Cognitive Network Fault Management Approach for Improving Resilience in 5G Networks

Borislava Gajic, Christian Mannweiler, and Diomidis S. Michalopoulos
Nokia Bell Labs, Munich, Germany
Email: {firstname.lastname}@nokia-bell-labs.com

***Abstract*—Resilience is one of the fundamental requirements of critical communication services such as ultra-reliable low latency (URLLC) services offered by 5G networks. In order to support the communication service, the 5G networks can take different approaches for deployment of network functions, i.e. the network functions can run on virtualized infrastructure (telco cloud) as well as on the specialized physical hardware instances (e.g. RAN functions). Irrespective of the deployment approach taken the adequate level of resilience needs to be supported on all parts of the network in order to achieve required level of service resilience. In this work, we aim at improving the resilience level of communication services by applying network fault management techniques specialized for 5G slicing-enabled networks taking jointly into account the aspects of virtualized and physical infrastructure. We describe the novel approach of designing flexible and cognitive fault management functions that can dynamically adapt their behavior based on the actual network slice requirements and current network context. We highlight the benefits of such an approach in achieving the required level of resilience especially addressing the telco cloud domain.**

***Index Terms*—Network resilience; network function virtualization; virtualized architecture; network slice fault management**

I. Introduction

Resilience is the ability of the network to continue operating correctly during and after an occurred network problem. The network fault management contributes to the service resilience by identifying and handling the network failures. The fault management continuously monitors the network state, detects the network performance degradation, determines the actual problem i.e. the root cause of the degradation and finally provides suitable solution for occurred problems [3]. In legacy networks the fault management focused mainly on the RAN part of the network, where a major attention has been devoted to the anomaly detection, diagnosis and healing at the cell level [4]. However, the next generation networks will imply significant changes to the network architecture which need to be taken into account by fault management in order to perform suitable actions. Especially introduction of virtualization to the network deployment and the concept of network slicing impose new challenges to the network management and to the fault management as its integral part.

This work has been performed in the framework of the H2020-ICT-2016-2 project 5G-MoNArch. The au-thors would like to acknowledge the contributions of their colleagues. This information reflects the consortiums view, but the consortium is not liable for any use that may be made of any of the information contained therein.

While virtualization [1] of traditional network elements broke up the tight coupling between hardware and software, it introduced additional complexity in handling the faults of network functions. In virtualized networks three layers of deployment can be identified: network function logic, virtual infrastructure (e.g. virtual machines, containers, etc) and physical infrastructure (e.g., commercial off-the-shelf (COTS) servers, compute and storage components) [5], as illustrated in Figure 1. In such an environment there might be different implementation and deployment options for network functions, i.e. there can be many to many relationships between layers of network functions logic, virtualized infrastructure and physical infrastructure where the resources allocated to the network function reside. Such layered implementation of network function requires enhanced fault management logic which considers the actual deployment and interrelations between these layers. As the faults can be related to the same layer and/or different layers of network function deployment, the correlation between fault occurring at the same as well as different layers is essential for root cause analysis in virtualized networks. The main aim of fault management in a virtualized environment is to discover the fault before the major effects take place and/or propagate among different layer as well as to handle the faults at the layer where it occurs, thus avoiding its propagation and effects to other parts of the network.

Furthermore, the fault management for next generation mobile networks needs to be adapted and extended towards the 5G network slicing context. The fault management characteristics and parameters need to be adjusted to the actual service that is supported. Due to the envisioned high dynamics of slice-enabled networks where the network slices can have much shorter life-cycle than the services supported by legacy networks, as well as versatility of offered services the fault management needs to be able to dynamically adapt to the actual service requirements and the current network context. The automation of such an adaptation is a crucial aspect for achieving the scalability of the solution. In a nutshell, the fault management procedures need to be defined in a way to comply with specific network slice requirements and current network context. Thus, the fault management needs to implement the cognition to learn from its environment in an automated way and to adapt its procedures accordingly. The contribution of this work is directed towards this end; as explained later in Section III, this is materialized by a proposal for a fault management framework via cognitive network management functions.

