Cognitive Network Management for 5G

The path towards the development and deployment of cognitive networking

Whitepaper –draft for consultation
By 5GPPP Working Group on Network Management and QoS

Version:	1.0
Date:	7-October-2016
Document Type:	Draft
Confidentiality Class:	P - Public

Project:	5GPPP Phase 1
Editors	Robert Mullins/ Michael Barros
Contributors:	5GPPP Phase 1 Projects
Approved by / Date:	

Introduction

5G Network Management is a non trivial endeavour which faces a host of new challenges beyond 3G and 4G. The number of nodes, the homogeneity of the access technologies, the conflicting management objectives, 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 network can be expanded to also be considered 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 historical data, which means in 5G, that the network will fully understand system variables and optimally adjust their values for achieving a superior network configuration. In the end, **cognitive network management** is introduced as the next generation of network management and key driver of 5G success.

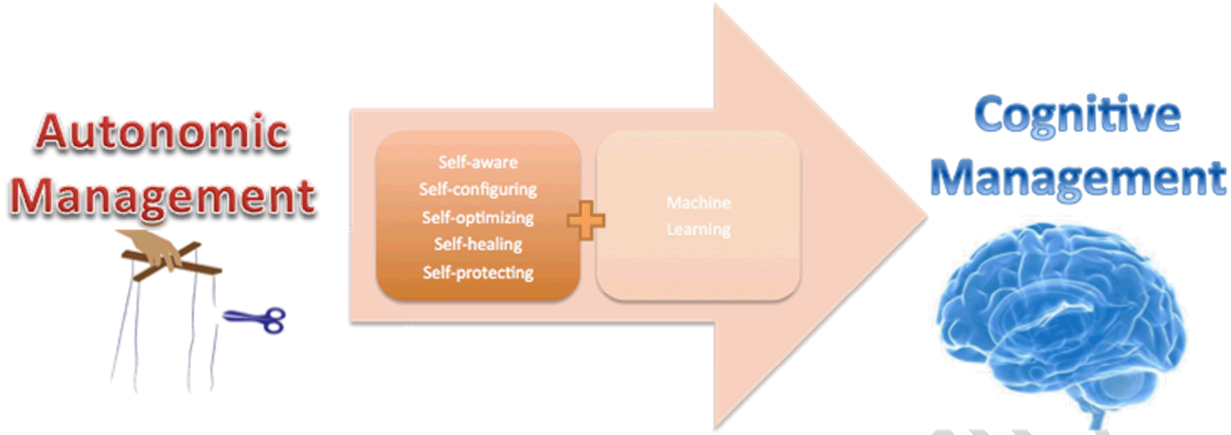


Figure 1. Evolution from autonomic network management to cognitive network management.

On the other side of the management there is **orchestration** of the network, which regulates network resources based on its management decisions. Optimisations and tradeoffs 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 accurate detection of 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.

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.

