

CogNet: A Network Management Architecture Featuring Cognitive Capabilities

Lei Xu¹, Haytham Assem¹, Imen Grida Ben Yahia², Teodora Sandra Buda¹, Angel Martin³, Domenico Gallico⁴, Matteo Biancani⁴, Antonio Pastor⁵, Pedro A. Aranda⁵, Mikhail Smirnov⁶, Danny Raz⁷, Olga Uryupina⁸, Alberto Mozo⁹, Bruno Ordozgoiti⁹, Marius-Iulian Corici⁶, Pat O’Sullivan¹, Robert Mullins¹⁰
¹Cognitive Computing Group, Innovation Exchange, IBM, Ireland, ²Orange S.A., France, ³Vicomtech, Spain, ⁴INTERROUTE S.P.A., Italy, ⁵Telefonica I+D, Spain, ⁶Fraunhofer-Gesellschaft, Germany, ⁷Bell labs, Nokia, Israel, ⁸University of Trento, Italy, ⁹Universidad Politécnic de Madrid, Spain, ¹⁰TSSG, Waterford Institute of Technology, Ireland

Abstract—It is expected that the fifth generation mobile networks (5G) will support both human-to-human and machine-to-machine communications, connecting up to trillions of devices and reaching formidable levels of complexity and traffic volume. This brings a new set of challenges for managing the network due to the diversity and the sheer size of the network. It will be necessary for the network to largely manage itself and deal with organisation, configuration, security, and optimisation issues. This paper proposes an architecture of an autonomic self-managing network based on Network Function Virtualization, which is capable of achieving or balancing objectives such as high QoS, low energy usage and operational efficiency. The main novelty of the architecture is the Cognitive Smart Engine introduced to enable Machine Learning, particularly (near) real-time learning, in order to dynamically adapt resources to the immediate requirements of the virtual network functions, while minimizing performance degradations to fulfill SLA requirements. This architecture is built within the CogNet European Horizon 2020 project, which refers to Cognitive Networks.

I. INTRODUCTION

Propagations of new types of devices such as smart phones and smart tablets, along with the extensive improvement on mobile communication networks, have produced an eruption of new applications that consume resources from mobile networks and led to an explosive increase of network traffic. Meanwhile, it is expected that the machine-to-machine communication will be increased tremendously in the future to complement the dominating human-centric communication of today [1] [2]. This will lead to a huge diversity of communication characteristic. Both trends will raise new requirements on network scalability, data rates and more stringent latency and reliability [1].

Present wireless-based technologies, such as high-speed packet access (HSPA), long-term evolution (LTE) and Wi-Fi are evolving continuously. The huge demand increase on network capacity, and strict and diverse requirements raised from new communication patterns may not be adequately addressed along with the evolution of ongoing existing technologies. Therefore, the investigation on the new wireless technologies is necessary in order to complement current technologies.

Virtualisation will play an important role in the new generation of networks as the 5G network will need the capability to provision itself dynamically to meet changing demands on resources. The Network Function Virtualisation (NFV) is a key

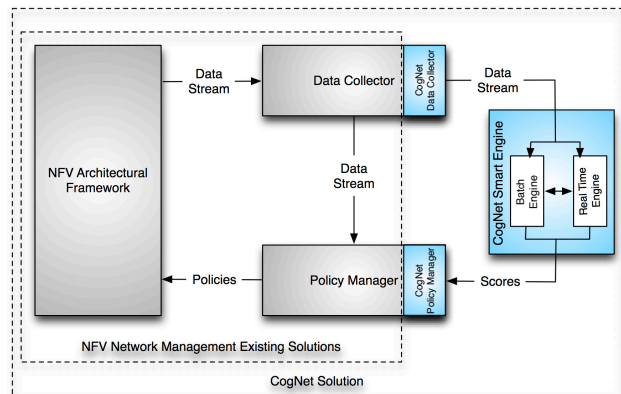


Fig. 1. Architecture Overview.