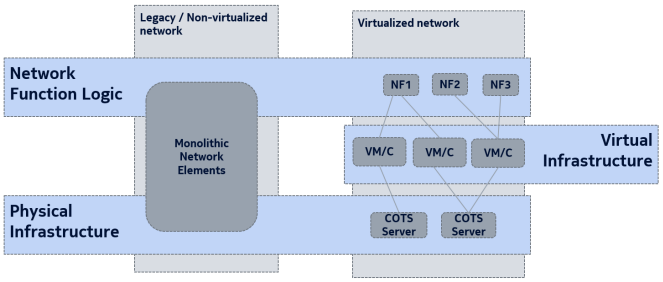

Fig. 1. Three deployment layers identified in virtualized networks: network function logic, virtual infrastructure and physical infrastructure [5].

The remainder of this paper is structured as follows. In Section II we give an overview on the Self-Organizing Network (SON) concept and its evolution towards Cognitive Network Management (CNM) which is used as a baseline for our work. Section III describes our proposed solution for fault management in slice-enabled 5G networks. Section IV illustrates some deployment options relevant for our solution, while our final concluding remarks are given in Section V.

II. Cognitive Network Management and Fault Management Cognitive Functions (FM CFs)

Network slicing imposes a clear need to adapt the network operation and management to the slice requirements and existing Service Level Agreements (SLAs) with the tenant (the entity that requests a specific network slice in order to offer services to its end customers). The slice-tailored network management framework implies slice-specific fault management and consequently slice-specific self-healing. The main goal of the slice-specific fault management is to provide the resilience to the network faults according to the actual slice requirements as requested by the tenant. Due to foreseen challenges with respect to network slices, e.g., shorter life cycle and variety of slice requirements, the according network and fault management procedures need to be automated as much as possible.

The first step in terms of automating the Operation, Administration and Maintenance (OAM) of mobile networks, and in particular, mobile radio networks, was the introduction of the SON solution [3], which allows the autonomic optimization of certain network configuration parameters based on measurements from the individual network elements. In order to enable a joint operation of a different SON function instances at the same time (one for each network slice instance), concepts for SON coordination and SON management have been introduced. However, a rather static nature of the logic of deployed SON function seems as unsuitable in the context of network slicing.

While SON management allows a modification of some parameters of a SON function such that the behavior of the SON algorithm can be slightly modified (and thereby its effects on the network configuration), the SON algorithm as such (including the algorithm inherent state machine and state transitions) remains unchanged. More sophisticated adaptations of the SON algorithms therefore need to be done manually through the SON manufacturer. Such manual intervention might not be acceptable for highly dynamic nature of 5G networks, thus the new solution that enable more automation in SON adaptation needs to be developed.

In this regard, the aim of Cognitive Network Management (CNM) [6] is to make the automation of OAM processes in mobile networks more flexible and adaptable to current network context. The main idea of CNM is to better extract the characteristics of the network environment so that it can decide on the most suitable configurations of network functions having in addition the information about current network states and failures. The CNM introduces the so-called cognitive functions (CF), which represent more intelligent SON functions that learn from historical data on network operation in different contexts. Furthermore, the CFs can be designed in a way that their logic can be adapted automatically by extending the knowledge space for a certain network context. Such a knowledge extension can be gained by applying the network setups that were not used before and learn from the corresponding network performance in a given context. Thus, the CFs of the CNM go beyond traditional SON solutions where each SON function merely matches combinations of KPIs to pre-configured network constellations.

Hereby we focus on the design of CFs that implement the Fault Management (including self-healing) operations for slicing-enabled 5G networks. We further refer to such functions as Fault Management Cognitive Functions (FM CFs). Depending on the slice/tenant requirements and priorities, criticality of individual network functions etc., the FM CFs need to be adapted. Furthermore, the interaction between different FM CFs at different deployment layers, subnets and network slices need to be carefully designed. In order to meet the stringent latency requirements, where applicable the troubleshooting should be done more locally/distributed, avoiding thus the case of transmission of all relevant data to hierarchically higher management entities and performing centralized data processing and troubleshooting.

