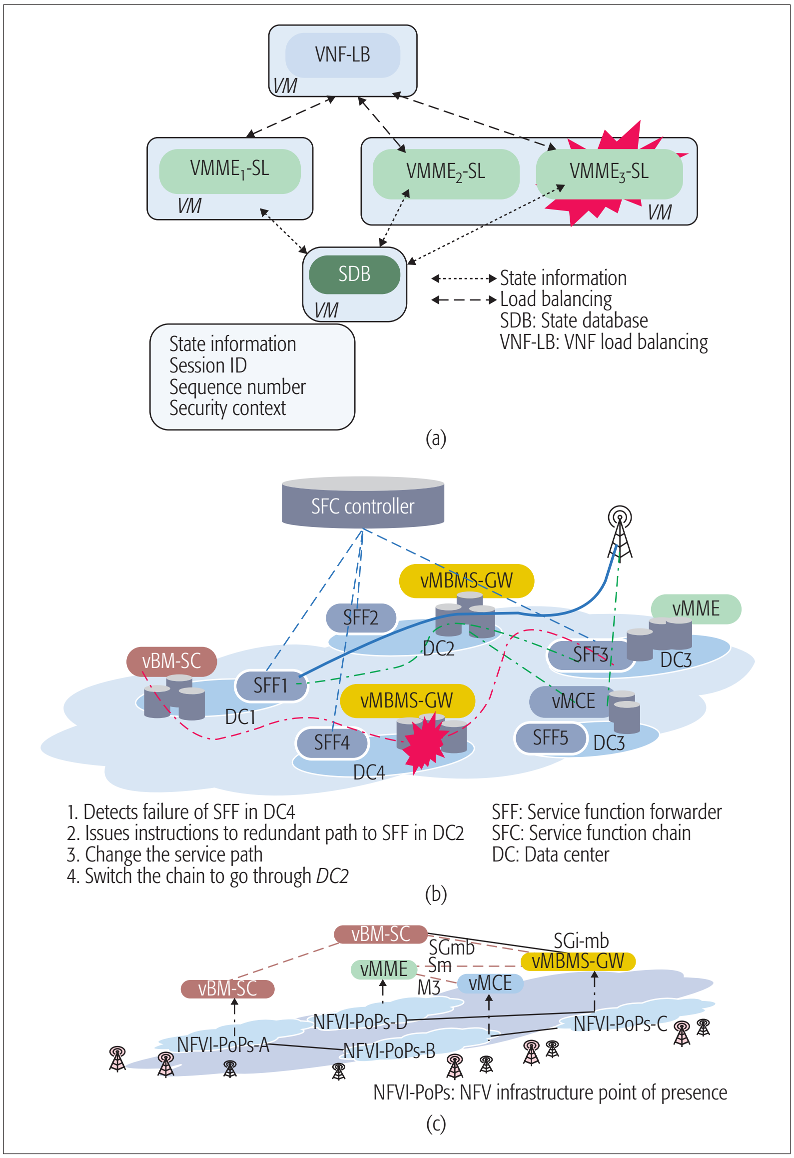

Figure 3. Example of the deployment of VNFs to provide resilience in an LTE-broadcasting service: a) service continuity with stateful VNF resilience; b) service resilience enabled by service function chain; c) service resilience provided by transparent topology with resources allocated in a multi-cloud environment in a geo-distributed location.

The solution to the reliability redundancy allocation problem determines the number of required redundant VNF components in each subsystem, as a function of the reliability requirements of the service, e.g., the optimal number of vMCEs needed to manage radio resource in a specific MBSFN service area.

