Cognitive Network Management for 5G

The path towards the development and deployment of cognitive networking

Cognitive Network Management for 5G

The path towards the development and deployment of cognitive networking

By 5GPPP Working Group on Network Management and QoS

Version:	1.02
Date:	9-March-2017
Document Type:	Final
Confidentiality Class:	P - Public

Project:	5GPPP Phase 1
Editors	Robert Mullins/ Michael Taynnan Barros, Waterford Institute of Technology
Contributors:	5GPPP Phase 1 Projects
Approved by / Date:	

List of Contributors

Name	Company / Institute / University	Country
Editorial Team		
Barros Michael Taynnan	Waterford Institute of Technology Network Management & QoS WG Chair	Ireland
Mullins Robert	Waterford Institute of Technology	Ireland
Contributors		
Barros Michael	Waterford Institute of Technology	Ireland
Belikaidis Ioannis-Prodromos	WINGS ICT Solutions	Greece
Casetti Claudio	Politecnico Di Torino	Italy
Costa Luciana	Telecom Italia	Italy
Demestichas Panagiotis	University of Piraeus	Greece
Galis Alex	University College London	UK
Grida Ben Yahia Imen	Orange	France
Klaedtke Felix	NEC Labs	Germany
Martinez Perez Gregorio	University of Murcia	Spain
Mullins Robert	Waterford Institute of Technology	Ireland
Neves Pedro Miguel	AlticeLabs	Portugal
Pérez-Romero Jordi	Universitat Politècnica de Catalunya	Spain
Siddiqui Shuaib	Fundacio i2CAT	Spain
Wary Jean-Philippe	Orange	France
Weigold Harald	Rohde Schwarz	Germany
Reviewers		
Al-Dulaimi Anwer	Brunel University	UK
Barattini Paolo	Kontor46	Italy
Behrmann Malte	BBW Hochschule	Germany
Bihannic Nicolas	Orange	France
Bouras Christos J.	University of Patras	Greece
Foster Gerry Foster	University of Surrey	UK
Guidotti Giovanni	Leonardo	Italy
Leguay Jérémie	Huawei	France
Mascolo Saverio	Politecnico di Bari	Italy
Papazois Andreas	GRNET S.A.	Greece
Rodriguez Juan	ISOIN	Spain

Table of Contents

Table of Contents

Summary

5G represents a complete revolution of mobile networks for accommodating the over-growing demands of users, services and application. In contrast to previous transitions between mobile networks generations, in 5G there will be a much complex management requirements based on the softwarization of network resources. This ultimately will lead to a system that requires real-time management based on a hierarchy of complex decision making techniques that analyse historical, temporal and frequency network data. Cognitive network management has been proposed as the current solution for this problem, in which the use of machine learning to develop self-aware, self-configuring, self-optimization, self-healing and self-protecting systems will enable cognitive network management. This technology is needed for managing a demanding infrastructure but one that yet has to present scalability and flexibility, such as that needed in 5G. In this paper, the novelties for network management in 5G are presented, including: autonomicity, NFV, SDN, network slicing, architectures, security and KPIs. All these points are also explained in the context of the current development of network management solutions within the 5GPPP phase 1 projects, including: CogNet, Selfnet, SONATA and 5GeX. The development of such novelties will pave the way for not only the future of cognitive network management, but for 5G and also the future mobile network generations.

1. Introduction

5G Network Management is a non-trivial endeavour that faces a host of new challenges beyond 3G and 4G, covering all radio and non-radio segments of the network. The number of nodes, the heterogeneity of the access technologies, the conflicting management objectives, the resource usage minimization, and the division between limited physical resources and elastic virtual resources is driving a complete change in the methodology for efficient network management.

In the past, a distinction was typically made between the control and data plane of the network. However, the model of the 5G networks can be expanded in terms of a “Service and Softwarisation plane”, where the management of the network services and the virtualised devices is an integral part of the overall network. This model can be used for extending the idea of network management to the reliance on an increased overall capacity of computational resources to create a robust solution.

Historically, autonomic management has gone as far as developing complete automated solutions into the network. The concept of "the selves" was introduced, in which network management is expressed through a mixture of the approaches including: *self-awareness*, *self-configuration*, *self-optimization*, *self-healing* and *self-protection*. With the advancements of the infrastructure technology for accommodating the next generation of networks, the next level of network management has to incorporate the flexible manipulation of network resources and leverage it with the number of users, the network traffic, the SLAs, and the demanded system performance.

Fig. 1 presents the vision of this paper, which is elevating the level of cognitive abilities in the “selves“ using **machine learning**. Machine learning has the capability of adapting an entire system based on historic data, which means that in 5G, the network management will monitor key metrics with the network, understand the configurable parameters and optimally adjust their values for achieving a superior network configuration, indicated through a set of key performance indicators. In the end, **cognitive network management** is introduced as the next generation of network management and key driver of the 5G success.

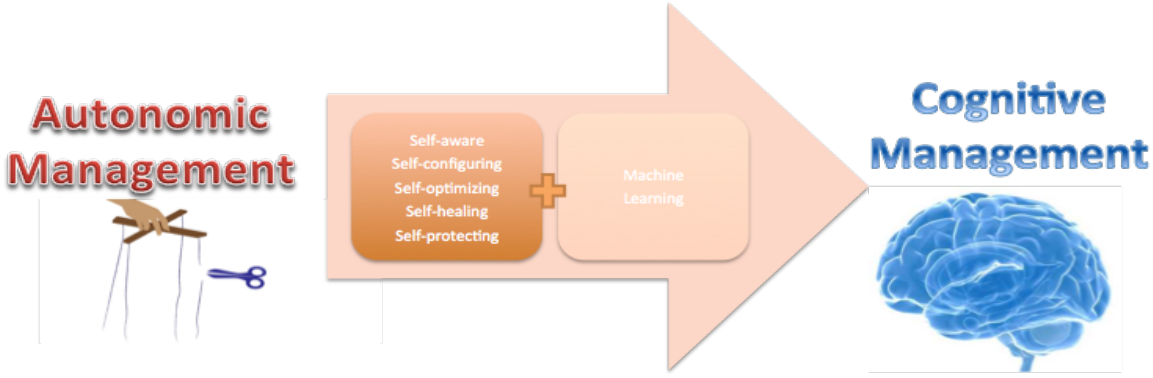

Figure 1: Evolution from autonomic network management to cognitive network management.

On the other side of management, the **orchestration** of the network regulates network resources based on its management decisions. Optimisations and trade-offs can be made, adapting the network over time through self-configuration and self-optimisation, self-healing and self-protection.

The orchestration of the network can be thought of as resolving a number of independent and in some cases interdependent management objectives across a number of key objectives such as:

1. *Provisioning***:** ensure that the network is adequately provisioned with resources sufficient to deal with current demand levels while maintaining QoS at an agreed level.
2. *Security Management:* Protect network data and its performance through accurately detecting intrusion, privacy and denial of service as well as autonomous anomaly detection.
3. *QoS support:* Network Slicing supports several defined QoS levels simultaneously and kept logically isolated by the same physical network.
4. *Fault Tolerance:* The network should be able to recognise emerging faults or error conditions and pre-emptively deal with them, or intercept unexpected faults or errors as quickly as possible to minimise any reduction in QoS.
5. *Energy Efficiency*: Optimising the source of energy, for example maximising the use of renewables and energy efficiency may be a factor in the selection of computing resources to support network functions.
6. *New components*: The operator should also be able to add new objectives and to change their priorities as well.

The challenge is in deploying the cognitive network management and its orchestration across multiple heterogeneous networks all of which have their own peculiarities and requirements, including: *Radio & Other Access Networks*, *Core & Aggregation*, *Edge Networks*, *Edge and Computing Clouds*, and *Satellite Networks*. The developed management technology has to meet such multiple party requirements in addition to being easily deployed, all of which will require much effort from industry to successfully achieve.

The above non-exhaustive lists of management objectives, network planes and network parts can be modelled as a three-dimension plane as in Fig. 2. Each of the entries represents a potential management entity or consideration, so in the above example, 80 separate management entities would be needed to support this. However, in practice the management across several entities may be combined. The orchestrator will provide management across all entities ensuring seamless and reliable network operation as well as the high quality user experience required by the many services.

To further the challenge, a means must be found to achieve the above in real time. The level of computation required for real time management is too expensive to be neglected at the development stage, considering also other related costs such as energy and equipment. A potential solution for real time management is the use of mathematical models to aid such real time decision making, while the models are computed offline but are used in real time and updated in near real time.

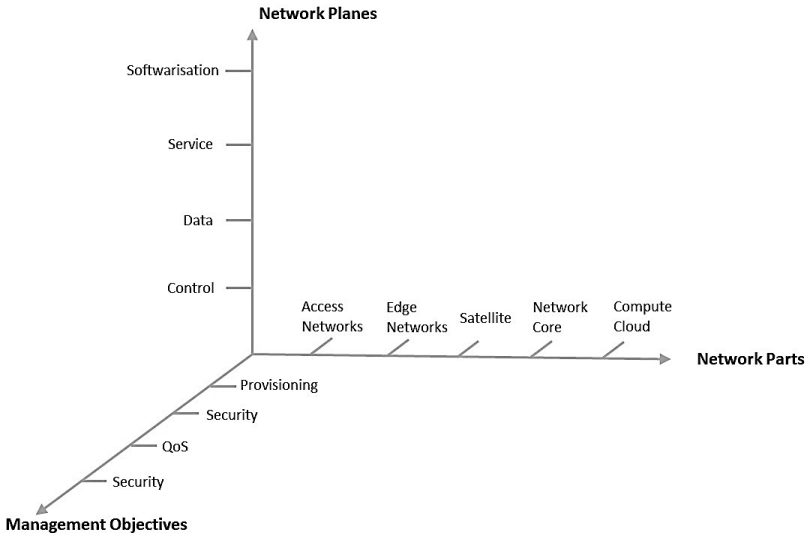

Figure 2: Plane across network planes, network parts and management objectives.

The presented document tries to introduce the vision of cognitive network management based on the 5G requirements and solutions, as well as extensively analysing the challenges briefly discussed previously. A cognitive network management architecture is presented, demonstrating the potentials of how this new level of management can be achieved. Lastly, new metrics are defined for capturing essential information about 5G performance based on the new direction mobile networks are moving towards to. All these points are also explained in the context of the current development of network management solutions within the 5GPPP phase 1 projects, including: CogNet, Selfnet, SONATA and 5GeX.

2. New Requirements for Network Management based on 5G

2.1 Network (Virtual) Functions

5G networks represent a shift in networking paradigms: a transition from today's "network of entities" to a "network of functions". Indeed, this "network of (virtual) functions (NVF)", resulting, in some cases, in the decomposition of current monolithic network entities will constitute the unit of networking for next generation systems. These functions should be able to be composed on an "on-demand", "on-the-fly" basis. In fact, a research challenge for managing Network (Virtual) Functions consists in designing solutions which identify a set of elementary functions or blocks to compose network functions, while today they implemented as monolithic. In addition uniform management and operations for NVF are becoming part of the dynamic design of software architectures for 5G.