enabler technology of virtualisation for 5G. Except for the network virtualisation, 5G network will also incorporate a number of technologies, such as network densification and infrastructure sharing, to address the challenges and requirements faced by today’s wireless networks. It is not hard to foretell that the complexity of network management can become one of the biggest challenges in 5G due to the conglomeration of technologies. To cope with similar challenges in 3G and 4G networks, self administering and self managing networks have been extensively researched [3]. The CogNet project applies autonomic network management based on Machine Learning as a key technology for 5G networks to reach the vision of automated management of telecoms network infrastructures.

The project will develop solutions to provide a higher and more intelligent level of network management to ensure quality of service (QoS), improve operational efficiencies and reduce operational expenditure of 5G networks. To achieve this goal and tackle the challenges in the area [4], this paper presents a high level architecture of the CogNet project with an overview presented in Figure 1. It brings a cognitive solution to NFV management. Compared with several related architectural frameworks that handle 5G network management, such as [5] [6], the proposed architecture is enhanced by both batch and real-time Machine Learning solutions to enable an elastic Big Data ecosystem that can scale horizontally or vertically to handle various 5G scenarios.

In the CogNet architecture proposed to extend the NFV

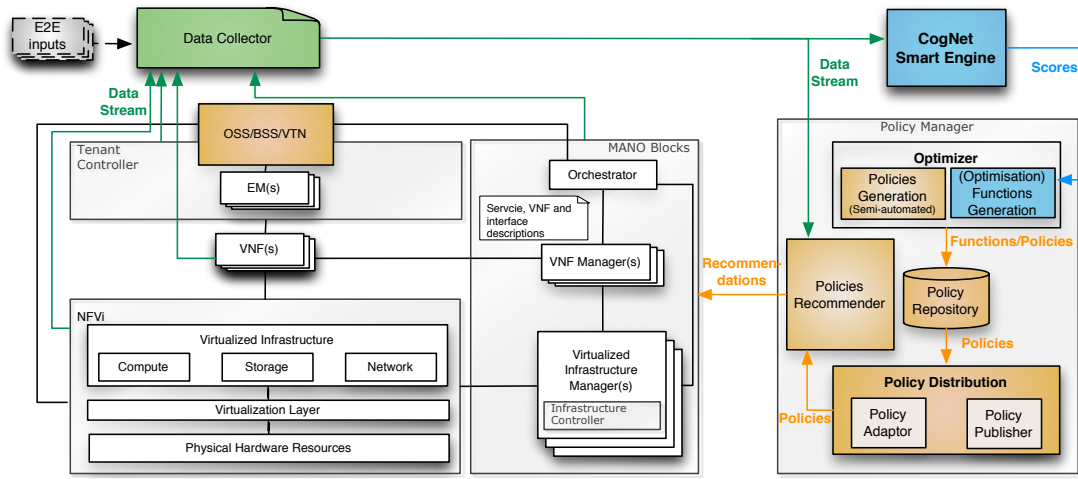


Fig. 2. Architecture of the CogNet Project.

reference architectural framework of European Telecommunications Standards Institute (ETSI) [7], hardware resources are orchestrated and managed in a layered architecture. The state and consumption records on the hardware resources are gathered in real-time from multiple functional blocks constituting the layered architecture. The collected records will be processed by the CogNet Smart Engine (CSE) in (near) real-time or periodically. Real-time processing is one of the core contributions of this work that we believe will be crucial to 5G network management since it aims to provide immediate response to changes.

Based on the output of the CSE, the Policy Manager generates control policies. These policies are helpful to produce an intelligent on demand driven network topology that provides high QoS without using excessive resources. They will be deployed and then recommended to the hardware resources and their management components by invoking related APIs.

II. REQUIREMENTS ANALYSIS

This architecture work aims to address management requirements on cost, efficiency and data analysis. They are discussed in the following sub-sections.