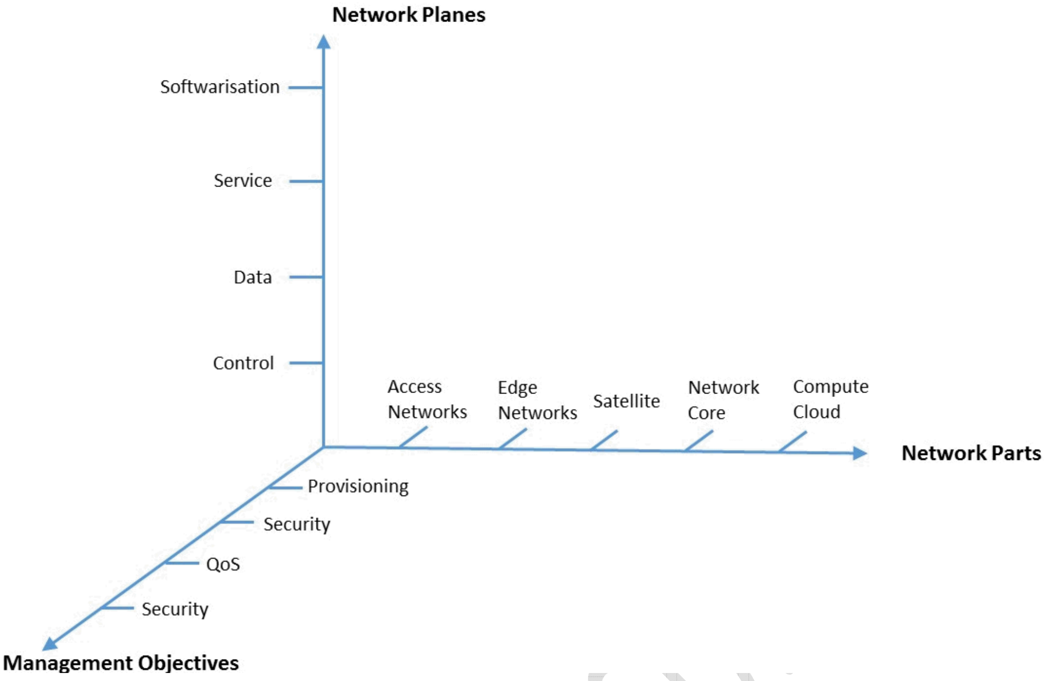


Figure 2. Plane across network planes, network parts and management objectives.

The above non-exhaustive lists of management objectives, network planes and network parts can be modelled as a three-dimensionplane as in Fig. 2 Above. 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 much 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.

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.

New Requirements for Network Management based on 5G

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 concept of slicing the network has been recently introduced for the upcoming 5G mobile networks and it is considered to be an integral part of 5G [Ericsson, 2014, 2015 (January), and 2015 (June)]. It is one of the main enablers and challenges 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 [[8] NGMN Alliance, NGMN 5G White paper, version 1, Feb 2015]. 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.

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.

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 which 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 which 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 for the user using the application. In the business world, customers may pay for a Service Level Agreement (SLA) which contractually defines the QoS which 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.

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…) which 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.

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 these 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.

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.

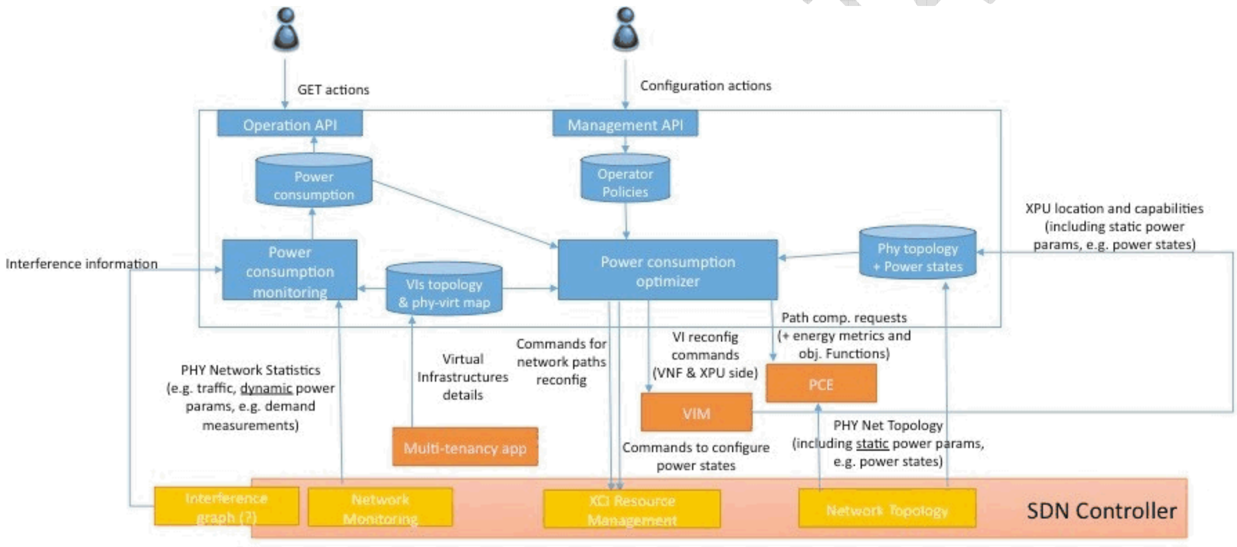

GET actions
Configuration actions
Operation API
Management API
Power consumption
Operator Policies
Interference information
Power consumption monitoring
VIs topology & phy-virt map
Power consumption optimizer
Phy topology + Power states
XPU location and capabilities (including static power params, e.g. power states)
PHY Network Statistics (e.g. traffic, dynamic power params, e.g. demand measurements)
Virtual Infrastructures details
Commands for network paths reconfig
VI reconfig commands (VNF & XPU side)
Path comp. requests (+ energy metrics and obj. Functions)
PCE
VIM
Multi-tenancy app
Commands to configure power states
PHY Net Topology (including static power params, e.g. power states)
Interference graph (?)
Network Monitoring
XCI Resource Management
Network Topology
SDN Controller