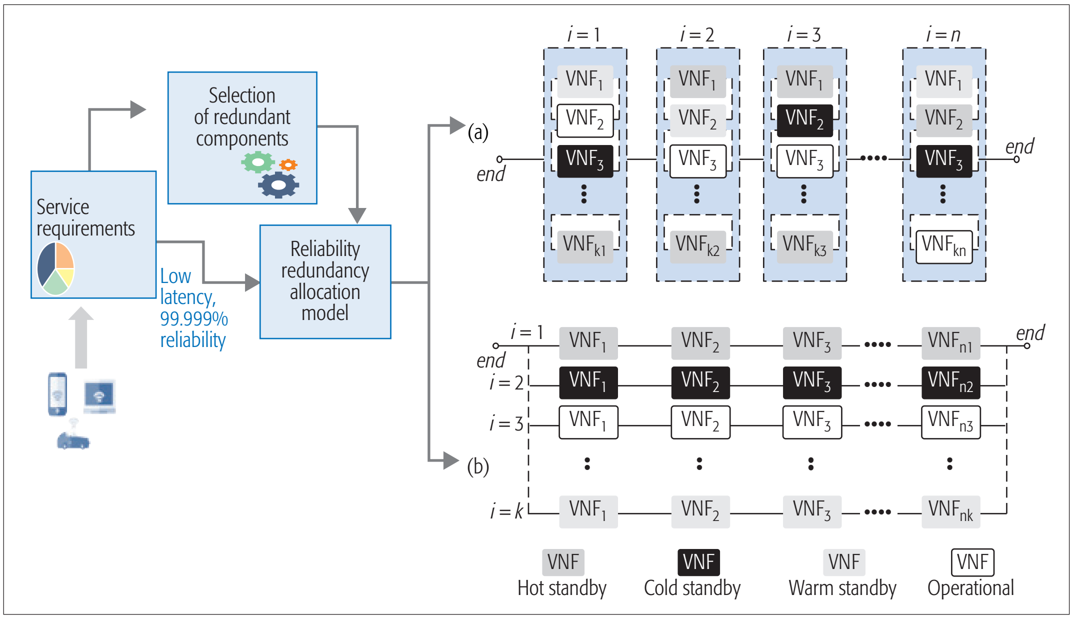

Figure 4. Reliability models with redundant VNF components for NFV deployment: a) Series-parallel; b) Parallel-series.

ponents available and their characteristics, such as processing delay. These two sources of input are provided to a module that solves a reliability redundancy allocation problem [15] to decide which scheme to adopt: a series-parallel one (Fig. 4a) or a parallel-series one (Fig. 4b), or possibly a combination of the two models.

The solution to the reliability redundancy allocation problem determines the number of required redundant VNF components in each subsystem as a function of the reliability requirements of the service, for example, the optimal number of vMCEs needed to manage radio resource in a specific MBSFN service area.

In Fig. 4, redundant VNFs are displayed in different modes: hot, warm, or cold. In the hot standby mode, a VNF should work in synchronization with the primary operational VNF and must be ready to replace a compromised primary VNF. Hot standby VNFs are subject to failure, with the probability of failure independent of the failure of the operational VNFs. In the cold standby mode, the VNFs are inactive, so they cannot fail until they are launched into the operational state to replace a compromised or failed VNF. In the warm standby mode, the VNFs are active but operate with a minimum number of tasks, consuming less processing power and memory than do those in the hot standby mode. VNFs in warm mode are also prone to failure, as they are in hot standby mode, although the probability of failure is lower than that in hot mode, since they are not fully loaded.

Figure 4a shows an example of a series-parallel system focusing on the VNF components (VNFs implementing a specific component function, e.g,, BM-SC, MME, MBMS-GW, or MCE), organized as *n* parallel subsystems in a series. Each subsystem contains redundant VNFs in parallel, and each VNF has an independent probability of failure as well as processing delay. The system becomes unresponsive with the failure of any one of the subsystems, and a subsystem fails when all of its constituent redundant VNFs fail. Each subsystem represents an individual resource pool of a specific LTE-broadcasting component function, for example, a pool of vMME, vBM-SC, vMBMS-GW, and vMCE.