2.2 The Network as a Service

The concept of the Network as a Service is essentially the killer use case for 5G and is enabled through a technology known as Network Slicing. The slicing concept has been recently introduced for the upcoming 5G mobile networks and it is considered to be an integral part of 5G [1]. It is one of the main enablers for 5G security. A network slice in the context of 5G consists of a collection of 5G network functions and specific Radio Access Technologies (RAT) that are combined together for a specific use case or business model [2]. In other words, a network slice is a logical instantiation of a network, with all the needed functionalities that the network needs in order to operate. Network slices can be considered more as networks on-demand, which will be created, deployed and removed dynamically. Ultimately, with network slicing it is possible to guarantee a certain level of quality and security to an application or a service.

2.3 Scalability

Scalability refers to the elasticity of the network to expand its capacity to meet the variability of services demand over time and location. A scalable network will always have sufficient capacity to deal with this service demand while not maintaining excessive unused capacity. While the traditional approach to this was to overprovision a network with resources to meet peak anticipated demand, this approach was wasteful. The sheer scale of 5G networks will mandate the conservation of resources and the ability to quickly adjust capacity through the scalability of infrastructure and control software and this capability will be mandatory for 5G technologies.

2.4 Quality of Service

Quality of Service (QoS) is a measure of the reliability and performance of the network's connections, particularly as perceived by the users on the network. QoS is a composite metric as it is based on a number of values that indicate the characteristics of the network transmission, and consequently reductions or improvements in QoS can be brought about through a number of factors.

QoS is very important for 5G Telecommunications and computer networking from a business perspective as different application types have their own QoS requirements, and various types of end users may also have specific requirements for QoS levels. Examples of applications that require high QoS include rich VoIP and video streaming. This is because any perceptible delay in transmission or lost packets reduces the Quality of Experience (QoE) for the user using the application. Augmented Reality, Virtual Reality and Vehicle-to-Vehicle applications specifically require very low latency communications. In the business world, customers may pay for a Service Level Agreement (SLA) that contractually defines the QoS that they expect for their connections and consequently their applications. End users will often pay a premium for higher levels of QoS and for this reason, QoS management is a key requirement for current and future networking technology, as it manages the process of guaranteeing levels of QoS to different applications and users simultaneously.

Network Slicing depends very highly on the application of and maintenance of QoS levels according to the parameters of the particular defined slices. These QoS levels will have to be maintained simultaneously on the same physical network and potentially using common virtual infrastructure. Designing and ensuring the correct operation of this will be one of the principal challenges for 5G Network Management.

2.5 Flexibility

Not so long ago, assembling and running a network would require designers, managers and providers to deal with “black boxes”, essentially pieces of hardware which, to some extent, implemented one or more functionalities typical of a specific network layer. These boxes are perfectly apt at supporting the data plane of a network: switching and routing packets at multi-Gb/s speeds, filtering content based on complex rules, contending and accessing busy shared channels. However, more often than not, they became true bottlenecks as far as the management and control plane are concerned. Fundamentally lacking real abstractions to make their task easier, network managers had to literally slug their way through a maze of protocols and network operating systems, none really designed with interoperability as a primary concern.

The availability of faster chips and the advances in virtualization techniques have since revolutionized the black box approach, bringing about the era of software virtualization, which in the case of networking translates into the realms of Software Defined Networking (SDN) and Network Function Virtualization (NFV). While SDN allows the creation of network abstractions, NFV consists in the virtualization and insulation of network functions (such as switching, firewalling, packet inspection, caching…) that become independent of the infrastructure they run on and the resources (computation, storage, and networking) they need. Although SDN can enhance the performance of NFV, ease its compatibility with existing deployments, and facilitate operation and maintenance procedures, SDN and NFV do not strictly require each other.

Virtualized network functions and their organic interaction (or chaining) concur in defining new virtualized network services that require a novel, ad hoc MANagement and Orchestration (MANO) framework. ETSI is at the forefront of the definition and standardization of such a framework through its NFV MANO working group, but many open source projects are in advanced deployment stages, as will be discussed in the next sections.

2.6 Sustainability

By monitoring the energy parameters of Radio Access Networks, fronthaul and backhaul elements, the VNFs supporting the internal network processes, and through estimating energy consumption and triggering reactions, the energy footprint of the network (especially backhaul and fronthaul) can be reduced while maintaining QoS for each VNO or end user. An Energy Management and Monitoring Application can be conveniently deployed along a standard ETSI MANO and collect energy-specific parameters like power consumption and CPU loads (see Figure 1). Such an Energy Management and Monitoring Application can also collect information about several network aspects such as traffic routing paths, traffic load levels, user throughput and number of sessions, radio coverage, interference of radio resources, and equipment activation intervals. All of this data can be used to compute a virtual infrastructure energy

budget to be used for subsequent analyses and reactions using machine learning and optimization techniques.

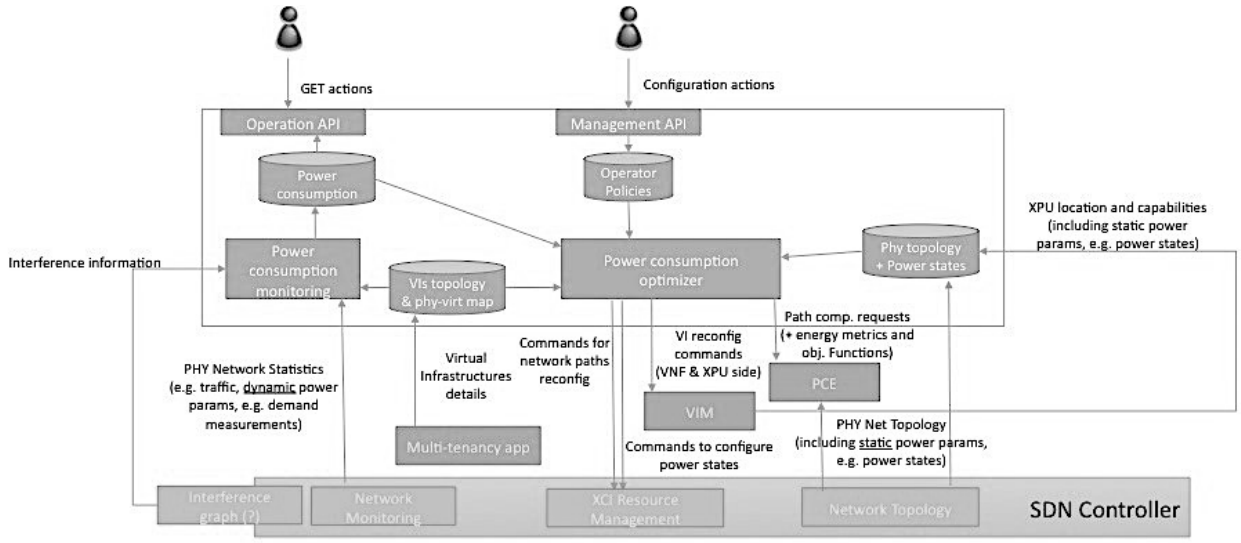

Figure 3: Functional description of an energy management and monitoring application

The application can optimally schedule the power operational states and the levels of power consumption of network nodes, jointly performing load balancing and frequency bandwidth assignment, in a highly heterogeneous environment. Also the re-allocation of virtual functions across backhaul and fronthaul will be done as part of the optimization actions, in order to move virtual network functions to less power-consuming or less-loaded servers, thus reducing the overall energy demand from the network. Arguably, such actions can achieve the target KPI of 10 times lower energy consumption.

Three main issues are to be addressed regarding energy efficient communications networks:

1. Environmental concerns, where finite resources increase harmful emissions
2. Operating costs, which telecommunication providers seek to reduce in order to offer more competitive services to their customers
3. High performance of the network itself.

Historically, improving hardware efficiency helped increase energy efficiency at device and infrastructure levels in mobile communications. Such gradual hardware advances will not reduce energy consumption sufficiently for 5G, given the expected increase in number of devices, data rates, and coverage. The hardware-based approach fails to address issues (2) and (3) above. A software-based approach offers a better solution to improve overall network management. Additionally, such an approach better fits the proposed 5G architecture, where data and control planes are decoupled. This software-based approach will be localized in the control plane of the 5G network. The infrastructure resources will be adjusted according to

- ❖ energy policies,
- ❖ QoS requirements of the data being transferred, and
- ❖ network resource conditions;

thereby, addressing issues (1), (2) and (3) outlined above

2.7 Context Awareness

Context awareness allows the network to adapt itself to changing external environmental conditions. These can be areas such as:

- Changing levels of demand, both predicable changes such as day of week or time of day, or temporary, non-typical changes in demand in specific areas, for example during a rock concert located in a football stadium, which causes a spike in demand in a specific location.
- Weather which can change the availability or cost of energy and which can feed into the availability and pricing of services, or the location of the computing resources used to provide networking functions.
- The emergence of threats such as a new type of virus or fraud which may threaten the network resources
- The network needs to be aware of the potential impact of external events. For example in the smart cities use case, it may need to assess the impact of some civil works on customer mobility in a proactive way. Network management can then benefit from interaction with such external platforms from vertical markets (e.g. Web of Things platforms like Fi-ware [3] as sponsored by the European Commission)

Developing such context awareness in the network to allow the network self adapt will be a major challenge for 5G but will greatly improve robustness and ultimately Quality of Service.

2.8 Security

An interesting security concept proposed for 5G and which aims to address some of the shortcomings in current network technology is Micro-Segmentation. This is a new security feature that has been introduced in data centres [4], but its use in mobile networks has not yet been considered. In data centres, the traditional security model is to regulate the north-south traffic at the edge of the data centre. This means that there is a single firewall at the perimeter: all incoming traffic to the data centre is considered untrusted and traffic inside is considered trusted. Consequently, once attackers gain access through the firewall at the perimeter, they are free to move and carry out their attacks. Micro-Segmentation aims to get rid of the single point of failure in data centre security by also taking into account the east-west traffic in the data centre, i.e., monitoring also the traffic inside the data centre and is generally considered an enabler for the Software Defined Data Centre. The 5GPPP project 5G Ensure does a thorough analysis of the role of Micro-Segmentation in 5G [5].

In the context of 5G, micro-segments can be considered as isolated parts of the 5G network dedicated for particular application services or users. Compared to network slices, micro-segments can provide more fine grained isolation and segmentation, specific access controls and stricter security policies. The mobile network is generally divided into smaller parts, where each unique micro-segment can have its own security controls defined, and services delivered. Only authenticated devices and network services can join the micro-segment and traffic inside the micro-segment should also be monitored. A micro-segment instance is not necessarily required to form a complete logical network. Other security issues are raised in the context of 5G and NFV like: the level of openness for management entities to be granted to vertical partners, how to manage security responsibility between all stakeholders involved in the VNF software delivery (whose lifecycle should be strongly shortened)

2.9 Open Management

There will be a requirement to grant access to parts of the network management for vertical stakeholders that use dedicated slices. This will lead to security issues being raised in the context of 5G and NFV like: the level of openness for management entities to be granted to vertical partners, how to manage security responsibility between all stakeholders involved in the VNF software delivery (whose lifecycle should be strongly shortened).