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

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

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, such as with SoftAir. This software-based approach will be localized in the control plane of the 5G network. The infrastructure resources will be adjusted according to

a) energy policies,
b) QoS requirements of the data being transferred, and
c) network resource conditions;

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

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 [VMware, 2014], 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 gains 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 center, i.e., monitoring also the traffic inside the data centre. Micro-segmentation is generally an enabler for the Software Defined Data Centre.

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.

Current standards

We will briefly outline the ETSI MANO reference model as detailed in [1], 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.

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 maintain 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.

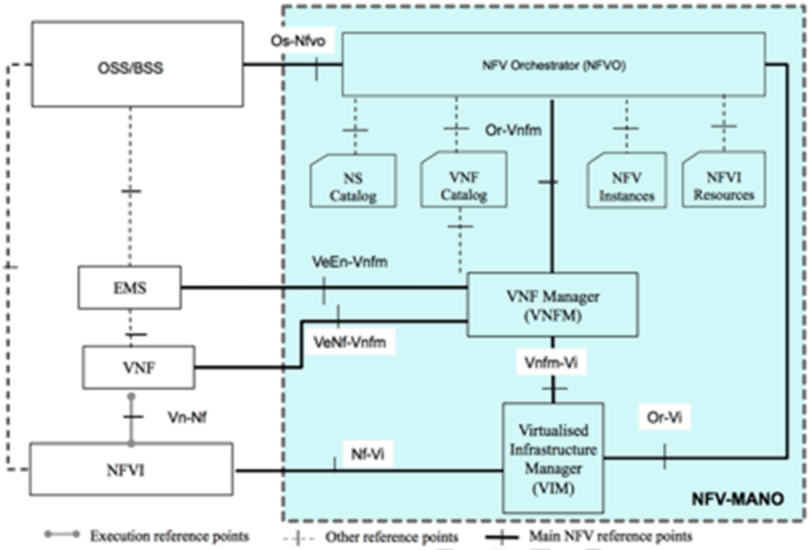


Figure 4: ETSI functional MANO architecture

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.

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.

A building block-type architecture was introduced by IBM for guiding developments of autonomic network solutions. The Monitor-Analyze-Plan-Execute over a shared Knowledge (MAPE-K) is a control theory-based feedback model for self-adaptive systems. Fig. 2. 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 analyzed further. Then the planning and execution pass the new actions to the actuators. This feedback loop is fully tied the knowledge base that is cross-linked to all other building blocks, serving as local network information database.