The series-parallel redundancy model allows the design of a reliable network by scaling up or down VNF components upon the occurrence of failure of a component, thus providing high levels of reliability with self-healing capabilities. Moreover, the proposed model can decrease the time required to react to failures in decentralized and virtualized LTE-broadcasting networks by locating virtualized components (vMBMS-GW, vMME, vMCE, and vBM-SC) closer to the end user.

Along with the above mentioned reliability model, deploying NFV in 5G mobile, MNOs should employ tools to measure the network state and control the network risk of failure, acting proactively and mitigating the spread of vulnerability.

In order to assess the effectiveness of the proposed series-parallel redundancy model, it was compared to a counterpart model with parallel-series redundancy. In the latter, an entire end-to-end service is replicated in parallel, and the end-to-end service becomes unresponsive if a single VNF component fails in any part of a redundant system. Figure 4b depicts an example of a parallel-series model.

In a parallel-series model formulated with redundant standby systems, each system consists of VNF components in a series, forming a service function chain. The sharing of VNF components is unlikely to be used for the support of different broadcasting services (e.g., mission-critical push-to-talk voice over LTE and on-demand firmware update services for IoT devices) in a parallel-series redundancy model since a compromised shared VNF component impacts the provision of the different services.

An NFV architecture based on multiple sites introduces geographically distributed resources in multiple locations, but it increases the complexity required to manage the service when reliabil-

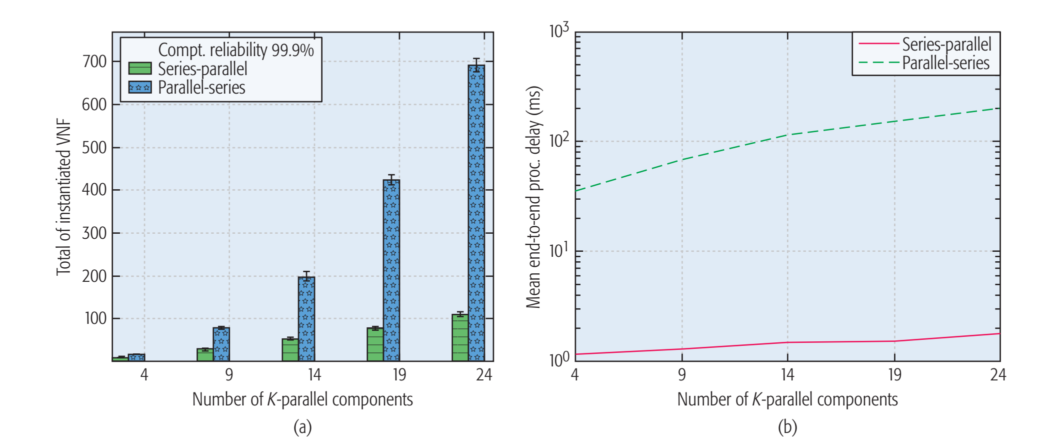

Figure 5. Comparison of the two redundancy models: a) the number of redundant VNFs instantiated in series-parallel and parallel-series models; b) the end-to-end delay service in series-parallel and parallel-series models.

The scalability and flexibility provided by the series-parallel redundancy model yield a reasonable number of instantiated VNF components yet produces low end-to-end delays. This model gives the MNOs the opportunity to increase the resilience of broadcasting services by using redundant components in hot-standbys.

ity is provided by a parallel-series model. More elaborate management is needed, and the time necessary to detect and recover from failures can increase significantly, which can potentially compromise the reliability of the service.

Performance Evaluation

When designing a redundant system, MNOs need to optimize the number of redundant VNFs to achieve high levels of reliability considering the processing delay of the VNFs. The problem is formulated as a reliability redundancy allocation problem, and it minimizes the total processing delay of the VNFs instantiated. By minimizing the total delay, an attempt is made to obtain a minimum number of VNFs and end-to-end service delays. Alternatively, one could solve a bi-criteria optimization problem to minimize both end-to-end delay and the number of VNFs, but such formulation would considerably increase the complexity of the solution of the problem. The problem involves a nonlinear optimization, and was solved by employing a particle swarm optimization algorithm.