Req.1 QoS & Cost Management

The QoS exposed to end-users is an important element in customer satisfaction and retention. These are crucial for a telco operator to stay competitive. Thus, networks managers of 5G networks will always need efficient tools to ensure that QoS is tailored to the services exposed to their Customers. Meanwhile, efficiency is another metric that will be considered carefully when designing a network and its management system. It is defined as the “relationship between the result achieved and the resource used” in ISO 9001 Standards. The CogNet solution should take into account both issues above and address the requirement:

- The CogNet solution needs to keep QoS tailored to user demand with adequate resource whilst generate the least cost on network equipments (CAPEX) and operations

including network resource allocation, service provision and monitoring, performance degradation and energy efficiency (OPEX).

Req.2 Big Data Analysis & Management

The CogNet project will conduct research in Self Organising Network and will apply the underlying concepts to NFV to produce an intelligent on demand driven network. This goal will be achieved by applying Machine Learning algorithms and their applications to network management. These algorithms and their applications will deal with huge amount of data from a variety of data sources based on current experience of 4G networks. The challenge on big data management will be compounded in 5G networks, and thus the requirement raised by big data analysis is crucial to consider when designing the CogNet architecture that is:

- Big data is identified by volume, velocity, variety and veracity [8]. The analysis part of the CogNet should take account into these features/challenges and yield insights from data efficiently and effectively.

III. COGNITIVE ARCHITECTURE FOR NETWORK MANAGEMENT

The high level architecture of the CogNet is depicted graphically in Figure 2. It includes the NFV architectural framework that will connect to legacy networks, and will constantly forward its state and usage records to the CSE. The CSE will collect data from both the NFV framework and systems that consume resources offered by the NFV framework. It will analyse gathered data for various purposes, such as dynamic resource allocation, security threats and performance degradation detection, demand prediction and then adjust resource offering accordingly. The output of the CSE is some key values that will be applied by the Policy Manager for policy generation. The Policy Manager not only translates rules from the CSE into policies, but also recommends the policies to the NFV architectural framework. The components constituting the CogNet architecture are specified in more detail in the following sub-sections.

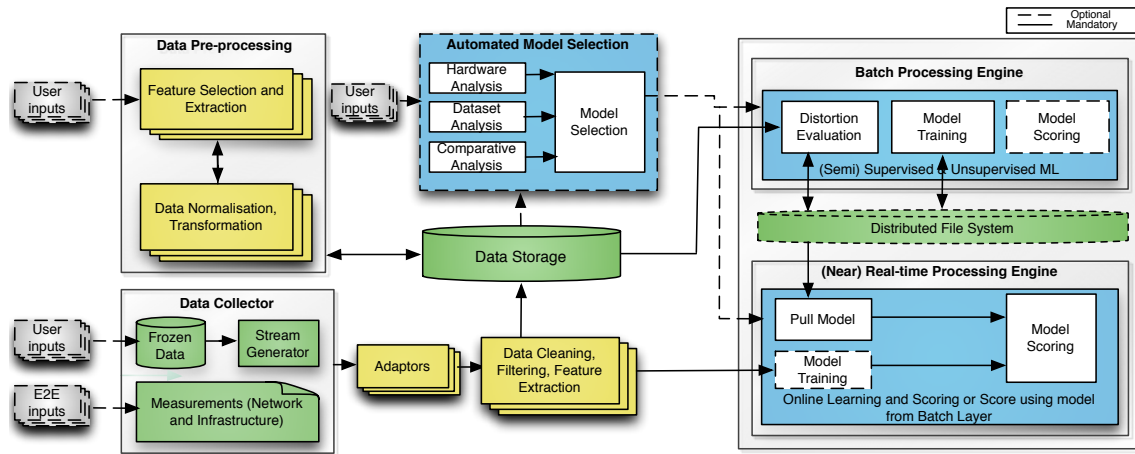


Fig. 3. CogNet Smart Engine.