2.10 Current standards

We will briefly outline the ETSI MANO reference model as detailed in [6], although single vendors may provide slightly different views of the ETSI model. Essentially, the ETSI MANO includes three functional blocks, as shown in the right side of Figure 1:

- The Virtualized Infrastructure Manager (VIM)
- The VNF Manager (VNFM)
- The NFV Orchestrator (VNFO).

Notwithstanding the order in which they are listed, no functional block is more important than the others and each leverages services offered by other functional blocks.

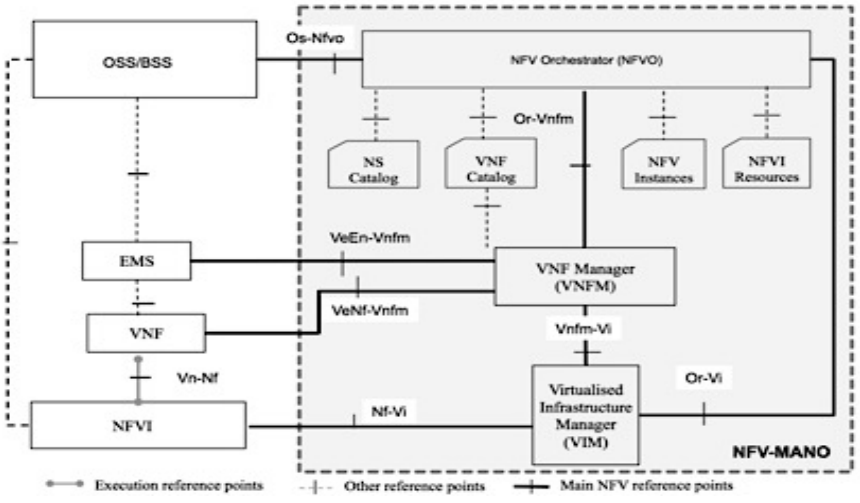

Figure 4: ETSI functional MANO architecture

The *Virtualized Infrastructure Manager* controls and manages the computing, storage and network resources in one domain of an NFV Infrastructure (NFVI). It is responsible for the life cycle of virtual resources by creating, managing and tearing down virtual machines (VMs), and maintains an inventory of which VMs are associated with which physical resources. Crucially, it exposes northbound APIs that allow other management systems to access physical and virtual resources. Its southbound interfaces interact with Network Controllers in order to perform the functionality exposed through its northbound APIs.

Instances of VNF installed on the VMs managed by the VIM are, in their turn, created, managed and torn down by the *VNF Manager*. The VNFM configures, updates, monitors the performance and trims CPU usage of VNFs. The deployment and operations of each VNF are captured in a template called Virtualised Network Function Descriptor (VNFD), stored in a VNF catalogue. For instance, VNFD describes the hardware resources needed for portability of VNF instances in multi-vendor environments.

A network PoP (Point of Presence) may include multiple instances of VIMs and VNFMs, and, in general, an operator needs to access and coordinate the resources exposed by different VIMs and instantiate the Network Services using VNF controlled by different VNFMs. These tasks are made possible by the *NFV Orchestrator*. The NVFO thus provides services that access resources in an abstract manner independently of any VIMs, and invokes VNF instances by coordinating with the appropriate VNFMs. The templates of any Network Service accessible through a NVFO are collected in a NS Catalogue, exposed through NVFO interfaces. Likewise, the NVFO can also expose VNFs in the VNF catalogue.

One challenge for a MANO-like system will be is ability to either address all those specific and complex policies (e.g. QoS handling, security, energy saving) in a single product, or to orchestrate a suite of managers that are each specialized per type of requirement.

3. 5G Service & Network Management and Orchestration

One of the main design objectives in 5G is to increase the flexibility and programmability networks with a novel Service & Network Platform and Orchestrators, and a novel Integrated and Uniform Infrastructure Management. This will maximize the predictability, efficiency, security, and maintainability of operational processes. As such bridging the gap between telecom business needs and operational management systems could be effectively realised. The expected key functionality and systems are represented by the Service Development Kit, the Management System and the Service Platform, including: a customizable Service Orchestrator, a Resource Orchestrator, a Service Information Base, and various Enablers as represented in Figure 3. The figure also shows the heterogeneity of the physical resources underlying the 5G infrastructures and related 5G network segments: radio networks, access networks, aggregation networks, core networks, software networks, data centre networks and mobile edge computing clouds. The Multi-Service Management is responsible for the creation, operation, and control of multiple dedicated communication network services running on top of a common infrastructure. Functionality for this plane includes: infrastructure abstraction; infrastructure capability discovery; catalogues and repositories; a large number of service and resource orchestration functions such as plugins; information management functionality; and enablers for automatic re-configuration of running services. Specific Service Management functionality includes DevOps functionality: Catalogues, Monitoring data analysis tools, testing tools, Packaging tools, Editors and primitives for Application & Service programmability. Figure 5 depicts the way in which 5G manages various underlying systems.

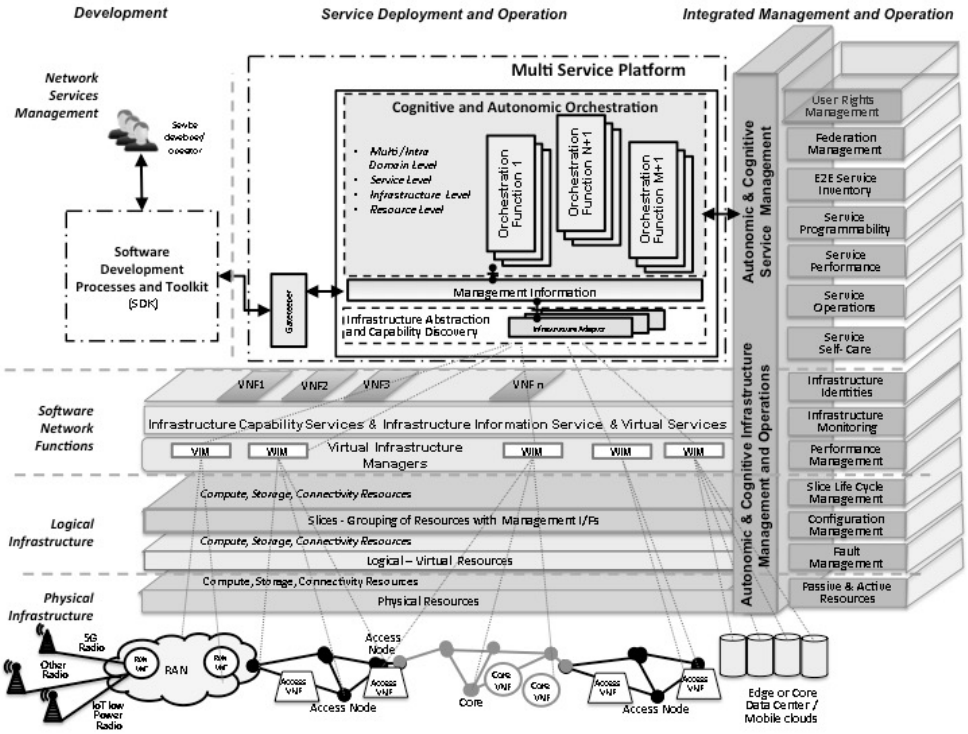

Figure 5: 5G Service & Network Management and Orchestration – SONATA Project

4. Multi-Domain Network Management and Orchestration

Multi-domain management and Orchestration refers to the automated management of services and resources in multi-technology environments (multiple domains involving different cloud and networking technologies) and multi-operator environments (multiple administrative domains) that includes operation across legal operational boundaries. The scope of the end-to-end multi-domain management and orchestration systems involves diverse concepts summarized in Figure 6. It represents the reference framework for organizing the components and interworking interfaces involved in end-to-end management and orchestration in multi-domain environments. At the lower plane there are resource domains, exposing resource abstraction on interface *I5*. Domain orchestrators perform resource orchestration and/or service orchestration exploiting the abstractions exposed on I5 by resource domains.

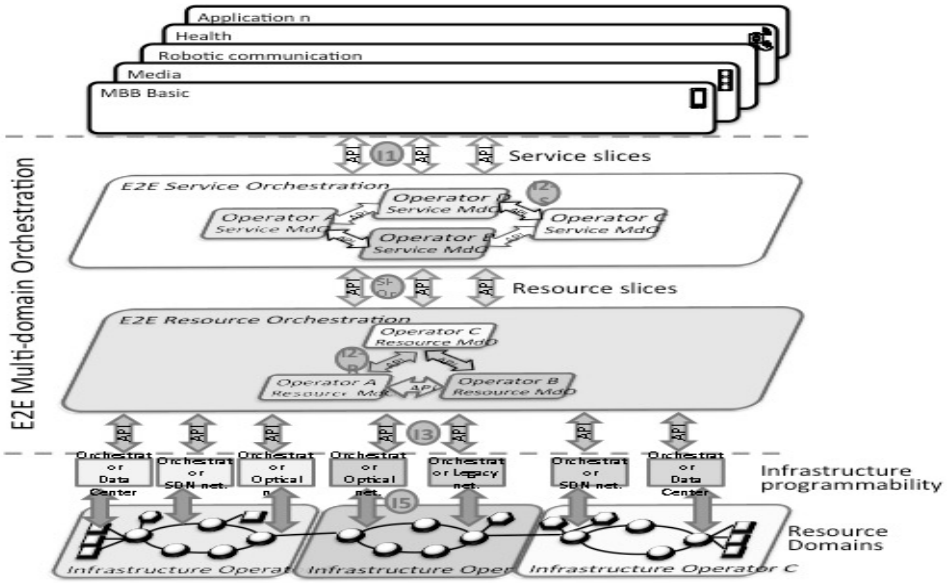

Figure 6: Multi-Domain Network & Service Management and Orchestration – 5GEx Project

A Multi-domain Orchestrator (MdO) coordinates resource and/or service orchestration at multi-domain level, where multi-domain may refer to multi-technology (orchestrating resources and/or services using multiple domain orchestrators) or multi-operator (orchestrating resources and/or services using domain orchestrators belonging to multiple administrative domains). The Resource MdO belonging to an infrastructure operator, for instance operator A, interacts with domain orchestrators, via interface *I3* APIs, to orchestrate resources within the same administrative domains. The MdO interacts with other MdOs via interface I2-R APIs (business-to-business or “B2B”) to request and orchestrate resources across administrative domains. Resources are exposed at the service orchestration level on interface SI-Or to Service MdOs. Interface *I2*-S (B2B) is used by Service MdOs to orchestrate services across administrative domains. Finally the Service MdOs expose, on interface *I1*, service specification APIs (Customer-to-Business or “C2B”) that allow business customers to specify their requirements for

a service. The framework also considers MdO service providers, such as Operator D in Figure 6, which do not own resource domains but operate a multi-domain orchestrator level to trade resources and services.