In the evaluation, different levels of reliability were used to characterize the reliability of each type of VNF component. The component reliability levels were assumed to be in the interval {99.9 percent; 99.99 percent}. A high-reliability level for the service was ensured by requiring a "five-nines" reliability. All components were considered to be in the hot mode. The processing delay of the VNFs was randomly drawn from a uniform distribution [0; 1.0] ms. The service function chain for the broadcasting service was composed of four components (vMCE, vBMSC, vMBMS-GW, and vMME). Combinations of the reliability level of each component and their processing delay were generated randomly. Results were obtained using an independent replication method with 95 percent confidence intervals.

Figure 5a shows the number of redundant VNFs instantiated in both series-parallel and parallel-series redundancy schemes for a k redundant system. Results show that the series-parallel scheme requires fewer components than does the parallel-series one. The difference is at least twice as much for $k = 9$, and it increases as the degree of redundancy increases, being seven times greater for $k = 24$. This is due to the fact that in the parallel-series redundancy scheme, the failure of a single VNF component can cause the failure of the entire end-to-end broadcasting service provided.

The parallel-series scheme ends up with more VNF components than does the series-parallel scheme, as it is unable to provide redundancy on a component basis. For instance, to avoid a bottleneck or prevent failure on a control plane component (e.g, vMME) the series-parallel scheme might provide a redundant vMME, while in the parallel-series scheme, the entire service function chain of the broadcasting service needs to be instantiated as a redundant path. The scalability provided by the two models directly impacts the end-to-end service delay. Results indicate that the use of a parallel-series system is not to be recommended for a 5G broadcasting network due to the high cost of maintaining the strict reliability requirements of that technology.

Figure 5b compares the end-to-end delay service produced by both redundancy models. This delay is the delay resulting from using redundant components in case of failure to establish the end-to-end path. The parallel-series system leads to end-to-end delays on the order of one magnitude greater than does the series-parallel one, being much greater than the delay requirements of 5G networks. This happens because the demanded number of components is much higher for the parallel-series model, which increases the chances that components with a high processing delay will be used, especially in heterogeneous networks when the delay resulting from different components is quite heterogeneous.

In summary, we point out that the series-parallel redundancy model efficiently supports broadcasting in a 5G NFV-based mobile network. The scalability and flexibility provided by the series-parallel redundancy model yield a reasonable number of instantiated VNF components but produces low end-to-end delays. This model gives MNOs the opportunity to increase the resilience of broadcasting services by using redundant components in hot standby mode.

The scalability and flexibility provided by series-parallel redundancy model yield a reasonable number of instantiated VNF components yet produces low end-to-end delays. This model gives the MNOs the opportunity to increase the resilience of broadcasting services by using redundant components in hot-standbys.

Conclusion

Emergent broadcasting services call for mechanisms to support the resilience requirements of 5G networks, as this is critical for service providers. In the present article, the challenges of providing broadcasting service in 5G NFV-based networks are discussed. This article proposes a series-parallel redundant model to provide resilience in 5G broadcasting services. The employment of a series-parallel model is compared with that of a parallel-series model, and the superiority of the former in minimizing the number of components demanded and limiting end-to-end delays is shown. A series-parallel model is thus recommended for the provision of reliability for broadcasting services since a parallel-series one is more expensive and leads to unacceptable end-to-end service delays.

Acknowledgment

The authors would like to thank PEC-PG-CAPES/CNPq Brazil and FAPESP (Proc no 15/24494-8) for financial support.

References

[1] GSMA, "The Mobile Economy," 2016; https://www.gsma.com/mobileeconomy/, accessed 7 Nov. 2017.