III. Design of Fault Management Cognitive Functions

In order to enable adaptable, slice-aware fault management, we opt at designing the Fault Management Cognitive Functions (FM CFs) that can be mapped to the different network entities, functions or parts of the infrastructure. For instance, a single FM CF can be responsible for the fault management of entire Network Slice Instance (NSI) or building blocks of the network slice, e.g. Network Slice Subnet Instances (NSSIs), network function chains, individual network functions as well as individual deployment layers of network functions. The exact mapping between FM CF and the entity for which it is responsible as well as the logic/algorithm of Fault Management Cognitive Function (FM CF) are determined by the slice characteristics/requirements along with SLAs agreed with the tenant. Furthermore, the requirements of individual network functions can be taken into account. The parameters that drive the design of FM CF are as follows (non-exhaustive list):

- Network slice requirements, particularly in terms of resilience
- Information about established service level agreements (SLA) with the tenant
- Type/criticality of the network function (e.g. user or control plane, centralized network controller) along with its resilience requirements (e.g. in terms of required time for restoration in the case of fault)
- Affinity among network functions e.g. if NFs are usually appearing together in the network function chain or if the output of one function is the direct input of the other function
- Deployment characteristics of the network functions in terms of the mapping between physical, virtual and functional deployment layers

It is worth mentioning that various design options for FM CFs might be applicable based on the concrete use case. As an example, the critical network functions that are part of the ultra-reliable low latency communication (URLLC) slice might have dedicated FM CFs in order to perform faster troubleshooting and healing. The algorithms of such dedicated FM CFs need to be designed in a way to be extremely reactive to any anomaly in the NF operation. Optionally, in order to minimize the effect of other NFs to that critical NF, the critical NF might be implemented on a single VM, container or physical server (so that fault localization and isolation/self-healing might be facilitated more swiftly).

The above implementation approach might be taken by the network orchestration entity after receiving the feedback from the running FM CFs on the feasibility to perform the healing operations. On the other hand, in eMBB slice with less stringent resilience requirements, one FM CF might be responsible for larger scopes, e.g. for an entire slice or subnet(s). The deployment of NFs belonging to such slices/subnets can span across multiple VMs/containers and physical servers with many inter-dependencies among infrastructure layers. In such cases the fault localization and isolation might be more complicated and thus take longer time. Additionally, the algorithm of such FM CFs can be adapted to more relaxed resiliency and fault recovery requirements, the FM CF might be designed to react only on a specific set or number of alarms/events, e.g. only those that can seriously endanger the network operation and thus the E2E service fulfillment.

In general, as network slice (and consequently subnets and NFs) realization in virtualized environment can include existence of different deployment layers, (that is physical, virtual, and functional layers), the responsible FM CF needs to perform a consolidated fault management by jointly considering the inputs from all the deployment layers. Alternatively, the dedicated layer-specific FM CFs can be mapped to different deployment layers.

As mentioned above, based on handled FM events and ability to localize, isolate and resolve the occurred issue the FM CF can give a feedback to the network orchestration entity in order to improve the NF placement/deployment. An example is the case of very critical control network function which constituent NFs were initially deployed across multiple physical servers, yielding to difficulties in fault detection and isolation, the FM CF of that network function might recommend to the orchestration entity to deploy the NF differently. Similarly, in the case that very critical control function of URLLC slice shares the infrastructure, e.g. the physical server with another NF that can be prone to failure or security threads, the critical NF can be affected by the problems caused by other NF. The corresponding FM CF (e.g. responsible for both NFs) will provide feedback to the orchestration entity that such deployments should be avoided.

IV. Fault Management Cognitive Function (FM CF) deployment options

The network slice and its constituent subnets and NFs can have different realizations in virtualized environment, and the FM CFs can have different mappings to network slices, network functions and deployment layers. Figure 2 illustrates such different options for FM CF deployment. Figure 2 a) shows one example implementation of FM CFs responsible for managing three network functions as well as corresponding mapping to the virtual and physical infrastructure on which NFs are deployed.

In the above example, it is noticeable that due to multiple overlaps in the layers mapping (e.g., both NF2 and NF3 components are implemented on the same physical server) a single FM CF (e.g., FM CF1), in order to perform the fault localization and isolation in the physical infrastructure layer, needs to exclude the input that might be related to the NF3 which is not under the responsibility of FM CF 1. This comes in addition to correlating the performance indicators and alarms related to different layers of the infrastructure.