5. Autonomous Network Management

Autonomic network management (ANM) was developed to introduce self-governed networks for pursuing business and network goals while maintaining performance. Flexibility is a further advantage of autonomic network management, and aligned with network technology, has paved the way for the network infrastructure that is found today.

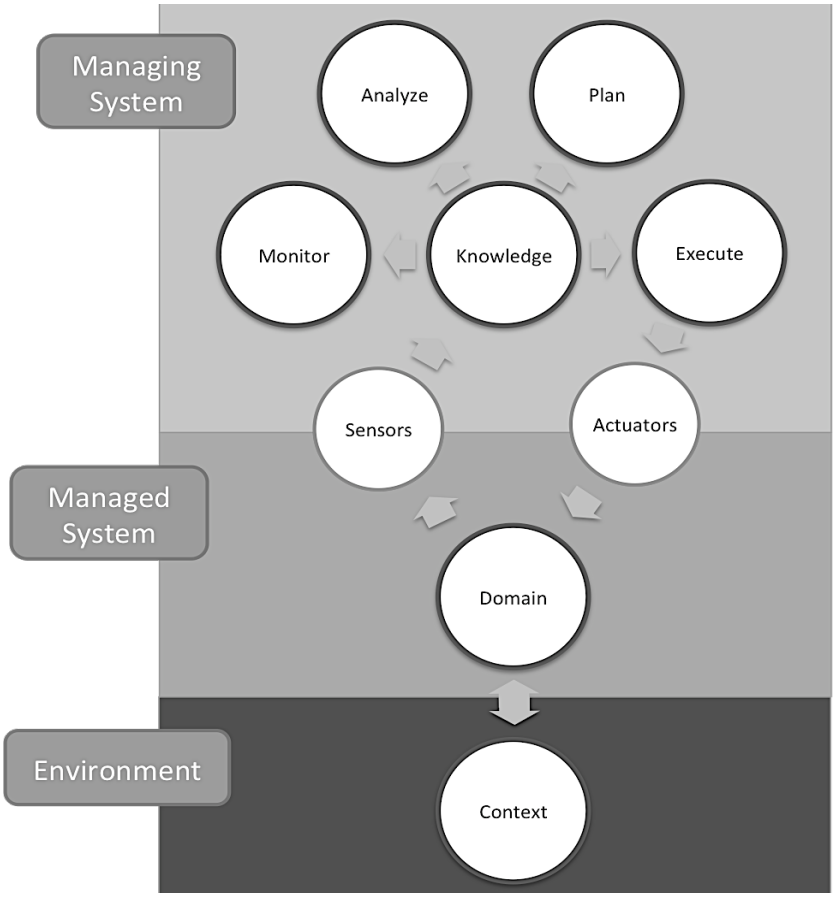

Figure 7: MAPE-K architecture.

IBM introduced building block-type architecture for guiding developments of autonomic network solutions. The Monitor-Analyse-Plan-Execute over a shared Knowledge (MAPE-K) is a control theory-based feedback model for self-adaptive systems. Fig. 7 presents the MAPE-K flowchart. The environment has full-duplex communication with a managed system that is controlled by managing systems. Sensors are used to gather data from the managed system, which is modified by actuators. The gathered data is used to monitor the managed system, which is being analysed further. Then the planning and execution pass the new actions to the actuators.

This feedback loop is fully tied to the knowledge base that is cross-linked to all other building blocks, serving as the local network information database.

The MAPE-K model can be expanded, as depicted for deeper understanding. The addition of ontologies and DEN-ng modelling enable enhanced capturing of network dynamics. A context manager is important for normalizing the information that is obtained in multiple domains. This whole process is comprehensible for building a knowledge base, and therefore, considerably improving the performance of the remaining building blocks. In the following, more detail about each building block is provided for a deeper understanding of autonomic network management and the MAPE-K.

5.1 Machine Learning

Machine Learning is one of the key technologies to be used in Cognitive Network Management to facilitate *self-adaption* - the adjustment of behaviour in response to the perception of the environment and the data that has already been processed. Adaptation is the “intelligent function” that can automatically select different functionalities by using different software components. Such functions cannot be based entirely on predefined handcrafted rules since predicting future working conditions caused by changes in the environment is too complex, and an automatic mechanism is needed.

ML utilises the power of Big Data and computing resources to look for patterns in historic data, and then uses these patterns as predictor functions when analysing future data. ML is typically based on using mathematical algorithms or statistical methods to analyse data sets, and using the results to modify how the software works. There are many variations of ML approaches, the two main being supervised or unsupervised. With Supervised ML, preexisting classifications are used to train a predictor function. Unsupervised ML does not use pre classified training data, rather it uses techniques such as data clustering to distinguish different conditions or situations.

5.2 Knowledge base

Building a representative knowledge base for network management is essential for its success. As previously depicted in Fig. 6, network information is shared across the whole MAPE-K architecture, and consultation and updates are often required from all building blocks to the network information server. Many approaches can be used to build knowledge of the network and its topology, including models from learning and reasoning, ontology and DEN-ng models. Even though the design of a proper knowledge base can involve multiple parties for different purposes, capturing structure knowledge, control knowledge and behaviour knowledge of the network requires an integrated solution.
Typically, a knowledge-based framework processes input data from multiple sources and extracts, through learning-based classification, prediction and clustering models, relevant knowledge to drive e.g., the decisions made by 5G Self Organizing Network (SON) functionalities such as self-planning, self-optimization and self-healing. Knowledge models can be categorized as cell-level, user-level and cell clustering [7].

5.3 Autonomic Monitoring

Collecting proper network information for building the desired knowledge base is the major challenge for identifying the overall status of the network. Accuracy, integrity and security of this information are the key factors to the success of the network management solution. Currently, the approaches found to date are: active, passive, centralized, distributed, granularity-based, timing-based and programmable. Regardless of the approach, the true lessons learned from the research built in this area rely on monitoring the right points on the network.

5.4 Autonomic Analysis

Analysing the obtained network data allow further additions to the knowledge base of the network, which is highly important for introducing concepts such as network information anticipation. Many approaches already exist mainly relying on probability and Bayesian models for knowledge anticipation, timing anticipation, mechanism anticipation, network anticipation, user anticipation, and application anticipation. The main challenge is to define a concentrated data set that comprehensively captures information across all anticipation points. Recent solutions rely on the usage of learning and reasoning to achieve such specific ends.

5.5 Autonomic Planning and Execution

The main objective of ANM is achieving network adaptation through governing the network resource, i.e. planning and executing a set of actions for dynamically adjusting the network. For that, the design of a network adaptation plan needs to address the following dimensions: knowledge, strategy, purposefulness, degree of adaptation autonomy, stimuli, adaptation rate, temporal scope, spatial scope, open/closed adaptation and security. The adaptation solutions differ broadly and there is no unanimity in defining proper planning and execution guidelines. However, it has been suggested that the most successful approaches rely on distributed network management solutions that are based on either evolutionary computing or feedback systems.

6. Cognitive Network Management for 5G

As the evolution of network management progresses, the use of machine learning to develop self-aware, self-configuring, self-optimization, self-healing and self-protecting systems will enable cognitive network management. As mentioned earlier, this technology is needed for managing a demanding infrastructure but one that yet has to present scalability and flexibility, such as that needed in 5G. In the following, some of the novelties for network management in 5G are presented, including: autonomicity, NFV, SDN and network slicing.

Some of the novel challenges of Network Management for 5G

Autonomicity - Autonomic computing or network management is not a new area and there have been many projects focused on this in the past, however there are many new challenges that come to 5G because of new technologies such as NFV and SDN, these include:

- New use cases for the network that use these new technologies such as multitenancy and network slicing.
- The additional depth of complexity introduced because while in the past, networks based on hardware components had a static topology, now the network can change dynamically and being able to maintain an accurate view of the state of the network in real time is a challenge
- Knowing how to manage changing topologies and how this impacts on management actions, i.e. effective actions for one topology may not work for another.
- As networks are expected to become more on-demand and therefore highly specialized to a given context, the operator may not necessarily benefit from a large amount of historic data as it may be too specific to the given context.

NFV - In the past most network functions (routers, switches, firewalls, gateways, protocol converters, IMS Cores etc.) were implemented as dedicated physical components, some with specialised hardware etc. Network Function Virtualisation moves all of this to the cloud and the network functions are now virtualised and run purely in software often on standard OSs such as Linux. The advantage of NFV is it allows the dynamic deployment of the network functions that makes the network scalable, it also allows the flexibility to retire functions if they go into an error state and dynamically replace with an equivalent function.

SDN - Software Defined Networking migrates the complexity and routing algorithms from the routers to the network controller. This allows the routers to be implemented as simple VNF (virtualised network functions) that can be controlled via the controller. The controller maintains a model of the network and using this can take a global view of the network, its topology and state. SDN is one of the key technologies underlying intelligent traffic management and network slicing.

Network Slicing - Network Slicing is the support of multitenancy in the network through its division into a number of virtual networks that are logically isolated with their own resources, security and QoS specifications, though in reality many of the physical resources such as radio spectrum, RAN and physical infrastructure are shared. Network slicing is seen as one of the key use cases for 5G and is partially enabled through NFV and SDN technology. Isolation among slices is a fundamental feature to ensure that the traffic of one slice does not negatively impact other slices. In addition, it is essential that this isolation is implemented in a way that leads to an efficient use of resources which is particularly challenging when considering the slicing of a multi-cell RAN, due to the inherently shared nature of the radio channel and the potential influence that any transmitter may have on any receiver. The split of radio resources among RAN slices can be based on different Radio Resource Management (RRM) strategies, such as the spectrum planning, the packet scheduling or the admission control. The selection of one or another option will impact on the degree of isolation and on the capability of customizing the different slices [8].

Knowledge-based Radio Resource Management – Knowledge-based RRM can be realized in 5G networks through the usage of certain contextual parameters related to the radio environment. By being able to retrieve data on the quality of channels, UEs can select in a

proactive manner the best possible channels, depending also on the type of services that will be served. Moreover, prioritization (in terms of available channels) should be given to mission-critical applications, followed by other applications such as mobile broadband, machine-type communications etc. Finally, through the building of knowledge, it will be possible to learn best available channels and/or bands (e.g., from a pool of licensed, lightly-licensed, unlicensed) and assign UEs accordingly.