[2] I. Giannoulakis *et al.*, "The Emergence of Operator-Neutral Small Cells as a Strong Case for Cloud Computing at the Mobile Edge," *Trans. Emerging Telecommun. Tech.* vol. 27, no. 9, 2016, , pp. 1152–59; http://dx.doi.org/10.1002/ett.3077

[3] 3GPP, "Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs," TS 26.346, Oct. 2014.

[4] H. D. Chantre and N. L. S. da Fonseca, "Redundant Placement of Virtualized Network Functions for LTE Evolved Multimedia Broadcast Multicast Services," *IEEE ICC 2017*, May 2017, pp. 1–7.

[5] G. Tseliou, F. Adelantado, and C. Verikoukis, "Scalable RAN Virtualization in Multitenant LTE-A Heterogeneous Networks," *IEEE Trans. Vehic. Tech.*, vol. 65, no. 8, Aug. 2016, pp. 6651–64.

[6] J. Liu *et al.*, "Reliability Evaluation for NFV Deployment of Future Mobile Broadband Networks," *IEEE Wireless Commun.*, vol. 23, no. 3, June 2016, pp. 90–96.

[7] M. Schller *et al.*, "Resilient Deployment of Virtual Network Functions," *Proc. 2013 5th Int'l. Congress Ultra Modern Telecommun. Control Systems Wksps.*, Sept. 2013, pp. 208–14.

[8] T. Taleb, A. Ksentini, and B. Sericola, "On Service Resilience in Cloud-Native 5G Mobile Systems," *IEEE JSAC*, vol. 34, no. 3, Mar. 2016, pp. 483–96.

[9] D. Cotroneo *et al.*, "Network Function Virtualization: Challenges and Directions for Reliability Assurance," *Proc. 2014 IEEE Int'l. Symp. Software Reliability Engineering Wksps.*, Nov. 2014, pp. 37–42.

[10] D. Lecompte and F. Gabin, "Evolved Multimedia Broadcast/Multicast Service (EMBMS) in LTE-Advanced: Overview and Rel-11 Enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, Nov. 2012, pp. 68–74.

[11] I. Local, "Access and Selected IP Traffic Offload (lipa-sipto) (Release 10)," TR23, vol. 829, 2011, p. V10.

[12] ETSI, "Network Functions Virtualisation (NFV); Resiliency Requirements," GS NFV-REL 001 V1.1.1, Jan. 2015.

[13] H. Hawilo *et al.*, "NFV: State of the Art, Challenges, and Implementation in Next Generation Mobile Networks (VEPC)," *IEEE Network*, vol. 28, no. 6, Nov./Dec. 2014, pp. 18–26.

[14] T. Taleb *et al.*, "Ease: EPC as a Service to Ease Mobile Core Network Deployment over Cloud," *IEEE Network*, vol. 29, no. 2, Mar./Apr. 2015, pp. 78–88.

[15] M. Caserta and S. Voss, "An Exact Algorithm for the Reliability Redundancy Allocation Problem," *Euro. J. Operational Research*, vol. 244, no. 1, 2015, pp. 110–16.

Biographies

Hernani D. Chantre received his B.S. degree in applied mathematics and informatics from the Peoples Friendship University of Russia in 2006, and his M.Sc. degree in computer science from Bridgewater State University, Massachusetts, in 2010. He is an assistant graduate professor at the University of Cape Verde. He is currently working toward his Ph.D degree in computer science at the State University of Campinas, Brazil. His research interests include network functions virtualization and network-based cloud computing.

Nelson L.S. da Fonseca received his Ph.D. degree from the University of Southern California in 1994. He is a full professor at the Institute of Computing, State University of Campinas, Brazil. He is the IEEE ComSocVice-President of Technical and Education Activities. He also served as ComSoc Vice-President of Publications, Vice-President of Member Relations, Director of Conference Development, Director of Latin America Region, and Director of Online Services. He is a past Editor-in-Chief of *IEEE Communications Surveys & Tutorials*.