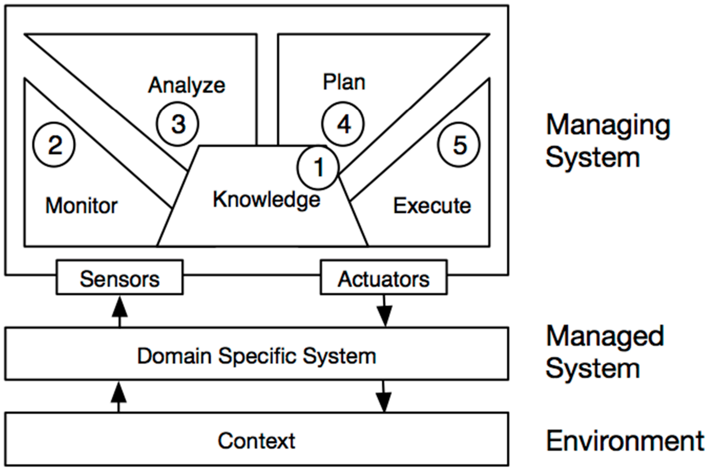
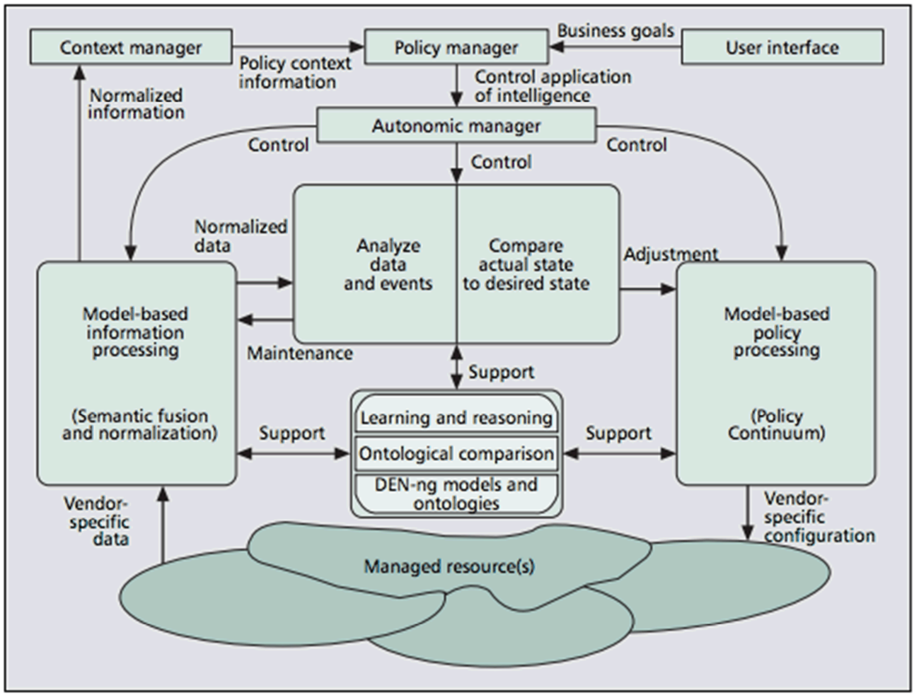


Figure 5. MAPE-K architecture.

The MAPE-K model can be expanded, as depicted in Fig. 6, for deeper understanding. The addition of ontologies and DEN-ng modeling 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.



■ **Figure 2.** *Conceptual representation of an autonomic network management system.*

Figure 6. Detailed ANM system based on MAPE-K.

In the following, more detail about each building block is provided for a deeper understanding of autonomic network management and the MAPE-K.

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 block 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.

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 relies on monitoring the right points on the network.

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.

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.

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 which 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, ie effective actions for one topology may not work for another.

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 which 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) which 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 which are logically separated 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.

Why does it fail 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 is 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 [ref]. 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**, network design is further enhanced by instantiating new virtual resources based on the existing network knowledge.

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 to be a major issue in 5G, due to the negative impact of these approach where there are high-performance requirements. In **autonomic diagnosis/anticipation**, predictive tools are used to align the network history profile with the current network trendings. 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.

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.

Proposed Framework and KPIs for 5G Management

There are two projects in the 5GPPP Phase 1 that deal with Network Management, these are CogNet and SelfNet. 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.

SelfNet again 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 which serve as its KPIs.