6.1 Challenges for Autonomic Network Management in 5G

5G networks are built on top of a flexible performance-demanding infrastructure and current autonomic network management technology is unlikely to fully support it. ANM emerged as a unification tentative of recent advancements and trends of many network research areas and not on the possibilities that 5G networks are capable of achieving. Concepts such as network softwarization and network slicing, inhibit the usage of the current approaches for ANM, which is limited to managing static network resources e.g. network topologies. New alternatives have to be based on the 5G vision, and more importantly, built with the proper autonomic layered principles, as suggested in [9]. In the following, three autonomic principles, that are complementary to the existing solutions for ANM, in which the evolution of network management of 5G needs to be based on are:

- Autonomic software-defined networks
- Autonomic diagnosis/anticipation
- Autonomic adaptation

Software-defined networking allows the possibility of the 5G vision by establishing itself as the evolution of communication networks through introducing new concepts such as dynamic topologies, network slicing etc. Current ANM solutions, however, are limited in their applicability to network-level fault management due to the lack of efficient techniques for network monitoring programmability. In **autonomic software-defined networks**, instantiating new virtual resources based on the existing network knowledge further enhance network design. Basically, optimization plus learning will come together to allow online network design with high performance and reliability.

ANM approaches are currently more responsive to network events rather than proactively managing the network based on anticipated future events. This will likely be a major issue in 5G, due to the negative impact of these approaches where there are high-performance requirements. In **autonomic diagnosis/anticipation**, predictive tools are used to align the network history profile with the current network trends. An enhanced knowledge base, made possible through social media analysis, will allow prediction of extraordinary events and dynamically change the network to accommodate those. Even though predictive analysis has a high associated overhead, the cloudification of 5G will allow proper conciliation of the mentioned technologies.

Beyond social media analysis, the concept of Web of Things (WoT) around semantic data or linked data is emerging for numerous IoT services like smart cities. There is certainly value to

interface ANM with such WoT platforms to collect additional information in order to predict events that may impact the network. For instance it may improve some processes about mobility management or RAN dimensioning/scalability in areas of the city facing civil works.

The whole flexibility of 5G relies on an adaptive infrastructure that is only possible with a very advanced network management solution. In **autonomic adaptation,** the idea of pro-activeness is further explored with different tools including control theory, evolutionary computing, and artificial intelligence. These tools from different disciplines can produce an integrated solution towards more autonomy for management systems.

7. 5G Management Architectures

There are three projects in the 5GPPP Phase 1 that deal specifically with Network Management; these are CogNet, SelfNet and Sesame. CogNet focuses on the application of Machine Learning (ML) to the management of the NFVI and the SDN through the dynamic configuring of management policies based on ML models. Depending on use cases, the application of ML may also find its way to the Network Orchestrator and the NFV Manager elements within the MANO stack.

7.1 SelfNet

SelfNet focuses on the management of some of the key technologies for 5G, NFV and SDN but with specific focus on the Self Organising Network paradigm including Self Monitoring, Self Optimisation, Self Protection and Self Healing. SelfNet also sets a number of Health Of Network (HON) metrics to measure the stability and performance of the network that serve as its KPIs. Its architecture is depicted in Fig. 8 and detailed in the following subsections.

Infrastructure Layer: Starting from the bottom of the figure, the Infrastructure Layer (IL) contains the Physical Sublayer and the Virtualization Sublayer. The Physical Sublayer contains all the physical elements of the network, as well as the physical servers available on the Data Centre (DC). On top of the Physical Sublayer is provided the Virtualization Sublayer which provides access to the virtual resources of the DC (compute, storage and network) through the hypervisor. The Virtualization Sublayer represents the NFVI (Network Functions Virtualization Infrastructure) as defined by the ETSI NFV terminology. It is assumed that a mesh of DCs, with different sizes and purposes, will be available. A number of high-capacity and centralized DCs will exist to host services that do not have significant real-time constraints. In addition to these centralized data centres, edge DCs will also exist in the operator's access networks in order to provide the real-time demanding services and network functions. Distributing network functions across several DCs also requires the system architecture to manage the inter-DCs network links, also known as Wide Area Network (WAN). SELFNET architecture considers support for the described distributed DC topologies by taking into account the inter-DC WAN connectivity, either it is composed by legacy network elements (e.g. MPLS-based routers) or by SDN-enabled/controlled network elements.

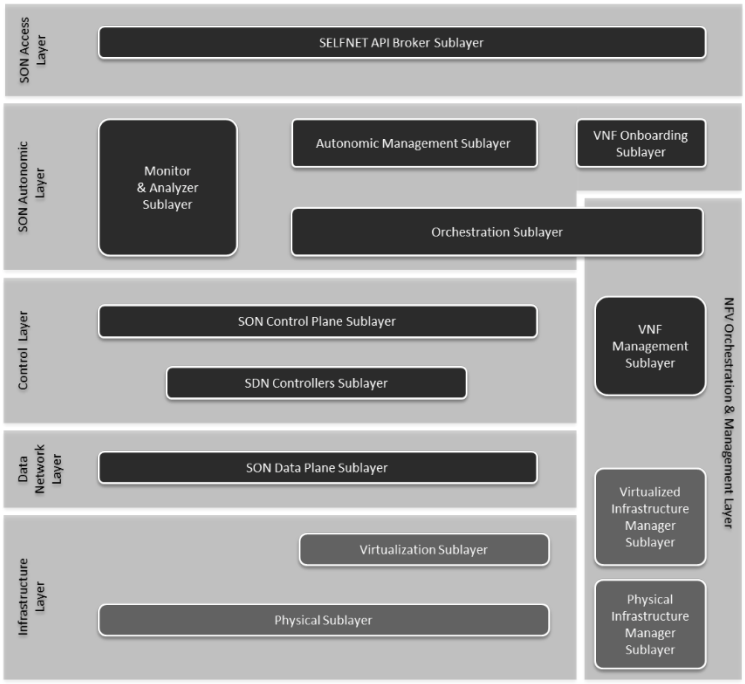

Figure 8: SelfNet Management Stack and components

Data Network Layer: On top of the Infrastructure Layer is located the Data Network Layer (DNL), which represents an explicit architectural evolution towards the SDN paradigm. The SDN paradigm decouples the control plane functions from the data plane functions, transforming the latter into a simple forwarding-based layer. Therefore, in order to be fully aligned with SDN, the DNL (more precisely the SON Data Plane Sublayer), is explicit in the architecture diagram. This sublayer supports SDN-controlled elements, as well as non-SDN-controlled elements.

Control Layer: As shown in Figure 7, on the SELFNET DNL is included the PNE and the VNE types, as well as the data component of the PNF and of the VNF, either they represent a sensor or an actuator. On top of the DNL is the Control Layer (CL), which includes two internal sublayers: SDN Controllers Sublayer and the SON Control Plane Sublayer. The SDN Controllers Sublayer comprises a group of horizontal and vertically distributed SDN Controllers, whereas the SON Control Plane Sublayer represents the network functions control plane, being either actuators or sensors. In terms of the set of network function types that are embedded in the SON Control Plane Sublayer, the PNF and VNF control components, as well as the SDN-Apps have been identified (see Figure 9).

NFV Orchestration and Management Layer: On the right side of the SELFNET architecture diagram is the NFV Orchestration and Management Layer (OML). It corresponds to the ETSI NFV Management and Orchestration (MANO) layer and is responsible for orchestrating and managing the whole set of virtual functions that are embedded on the SON Control Plane Sublayer and on the SON Data Plane Sublayer. As sublayers, it includes (partially) the Orchestration Sublayer, the VNF Management Sublayer and the Virtualized Infrastructure Manager (VIM) Sublayer. The Orchestration sublayer part in the OML corresponds to the ETSI

MANO Network Functions Virtualized Orchestrator (NFVO), and is responsible for orchestrating the virtual resources and network functions. The VNF Management and the VIM sublayers, are responsible for the VNFs and virtual resources management, respectively. These sublayers also correspond to the ETSI MANO VNFM and VIM, respectively.

SON Autonomic Layer: The SON Autonomic Layer is the top-most layer of the SELFNET architecture. This layer provides the mechanisms to provide network intelligence. The layer collects pertinent information about the network behaviour, uses that information to evaluate the network condition, diagnose any pending/existing network issues, and decides what must be done to accomplish the system goals. It then guarantees the organized enforcement of the actions. In essence, the SON Autonomic Layer is split in four main sublayers: Monitor & Analyzer, Autonomic Management, Orchestrator, and VNFs Onboarding. These sublayers and their modules will be described in the upcoming subsections.

SON Access Layer: The topmost layer is the SON Access Layer, which encompasses the interface functions that are exposed by the framework. Despite the fact that internal components may have specific interfaces for the particular scope of their functions, these components contribute to a general SON API, managed by the SELFNET API Broker Sublayer, that exposes all aspects of the autonomic framework to external actors (Business Support Systems – BSS, Operational Support Systems – OSS and Administration GUI). The GUI provides the network administrator the capability to interact with and configure the framework components (e.g. stop, verify or manually enforce any of the actions that SELFNET is governing) and also obtain the complete status of the network.

7.2 CogNet

The diagram in Fig. 9 shows a high level overview of the role of the CogNet technology in the network management architecture for 5G. Some of the key 5G technologies and how CogNet components interact with them, SDN and NFV are shown, as is the MANO stack for managing the NFVI, VNFs and the underlying VMs that support these functions.

A key aspect of CogNet is the use of ML models derived from applying suitable ML algorithms to the network data and metrics collected from the NFVI and the control plane. These are then used to inform the code implementing the policies as shown below. In CogNet, the policies to be implemented in the network are described using the Simplified Use of Policy Abstractions (SUPA) specification. This is a high level abstraction of the desired policy that is not concerned with specific implementation details. However there is sufficient information in the policy to allow the intent to be translated into the specific semantics of the management. In CogNet this is done through code generation and deployed through continuous integration.

In Fig. 10, an ML algorithm has been applied to network data patterns over a period time to create a probability distribution of a VNF entering a failure state within a specific time window. The policy being implemented is to create a hot standby and or switchover depending on the probability level and time horizon of the potential failure.