Alternatively, the FM CFs can be deployed as layer-specific FM CFs as illustrated in Figure 2 b), e.g. FM CF-1 PHY. In such a case their scope of operation is limited to a specific deployment layer (e.g. physical or virtual). Such layer-specific FM CFs can to a certain extent autonomously act within a dedicated deployment layer e.g. in the case of physical Network Interface Card (NIC) outage the physical layer FM CF can automatically trigger the switching of traffic to another NIC. However, the layer-specific FM CFs need to exchange the info between each other and/or with the FM CF of the network function to which the layer-specific FM CFs are related to in order to achieve consistent troubleshooting.

Based on the actual deployment of NFs and their mappings to FM CFs the fault healing can happen on different levels without affecting other running NF instances. Figure 3 illustrates the case where FM CFs can implement the fault management algorithms tailored to the NFs (a single NF or a group of NFs with a similar resiliency requirements) they are managing and self-healing can be done on virtual infrastructure level. This is possible since VMs/containers on which one NF is deployed can be easily isolated without affecting other NFs. E.g. if the fault happens on the virtualization deployment level of NF1 (i.e. VM/Container1 or VM/Container2) the self-healing will be done on this level without affecting the NF2 and NF3.

Similarly, Figure 4. illustrates the case where the FM CFs can implement NF-tailored FM algorithms and isolation can be

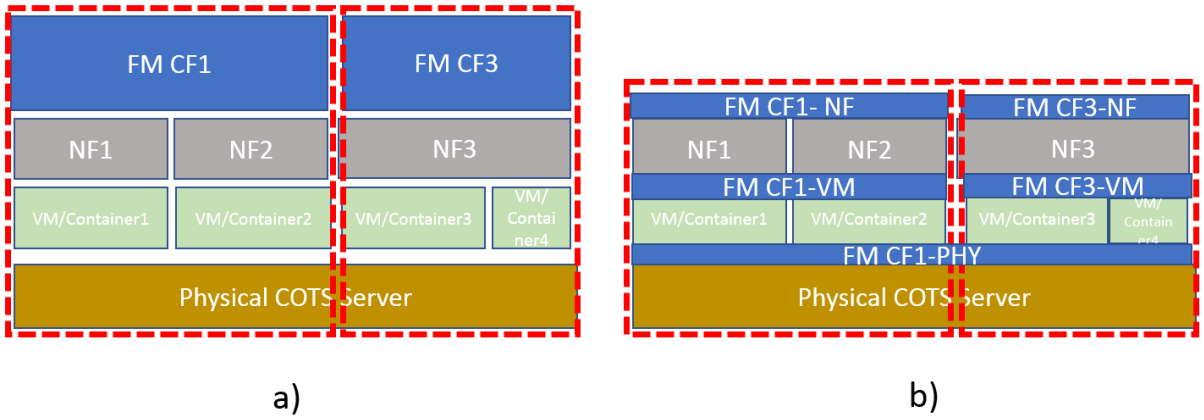

Fig. 2. Example deployment of FM CFs. a) an example of FM CF deployed at functional deployment layer, b) an example of FM CFs deployed on individual deployment layers.

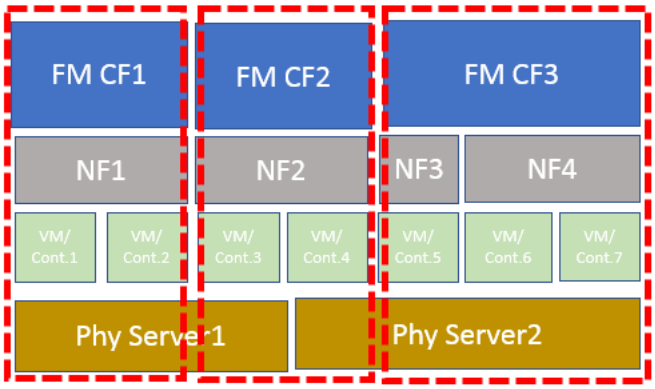

Fig. 3. Mapping of FM CFs to network functions (NFs) and infrastructure components and deployment layers where isolation is possible on virtual infrastructure deployment layer.