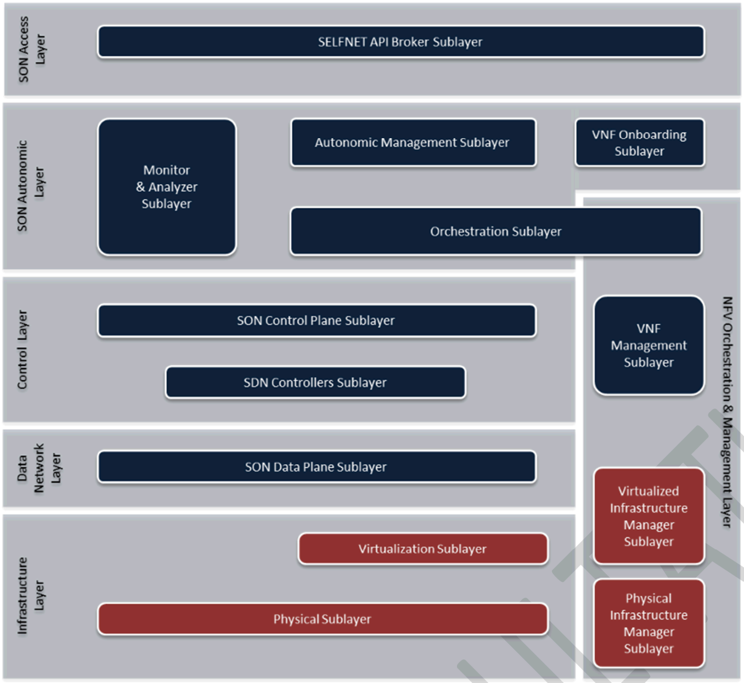


Figure 7. SelfNet Management Stack and components

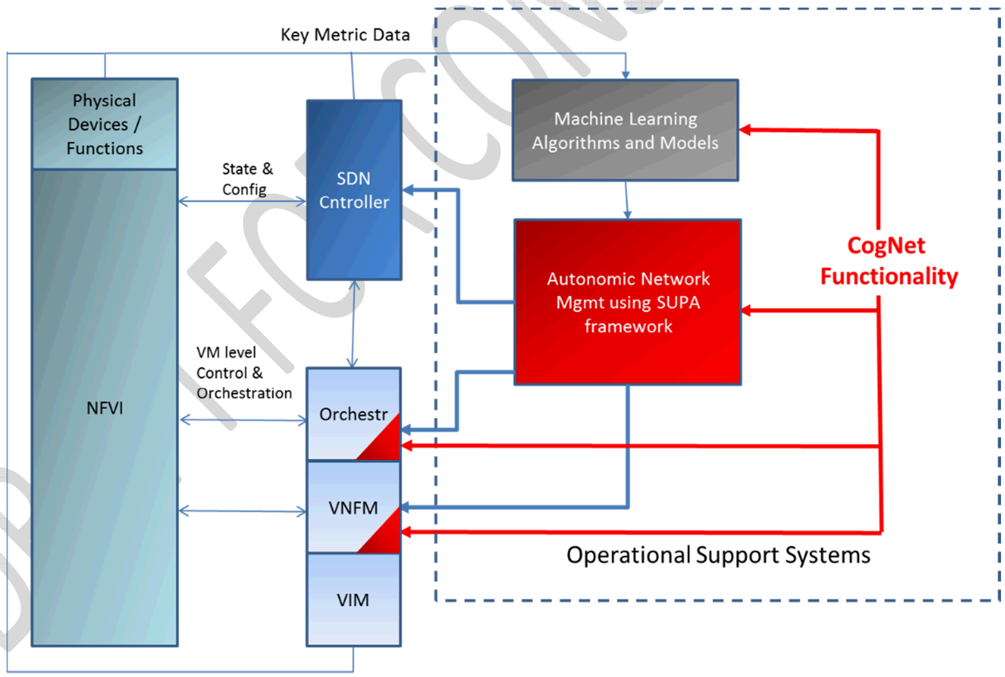

Key Metric Data
Physical Devices / Functions
NFVI
State & Config
SDN Cntroller
VM level Control & Orchestration
Orchestr
VNFM
VIM
Machine Learning Algorithms and Models
Autonomic Network Mgmt using SUPA framework
CogNet Functionality
Operational Support Systems

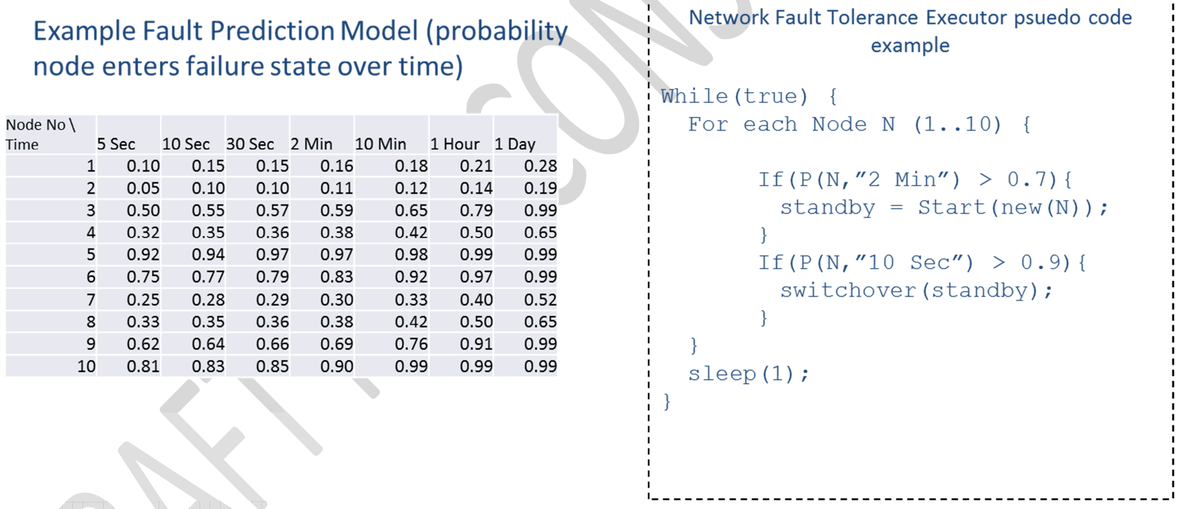
Figure 8. CogNet relationship with the 5G Infrastructure and Management components