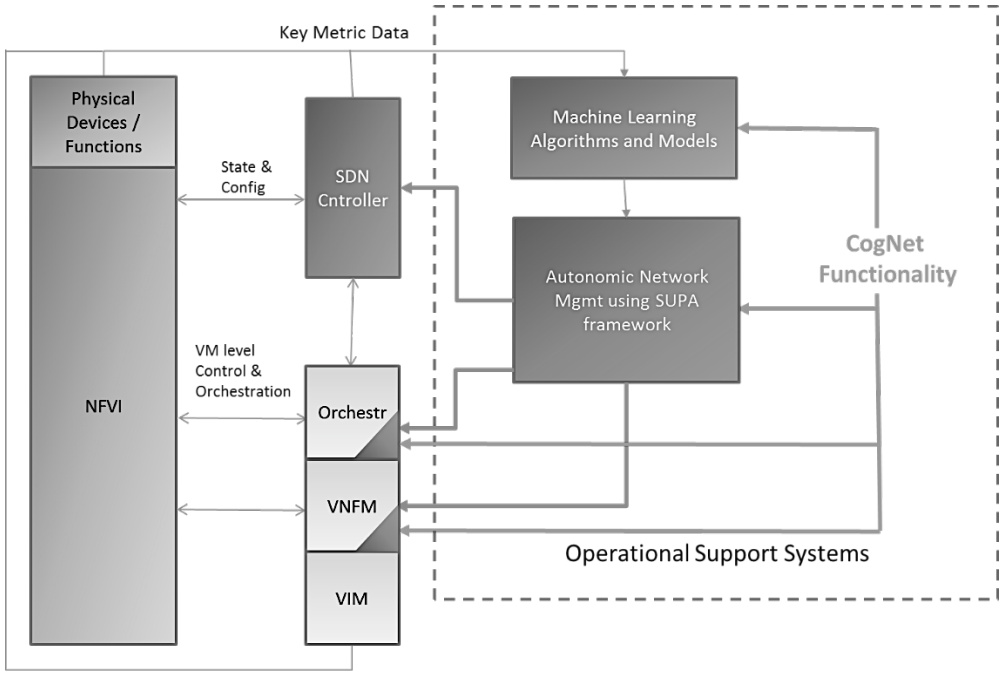

Figure 9: CogNet relationship with the 5G Infrastructure and Management components.

The module implementing this policy is referred to as a policy executor. There may be multiple such executors running simultaneously, each using or sharing ML models or combinations of models. These continuously running policy executors collectively provide a form of autonomic network management. The ML model does not remain static and at any time a new model is being trained which further refines the parameters of the model.

Example Fault Prediction Model (probability node enters failure state over time)

Node No \ Time	5 Sec	10 Sec	30 Sec	2 Min	10 Min	1 Hour	1 Day
1	0.10	0.15	0.15	0.16	0.18	0.21	0.28
2	0.05	0.10	0.10	0.11	0.12	0.14	0.19
3	0.50	0.55	0.57	0.59	0.65	0.79	0.99
4	0.32	0.35	0.36	0.38	0.42	0.50	0.65
5	0.92	0.94	0.97	0.97	0.98	0.99	0.99
6	0.75	0.77	0.79	0.83	0.92	0.97	0.99
7	0.25	0.28	0.29	0.30	0.33	0.40	0.52
8	0.33	0.35	0.36	0.38	0.42	0.50	0.65
9	0.62	0.64	0.66	0.69	0.76	0.91	0.99
10	0.81	0.83	0.85	0.90	0.99	0.99	0.99

Network Fault Tolerance Executor psuedo code example

```
While(true) {
  For each Node N (1..10) {

       If(P(N,"2 Min") > 0.7){
         standby = Start(new(N));
       }
       If(P(N,"10 Sec") > 0.9){
         switchover(standby);
       }
  }
  sleep(1);
}
```

Figure 10: Sample ML model and pseudo code using this to implement fault tolerant policy.

7.3 SESAME

SESAME project focuses on the management of multi-tenant small cells. The SESAME architecture is presented in Fig. 11 [10]. The Cloud Enabled Small Cell (CESC) is a complete

Small Cell (SC) with necessary modifications to the data model to allow Multi-Operator Core Network (MOCN) radio resource sharing. The CESC is composed by a Physical SC unit and a micro server. The physical aggregation of a set of CESCs (CESCs cluster) provides a virtualised execution infrastructure, denoted as Light Data Centre (Light DC), enhancing the virtualization capabilities and process power at the network edge. The functionalities of the CESC are split between SC Physical Network Functions (PNFs) and SC Virtual Network Functions (VNFs). SC VNFs are hosted in the environment provided by the light DC.

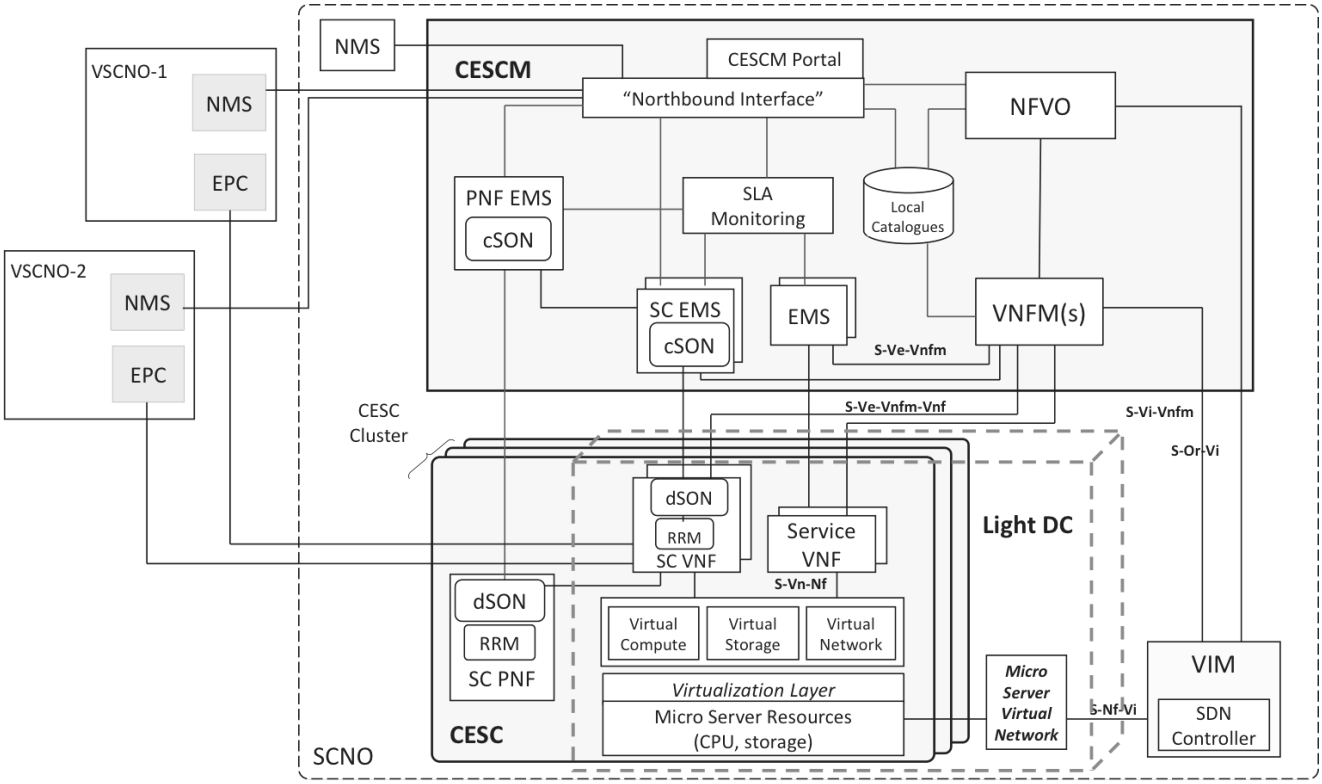

Figure 11: SESAME architecture

The CESC Manager (CESCM) is the central service management component in the architecture that integrates the traditional 3GPP network management elements and the novel functional blocks of the NFV-MANO (Network Function Virtualisation - Management and Orchestration) framework. Configuration, Fault and Performance management of the SC PNFs is performed through the PNF Element Management System (EMS), while the management of the SC VNFs is carried out through the SC EMS. The EMSs provide performance measurements to the Service Level Agreement (SLA) Monitoring module that assesses the conformance with the agreed SLAs. EMSs are connected through the northbound interface with the Network Management Systems (NMS) of the Small Cell Network Operator (SCNO) and the different tenants, denoted in SESAME as Virtual Small Cell Network Operators (VSCNOs), providing each VSCNO with a consolidated view of the portion of the network that they are able to manage. Finally, the CESCM includes a portal that constitutes the main graphical frontend to access the SESAME platform for both SCNO and VSCNOs.

Automated operation of CESCs is made possible by different SON functions that will tune global operational settings of the SC (e.g., transmit power, channel bandwidth, electrical antenna tilt) as well as specific parameters corresponding to Radio Resource Management (RRM) functions (e.g., admission control threshold, handover offsets, packet scheduling weights, etc.). As shown in Figure 10, the PNF EMS and SC EMS include the centralised self-x functions (cSON) and the centralised components of the hybrid SON functions. In turn, the decentralised (dSON) functions - or the decentralised components of the hybrid functions - reside at the CESCs. The dSON functions can be implemented as PNFs or, if proper open control interfaces with the element (e.g. the RRM function) controlled by the SON function are established, they can also be implemented as VNFs running at the Light DC. The mapping of the specific RRM and SON functions in the different components of the architecture depends in general on the selected functional split between the physical and virtualized functions.

8. 5G Management KPIs

Below are listed some of the key aspirations for 5G technology over and above today's 4G technology and supporting systems [11]. Many of the below will rely on enhanced and more efficient management of the network, in addition to improvements in underlying technologies and Radio Access bandwidth.

- 1000 times higher mobile data volume per geographical area.
- 10 to 100 times more connected devices.
- 10 times to 100 times higher typical user data rate.
- 10 times lower energy consumption.
- End-to-End latency of < 1ms.
- Ubiquitous 5G access including in low density areas.
- Deployment time to be reduced by 1000 in comparison with the current 4G system (basically from 90 days to 90 minutes)

To measure whether the management of the network is achieving the desired improvements, below is an overview of some of the key technologies and suggestions on how they may be analysed and measured.

Autonomicity: What is the level of autonomicity that can be achieved - how would you express this ? In terms of transactions or operations that need to have an operator involvement? Can this vary over time so for example if we are employing a ML system, over time this would be trained and presumably would be become more autonomic over time? It should be established how this can be measured? Perhaps transactions that require a human operator input. Maybe this could be classified by transaction type?

Network Resource Utilisation: Even with the use of NFV, we should be able to measure the resource utilisation of the network, so given that we can deploy or shutdown resources at certain threshold utilisation points, within these threshold points, can we calculate resource

utilisation for NFVI? This in turn would allow us measure whether our active network management is actually improving utilisation

Relative Efficiency: If we are using intense computational methods for network management this only makes sense if we can achieve an efficiency and resource savings that is greater than the cost of the computation over and above the efficiency of traditional deterministic approaches to managing networks. So for example if we needed to deploy 10 CPUs performing processing to achieve a 10% efficiency gain in a network consisting of 100 CPUs, this would only be a breakeven point. However, there may be economies of scale that may make this a non-issue.

Traceability: Where we are using either deterministic, intelligent or statistical methods, we need to be able to trace how the software makes a decision. This is required for justification, accountability and potentially for error tracking purposes. A KPI could be the percentage of transactions whose underlying reasoning and outputs can be fully accounted for. Ideally this should be 100

Quality of Service (QoS): QoS has to be measured objectively as different types of applications using the 5G network may have different QoS requirements. There are the traditional measures such as latency, bandwidth, jitter, error rates, no. of dropped or out-of-order packets etc. For an optimized network QoS has to be measured not only at the end user, but continually and everywhere in the network to enable the network management to optimise the traffic control decisions. However, the metrics in themselves do not tell us enough and are disjointed.