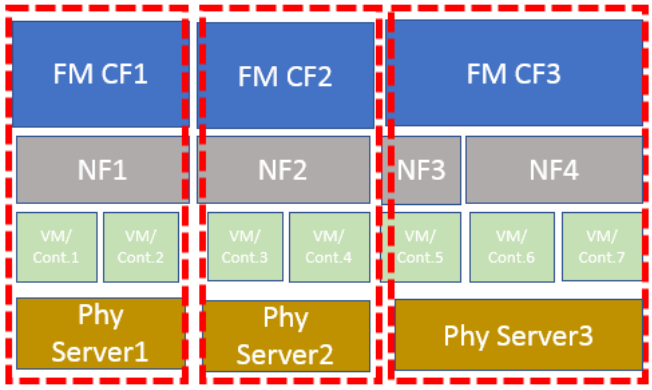

Fig. 4. Mapping of FM CFs to network functions (NFs) and infrastructure components and deployment layers where fault isolation is possible on virtual and physical infrastructure deployment layer.

done on physical infrastructure level as well, e.g., self-healing or isolation of NF2 either on virtual of physical level will not affect NF1 and NF3.

The different implementation approaches can be used for realization of FM CFs namely, a distributed, centralized or hybrid implementations can be suitable. However, certain advantages and disadvantages are bound to every implementation option, especially in terms of ability to detect and isolate faults locally using NF-specialized algorithms, as well as the need to coordinate the operation with other instances of FM CFs. For example, the advantage of distributed (NF-specific) implementation of FM CFs is more efficient handling of FM events both in terms of applying the algorithms suitable for specific type of network function (or a group of NFs with same/similar resilience requirements) and its implementation/deployment as well as in terms or more local processing of fault event notifications and thus faster reaction to faulty events.

It should be noted, however, that the distributed, highly NF-specific FM CFs cannot work independently from each other due to many dependencies among network functions and overlapping deployments in the underlying infrastructure which might lead into contradicting decisions at the FM CFs. In other words, the distributed FM CFs do not have enough visibility on the overall situation in the network in order to act completely independent from one another. Therefore, a highly distributed implementation of FM CFs would require either very close interaction between FM CFs or the existence of coordination entity for consolidation of FM CFs operation and decisions. This means that the corresponding interfaces between FM CFs and coordination point as well as among FM CFs need to be defined in order to allow for exchange of data.

In a more centralized implementation of FM CF concept (e.g., by merging the different FM CFs into a single one, or assigning a single FM CF to entire network slice) such interaction becomes less critical as decisions are taken centrally at one single entity. As a result, such single entity has better visibility on overall network states and can perform decisions based on such knowledge. However, the responsiveness of such implementation might become lower, as all the information needs to be collected and processed centrally.

Due to trade-offs between distributed and centralized approaches neither a fully distributed nor a fully centralized

implementation for FM CFs are optimal. The actual level of centralization/distribution of FM CFs depends on the use case, e.g. slice characteristics in terms of required reliability, responsiveness to faults etc.

V. Conclusions

In this paper we present the novel concept of Fault Management Cognitive Functions (FM CFs) enabling the 5G slice specific network fault management. This concept takes into account the information about slice requirements, agreed SLAs, and slice/subnet/NF deployment characteristics for performing tailored automatic fault detection, diagnosis and resolution. FM CFs aim at joint treatment of all deployment layers of virtualized networks, that is physical, virtual and logical, in order to achieve a consistent fault troubleshooting and perform most optimal healing actions. The presented approach provides means for dynamic, adaptable and use-case tailored fault management in next generation slicing enabled virtualized networks.

References

[1] ETSI NFV, *Network Functions Virtualisation (NFV): Network Operator Perspectives on Industry Progress*, White Paper, 2014.

[2] Nokia Networks, *5G use cases and requirements*, White Paper, 2014.

[3] S. Hämäläinen, H. Sanneck, C. Sartori, *LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency*, John Wiley and Sons, Dec. 2011.

[4] S. Novaczki, *An Intelligent Anomaly Detection and Diagnosis Assistant for Mobile Network Operators*, 9th International Conference on the Design of Reliable Communication Networks (DRCN), Budapest, Hungary, June 2013

[5] D. S. Michalopoulos, B. Gajic, B. G-N. Crespo, A. Gopalasingham, and J. Belschner, "Network Resilience in Virtualized Architectures", Interactive Mobile Communication Technologies and Learning, pp.824-839, February 2018

[6] S. Mwanje, G. Decarreau, C. Mannweiler, M. Naseer-ul-Islam, C. Schmelz, *Network Management automation in 5G: Challenges and opportunities*, 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain, September 2016