The above diagram 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.

5G Management Strategies

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 the below example, 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.



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 9. Sample ML model and pseudo code using this to implement fault tolerant policy

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.

5G Management KPIs

How can we really 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 become more autonomic over time? It should be established how this can be measured? Perhaps % transaction 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 to 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 which 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 break even point. However are there economies of scale in this?

Traceability

Where we are using either deterministic, intelligent or statistical methods, we need to be able to trace how a decision is made by the software. 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%

Achieved 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 etc. However these do not tell us enough and are disjointed:

A potential measure 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.

Talking this 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.

These QoS KPIs 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.

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.

Reliability and availability
This is the old ‘5 9s’ or 99.999% availability. But this 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.

Limitations and Challenges

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. Many experts understand, however, the positivity in the 5G community and they are alerting the community to the potential failures deploying all the proposed technologies to achieve a fully functioning 5G network.

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

Distributed network management architectures were suggested for allowing fully 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.

Challenges to collect network knowledge

Changing completely a networking paradigm for building the next generation, 5G, is challenging but promising performance improvements even with envisioned applications with 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.

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 these data for efficiently developing learning algorithms.

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 be, in this time, mixed with software and hardware flexibility. However, there are no solutions that take advantage of this powerful possibility for network management that 5G is opening.

Network nodes need to be able to scale up and down resources based on network agreements 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 has to appear for adapting network nodes to the incoming network traffic convoluted with network policies.

Integration with existing infrastructure

The transition towards 5G infrastructure world-wide 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 the 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 hardly integrated with existing infrastructure. Novel integration solution must emerge, and also 5g architecture must consider the existing infrastructure for service aggregation.

Security and Trustworthy of AI integration

Building an interdisciplinary approach for network management requires many efforts involving conciliation of 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.

Fully Automated Network Management

Questioning full network automation is very relevant where the number of the network nodes is increasing exponentially. Network operators are not able to make real time decisions based on reading network statistics anymore, or even network logs.

A high autonomic degree is required in this particular scenario towards, a priori, not fully automated. Security, as discussed before, is one major impairment for allowing full automation of network management. On top of that, survivability of the network is usually handled by human resources. Therefore, it is likely that there will remain a need for network operators, which will perform machine-learning assisted network management.

Conclusions

Moving through the generations of mobile networks, they have historically relied on hardware technology advancements but that is not the same for 5G. Software technology advancements are now also required, while the same has to be applicable to the network. One example 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 resource and maximization of network efficiency. This will require learning algorithms, which can quantify the current traffic in the network precisely, allowing efficient 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 defined new network performance metrics specialized on the 5G KPIs. Cognitive network management is likely to pave the way for 5G success as a key enabler of 5G performance expectations.

APPENDIX

PPP projects and open source initiatives (3 pages)

Open Source MANO

Open Source MANO [2] 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 dataplane 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) [3] was announced. It is focused on delivering an Open Source NFV Management and Orchestration software stack for production NFV networks [4].

OpenBaton

OpenBaton [5] 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 center 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.

OpenStack

OpenStack [6] 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 [7] 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 [8] is an application catalog, enabling application developers and cloud administrators to publish various cloud-ready applications in a browsable categorized catalog. The key goal of the Murano project is to provide UI and API which allows to compose and deploy composite environments on the Application abstraction level and then manage their lifecycle.
- **Mistral** – OpenStack's Mistral project [9] 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 [10] 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.

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 [11] 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.