A potential approach to using these values may be to take a number of these and take the average measured rate as a percentage of the theoretical maximum or minimum and also indicate the standard deviation of the measurements as an indication of how reliable the particular measure is.

There could also be some measure of cross correlation between characteristics, measure using something like covariance. So for example high latency may be highly correlated with jitter or error rate etc.

To represent different QoS requirements across application types and to enable the network management to assign an IP data packet to a QoS level, IEEE802.1Q standarized a Priority Code Point (PCP) in the Ethernet header to represent 8 classes of service defined in IEEEP802.1p.
These services are

- background (lowest priority)
- best effort
- excellent effort
- critical applications
- video
- voice

- internetwork control
- network control (highest priority)

The design of cognitive network management has to take into account whether this already defined classification is sufficient or if there has to be further service data transmitted to provide the required QoS.

These QoS metrics could be considered to be a potential optimisation function for the Network Management and the quality of the network management could be measured against these, if QoS is the metric to optimise against.

Quality of Experience (Perceived Quality of Service (PQoS)): An important KPI for a cognitive managed network is to provide advantages for the customer. These advantages can be described by the Quality of Experience (QoE) KPI. QoE describes the overall customer perception of the application (service). Of course, this reflection includes the time to set up the service, design of the application, usability and handling of the service besides the quality parameters on the technical basis which are described with the term Quality of Service (QoS). In contrast to QoE, QoS can be measured objectively.

Ease of Deployment of New Applications: A KPI could be around how easy it is to launch new applications in the network or how easy it is to integrate new applications. Is there some KPI from the software world for measuring this as it could be brought over into 5G? From a Network Mgmt perspective the ideal is that new applications and the network which manages them should be as self configuring as possible to achieve a “plug and play” of new applications and components/devices.

Stability: This is similar to the old ‘5 9s’ or 99.999% availability. But this measure was applied to telecoms infrastructure that was very much a siloed closed loop. With NFVI, the telecoms infrastructure is now potentially shared with other types of cloud infrastructure and ensuring and measuring stability and/or availability becomes more challenging.

9. Limitations and Challenges

9.1 Network Neutrality

Neither network neutrality nor network slicing is defined by the EU legislation. Both concepts, however, are regulated by the EU Regulation laying down measures concerning open internet access and amending Directive 2002/22/EC on universal service and users’ rights relating to electronic communications networks and services and Regulation (EU) No 531/2012 on roaming on public mobile communications networks within the Union (Telecom Single Market Regulation 2015/2120 - TSM). The rules preserve the principle of net neutrality that aims at maintaining an open, end to end network, offering scope for providing "specialized services" which are defined in the art. 3(5) of TSM as “services other than internet access services which are optimised for specific content, applications or services, or a combination thereof, where the

optimisation is necessary in order to meet requirements of the content, applications or services for a specific level of quality." Examples of specialised services include Voice-over-LTE (high-quality voice calling on mobile networks) and broadcasting IPTV services with specific quality requirements, as well as high-definition videoconferencing or real-time healthcare services like remote surgery. The support of these specialized services, each one with specific requirements and a level of services guarantees, requires some sort of services prioritization in terms of flexible configuration of resources in the networks, to adapt to changes in end-user/application and traffic demand. However the economic sustainability of the Internet ecosystem, raises questions about the applicability of traffic management practices and the new investments in networks, to meet the increasing demand for bandwidth generated by the explosion of data traffic. The support of new business applications each one with specific requirements and a level of services guarantees, especially in the context of 5G, requires some sort of services prioritization in terms of flexible configuration of resources in the networks, to adapt to changes in end-user/application and traffic demand.

Network slicing is a key enabler for supporting specialized services, allowing on a single infrastructure the creation of multiple virtual networks, each one configured to suit different traffics or application types with a specific level of services guarantees and requirements.

The use of slicing is also mentioned in the BEREC Guidelines on the Implementation by National Regulators of European Net Neutrality Rules (BoR (16) 127) saying that "Network-slicing in 5G networks may be used to deliver specialised services", In addition BEREC states that the requirement of an application can be specified by the provider of the specialised service, or it may also be inherent to the application itself. For example, a video application could use standard definition with a low bitrate or ultra-high definition with high bitrate, and these will obviously have different QoS requirements. A typical example of inherent requirements is low latency for real-time applications (remote surgery in the case of telemedicine or high density platooning in automated mobility).

Therefore, it is assumed that slicing will always deliver services that need optimization. This is not completely true since their delivery is conditioned by the following aspects:

- optimization or a specific level of quality can be provided to a specialized service when a specific level of quality cannot be assured by the standard best effort delivery. Article 3(5) second subparagraph states that "Providers of electronic communications to the public, including providers of Internet Access Services (IAS), may offer or facilitate such services only if the network capacity is sufficient to provide them in addition to any internet access services provided. Such services shall not be usable or offered as a replacement for internet access services, and shall not be to the detriment of the availability or general quality of internet access services for end-users." Therefore, specialised services shall only be offered when the network capacity is sufficient such that the IAS is not degraded (e.g. due to increased latency or jitter or lack of bandwidth) by the addition of specialised services. Both in the short and in the long term, specialised services shall not lead to a deterioration of the general IAS quality for end users. Given

the assumed increased throughput offered by the future 5G networks, it is unlikely that the 5G networks will not offer sufficient capacity to offer specialized services.
However, it is interesting to observe what happens, if the 5G coverage is not full and there is a sudden network handover to older generation of networks (4G) or other types of networks with lower characteristics (satellite, WiFi). This situation will create immediate shortage of capacity and a possible slice with guaranteed QoS would cannibalize the existing best effort delivery available on the 4G network. Such example would be incompatible with the regulation.

- specialised services cannot replace the IAS. In point 110 of its Guidelines, BEREC states that: “Specialised services do not provide connectivity to the internet and they can be offered, for example, through a connection that is logically separated from the traffic of the IAS in order to assure these levels of quality.”

- Specialised services need to be in line with the regulation Provisioning of specialised services will be closely monitored by the National Regulatory Authorities. NRAs could request from the provider relevant information about their specialized services, using powers conferred by Art. 5(2) of TSM. In its responses, the providers should give information about its specialised services, including what the relevant QoS requirements are (e.g. latency, jitter and packet loss), and any contractual requirements. Furthermore, the “specific level of quality” should be specified, and it should be demonstrated that this specific level of quality cannot be assured over the IAS and that the QoS requirements are objectively necessary to ensure one or more key features of the application.

9.2 Performance

Several performance requirements have been set for 5G for which it is foreseen that they will be achieved in the middle deployment stage, at least theoretically. Experts understand, however, the optimism of the 5G community and they are being alerted to the potential failures in deploying the proposed technologies required to achieve a fully functioning 5G network.

Network management is among the novel technologies essential for guaranteed performance architectures, either autonomic or cognitive. This solution is one of those responsible for the overall network performance but its efficiency is not yet clarified.

Distributed network management architectures were suggested for allowing full autonomicity of network configuration, but their performance is diminished by distributed protocols. Also, the pace of the network adaptation will dictate whether or not a novel architecture corresponds with the performance demands. Successful solutions must rely in efficient gossip techniques for monitoring aside with management protocols for network setup. It is, as well, noteworthy to use proper management tunnelling for non-delay tolerant applications, bypassing the network management procedures.

9.3 Defining and calculating relevant KPIs can be challenge

Changing completely the networking paradigm for building the next generation, 5G is challenging but promising performance improvements even with envisioned applications and the exponential increase in the number of network nodes. From a management perspective, obtaining monitoring data from the network can be more difficult, unless the design of new KPIs can be achieved.

The network knowledge base must be obtained from relevant data and efficiently filtered for easily accessing reliable information about the infrastructure status. This new information can be translated to novel KPIs, and as already explored in this document, higher degrees of cognition depend on this data for efficiently developing learning algorithms.

9.4 Cross-Layer network management

5G researchers have promised to deliver a flexible infrastructure that will leverage the network performance requirements for future applications. This flexibility will utilise software and hardware flexibility. However, there are as yet no solutions that take advantage of this powerful possibility for network management that 5G networks are opening.

Network nodes need to be able to scale up and down resources based on agreed targets for performance, maintenance, sustainability and reliability. Without a cross-layer approach, network management will be limited to logic operation only in the connections between nodes, but not in the nodes themselves. For this, distributed network management solutions have to appear for adapting network nodes to the incoming network traffic convoluted with network policies.

9.5 Integration with existing infrastructure

The transition towards world-wide 5G infrastructure possibly will take 5-10 years, and is going to take different deployment times in different parts of the world. On top of that, likely customers will not accept the new technology straight way, even with its numerous benefits. For this, 5G has to accommodate past mobile network generations for smoother transitions.

In search for more flexible infrastructures, 5G is being built on a total different concept which is not integrated with existing infrastructure. Novel integration solutions must emerge, and also 5G architecture must consider the existing infrastructure for service aggregation.

9.6 Security and Trustworthy of AI integration

Building an interdisciplinary approach for network management requires much effort reconciling various concepts, but more importantly, building a secure and reliable solution. A deep knowledge of the areas involved in the interdisciplinary approach is also very appealing, and one must fully understand how they are going to work together and what limitations are encountered.

However, dealing with cognitive solutions requires a higher degree of security and trustworthiness. Learning algorithms can erroneously interpret performance degradation based on biased training data or network configuration. On top of that, new types of network attacks might emerge when these algorithms are damaged by some underlying activity on the network that can negatively impact on the learning models or the built knowledge base.

9.7 Fully Automated Network Management

Questioning the feasibility of manual network operation is very pertinent where the number of the network nodes is increasing exponentially. Human network operators are no longer able to make real time decisions based on reading network statistics or even network logs.

A highly autonomic degree is required in this particular scenario towards, a priori, rather than fully automated. Security, as discussed before, is a major impairment to allowing full automation of network management. On top of that, human resources usually handle resilience of the network. Therefore, it is likely that there will remain a need for human network operators, who will perform machine-learning assisted network management, rather than a fully autonomic type solution.

10. Conclusions

Moving through the generations of mobile networks, they have historically relied on hardware technology advancements but that will not be the same for 5G. Software technology advancements are now also required, while the same has to be applicable to the network. One example of this is network management, which through recent years has built an entire framework that allows full automation when handling network resource usage, namely autonomic network management. However, 5G plans require a more robust paradigm for network management. The number of devices, the demanding services traffic, the performance requirements of the network require a more optimized yet specialized network management solution, capable of dealing with flexibility of resources and maximization of the network efficiency. This will require learning algorithms, which can analyse and quantify the current traffic in the network precisely, allowing for improved efficiency, dynamic scaling, resilience and reliable and secure network slicing. This solution aligned with concepts such as self-awareness, self-configuration, self-healing, self-optimization and self-protection can be defined as cognitive network management. In this paper, we have explored the characteristics of cognitive network management, with its limitation and challenges, as well as defining new network performance metrics specialized on the 5G KPIs. Cognitive network management is likely to pave the way as a key enabler of 5G performance expectations.

APPENDIX

A: Open Source Initiatives

A.1 OpenMANO and Open Source MANO

Open Source MANO [12] is an open source project that aims to provide a practical implementation of the reference architecture for NFV management and orchestration proposed by ETSI NFV ISG. The OpenMANO framework consists of three major components: *openvim*, *openmano*, and *openmano-gui* all available under Apache 2.0 license. The first component is not directly related to the orchestration task and focuses on building a virtual infrastructure manager (VIM) that is optimized for VNF high and predictable performance. Although it is comparable to other VIMs, like OpenStack it includes control over SDN with plugins (floodlight, OpenDaylight) aiming for high performance data plane connectivity. It offers a CLI tool and a northbound API used by the orchestration component *openmano* to allocate resources from the underlying infrastructure, this includes the creation, deletion and management of images, flavours, instances and networks. Openvim provides a lightweight design that does not require additional agents to be installed on the managed compute nodes.

The orchestration component itself can either be controlled by a web-based interface (*openmano-gui*) or by a command line interface (CLI) through its northbound API. OpenMANO's orchestrator is able to manage entire service chains that are called *network scenarios* and correspond to ETSI NFV's *network services* at once. These *network scenarios* consist of several interconnected VNFs and are specified by the service developer with easy YAML/JSON descriptors. It offers a basic life-cycle of VNF or scenarios (define/start/stop/undefine). This goes beyond what simple cloud management solutions, like OpenStack, can handle. The easy to install framework includes both, catalogues for predefined VNFs and entire network services including support to express EPA (Enhanced Platform Awareness) requirements.

OpenMANO does not provide interfaces for the integration of service development tools, like feedback channels for detailed monitoring data to be accessed by service developers. This limits the current system functionalities to orchestration and management tasks only.

More recently, a new project namely Open Source MANO (OSM) [13] was announced. It is focused on delivering an Open Source NFV Management and Orchestration software stack for production NFV networks.

A.2 OpenBaton

OpenBaton [14] aims to provide a NFVO framework that is fully compatible with the ETSI NFV ISG specifications. It uses OpenStack as underling VIM and provides a plugin mechanism to support additional VIM types. The same mechanism is provided to integrate either the default virtual network function manager (VNFM) or a VNFM provided by a third party. These VNFMs can communicate with OpenBaton by using a message queue system or a RESTful JSON interface. OpenBaton uses the ETSI NFV description format to specify VNFs and network services consisting of multiple VNFs. It can manage the end-to-end deployment of these

services across multiple data centre instances (NFV PoPs) and provides basic slicing support for multi-tenant environments. The system is implemented in Java and provides a web-based dashboard and a command line interface (CLI) for user interactions.

The current version does still focus on providing the basic network service provisioning and management functionalities and there is no support for auto-scaling or fault management at the moment. OpenBaton does not offer built-in VNF monitoring functionalities to directly support the service optimization process.

A.3 OpenStack

OpenStack [15] is an open-source cloud computing platform for public and private clouds. It is built out of a series of interrelated projects that deliver a cloud infrastructure solution. It is one of the leading cloud platforms used by several governments and major carriers. OpenStack is managed by the OpenStack Foundation, a non-profit, vendor-neutral, multi-stakeholder effort to help build and promote the OpenStack platform which oversees both development and community-building around the project. While OpenStack in 2010 was made up of two companies, the OpenStack Foundation in 2015 numbers well over 100 members. OpenStack's APIs are a *de facto* standard for IaaS APIs for both private and public cloud and it is the most commonly IaaS used by both enterprises and telecoms. The OpenStack's projects that are most relevant are:

- **Tacker** – OpenStack’s Tacker project [16] aims on developing a general-purpose orchestrator and VNF manager for OpenStack that is compatible to the MANO design of ETSI reference architecture. The goal is to support the end-to-end orchestration and management of network services composed of several VNFs deployed on multiple OpenStack instances. Tacker uses TOSCA's NFV profile schema to describe VNFs and services. As default, it uses the OpenStack Heat component to interact with the underlying VIMs by translating parts of the TOSCA definition to the Heat specific template language. The project provides a management driver framework that can be used to inject initial configurations to VNFs and to update configurations during operation. This framework provides an extendable design so that vendors can include their own management and configuration tools.
- **Murano** – OpenStack’s Murano project [17] is an application catalogue, enabling application developers and cloud administrators to publish various cloud-ready applications in a browsable categorized catalogue. The key goal of the Murano project is to provide UI and API that allows to compose and deploy composite environments on the Application abstraction level and then manage their lifecycle.
- **Mistral** – OpenStack’s Mistral project [18] is a workflow service - any process can be described as a set of tasks and task relations, once this description is upload to Mistral, Mistral takes care of the state management, correct execution order, parallelism,

synchronization, and high availability. Mistral also provides flexible task scheduling so processes can run according to a specified schedule (instead of running it immediately).

- **Congress** – OpenStack's Congress project [19] provides policy as a service across any collection of cloud services in order to offer governance and compliance for dynamic infrastructures. Congress aims to provide an extensible open-source framework for governance and regulatory compliance across any cloud services (e.g. application, network, compute and storage) within a dynamic infrastructure. It is a cloud service whose sole responsibility is policy enforcement.

These projects currently support only OpenStack as underlying infrastructure (and not any VIM). Furthermore, these projects are currently under development and their components are not yet finalized.

A.4 OpenDayLight

A key abstraction of the SDN paradigm is the separation of the network control and forwarding planes. The control logic is implemented on top of a so-called SDN controller. The controller is a logically centralised entity which is responsible for a set of tasks, including the extraction and maintenance of a global view of the network topology and state, as well as the instantiation of forwarding logic appropriate to a given application scenario. This central approach opens the door for efficient network configuration and monitoring opportunities in SDN enabled networks. In practice the controller manages connections to all substrate switches using a southbound protocol such as OpenFlow, and installs, modifies and deletes forwarding entries into the forwarding tables of the connected switches by using protocol specific control messages. While conceptually SDN controllers are centralised, in real world deployments the controller functionality may be distributed across multiple devices to ensure scalability and failure resilience.

OpenDaylight [20] is currently the latest and also largest SDN controller platform. It is backed by the Linux Foundation and developed by an industrial consortium, which includes Cisco, Juniper and IBM, among many others. OpenDaylight includes numerous functional modules, which are interconnected by a common service abstraction layer. It provides an extendable software platform on top of which SDN applications may be developed and deployed thus offering easy to use (northbound) APIs to the functionality provided by the SDN substrate. As a result, OpenDaylight controller may be regarded as a layer between the SDN substrate and the SDN application layer, which implements the logic for concrete network services.

OpenDaylight also provides a flexible northbound interface using Representation State Transfer APIs (REST APIs), and includes support for the OpenStack cloud platform. More specifically, the current OpenDaylight is built upon four "layers", i.e.:

- Technology-specific plug-ins, for managing SDN and non-SDN devices with various network configuration protocols;
- A Service Abstraction Layer, unifying the capabilities of the underlying technology-specific plug-ins;

- A core of basic network services, such as topology management, host tracking etc.;
- A set of northbound APIs (REST-based) for communicating with network management applications.

A.5 ONOS

The Open Network Operating System (ONOS) [21] is the first open source SDN network operating
System targeted specifically at the Service Provider and mission critical networks. ONOS is purpose built to provide the high availability (HA), scale-out, and performance these networks demand. In addition, ONOS has created useful Northbound abstractions and APIs to enable easier application development and Southbound abstractions and interfaces to allow for control of OpenFlow ready and legacy devices. Thus, ONOS will :

- bring carrier grade features (scale, availability, and performance) to the SDN control plane
- enable Web style agility
- help service providers migrate their existing networks to white boxes
- lower service provider CapEx and OpEx

ONOS has been developed in concert with leading service providers (AT&T, NTT Communications), with demanding network vendors (Ciena, Ericsson, Fujitsu, Huawei, Intel, NEC), R&E network operators (Internet2, CNIT, CREATE-NET), collaborators (SRI, Infoblox), and with ONF to validate its architecture through real world use cases.

CORD (Central Office Re-architected as a Datacentre) [22] combines NFV, SDN, and the elasticity of commodity clouds to bring datacentre economics and cloud agility to the Telco Central Office. CORD lets the operator manage their Central Offices using declarative modelling languages for agile, real-time configuration of new customer services. Major service providers like AT&T, SK Telecom, Verizon, China Unicom and NTT Communications are already supporting CORD. ONOS supports an implementation of CORD called M-CORD which integrates the open CORD framework into service providers' mobile network architecture to enable an agile service-driven environment that can dynamically respond to real-time subscriber demands.

REFERENCES

[1] https://www.ericsson.com/networks/topics/network-slicing
[2] NGMN Alliance, NGMN 5G White paper, version 1, Feb 2015
[3] https://www.fiware.org/
[4] VMWARE Microsegmentation Solution Overview
[5] www.5gensure.eu/
[6] ETSI, Network Functions Virtualisation (NFV); Management and Orchestration
[7] J. Pérez-Romero, O. Sallent, R. Ferrús, R. Agustí, "Knowledge-based 5G Radio Access Network Planning and Optimization", The Thirteenth International Symposium on Wireless Communication Systems (ISWCS-2016), Poznan, Poland, September, 2016.
[8] O. Sallent, J. Pérez-Romero, R. Ferrús, R. Agustí "On Radio Access Network Slicing from a Radio Resource Management Perspective", IEEE Wireless Communications, accepted, November, 2016.
[9] Movahedi, Ayari, Langar, Pujolle "A Survey of Autonomic Network Architectures and Evaluation Criteria", in *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, pp. 464-490, Second Quarter 2012.
[10] I. Giannoulakis (editor) "SESAME Final Architecture and PoC Assessment KPIs", Deliverable D2.5 of SESAME, December, 2016.
[11] 5G Vision: the next generation of communication networks and services - pages 8,9 - https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf
[12] http://www.tid.es/long-term-innovation/network-innovation/telefonica-nfv-reference-lab/openmano
[13] http://www.etsi.org/technologies-clusters/technologies/nfv/open-source-mano
[14] https://www.fokus.fraunhofer.de/en/fokus/news/openbaton_2015_10
[15] https://www.**openstack**.org
[16] https://wiki.openstack.org/wiki/Tacker
[17] https://wiki.openstack.org/wiki/Murano
[18] https://wiki.openstack.org/wiki/Mistral
[19] https://wiki.openstack.org/wiki/Congress
[20] https://www.opendaylight.org/
[21] http://onosproject.org/
[22] http://opencord.org/