A. CogNet Smart Engine

The CSE, depicted in Figure 3, is responsible for receiving the state and resource consumption records, pre-processing the records, selecting suitable algorithms, and then applying selected models to further process the received data. The input of this component will be a data stream on the relevant events whilst its output will be scored on the states of given components in the architecture. The CSE collects data from both resource provider-side and consumer-side. This is intended to increase the openness and transparency of services delivered by 5G networks, and subsequently provide better user experiences. The CSE consists of following sub-components:

- *Data Collection & Adaptors* – it collects data from multiple resources, and maps collected data into those that can be processed directly by the following components.
- *Data Cleaning & Filtering* – it cleans and refines received data, and then stores it into the Data Storage or forwards it to the (Near) Real-time Processing Engine.
- *Data Storage* – it stores historical data, and makes them available for multiple components constituting the CSE.
- *Data Pre-processing* – it can work in either automatic or manual mode to pre-process collected data stored in the Data Storage and make them ready for the Batch Processing Engine. Feature extraction will be implemented with Deep Neural Networks, which will allow to generate highly informative features automatically. This is essential to allow for the overall flexibility of the CogNet models, to keep them adjustable to constantly changing network environment. Feature selection, in turn, will control the complexity of the model, to reduce the noise and improve the processing time in the big data context.
- *Algorithm Selection* – similar to the Data Pre-processing, this component is able to identify the model(s) that will be deployed on both processing engines automatically or based on customer requirements. In the automatic mode, it will take account into the features of given datasets,

the performance of candidate algorithms, and available resources of the processing engine.

- *Batch Processing Engine* – it retrieves consumption and state data from the Data Storage, and applies these data to train a model or generate scores. In the former case, the Batch Processing Engine will evaluate the distortion of current model. If the model has become “stale” or no model is available, it will generate a new model from scratch to facilitate the work of the (Near) Real-time Processing Engine. In the latter case, this engine works independently to analyse collected records in a more accurate but higher latency manner. Note that the scoring in both the Batch Processing Engine and (Near) Real-time Processing Engine is not to simply apply one Machine Learning model but may involve a sequence of models associated with post-processing. For example, to detect network anomaly, we may need to score a number of records and then make a conclusion based on a linear combination of generated scores.
- *Distributed File System* – it stores models generated by the Batch Processing Engine that will be deployed on the (Near) Real-time Processing Engine. Note that this component is optional since the Batch Process Engine may forward generated models directly, such as through message queues/RESTful Web Services or the two processing engines may not share data between each other without writing it to an external storage system if they are implemented and deployed in some cluster computing systems, such as Apache Spark.
- *(Near) Real-time Processing Engine* – it consumes the data from the sources directly, and scores the received data within a short period of time. This can be achieved by applying the model generated by the Batch Processing Engine, or light-weight on-line learning approaches directly, such as some on-line clustering algorithms. Depending on the latency requirement of a network, this component can generate scores either in real-time if it is implemented and deployed in a distributed real-time

computation system, such as Apache Storm, or near real-time if it is powered by a mini-batch system, such as Spark Streaming.

B. Policy Manager

This component is mainly responsible for translating the output from the CSE to the policies that can be directly understood by the related components in the MANO Blocks, Tenant Controller and OSS/BSS/VTN. It consists of the following sub-components:

- *CogNet Optimizer* – it transforms the outputs of CSE that are typically some scores into optimisation functions, and then into certain policies in a semi-automated way. Note that human interaction might be needed in the translation of policies.
- *Policy Repository* – it stores policies generated by the CogNet Optimizer and forwards these policies to the Policy Distribution & Execution.
- *Policy Distribution & Execution* – The Policy Distribution translates the policies from the Policy Repository to those that can be directly understood by the Policy Execution. The Policy Execution invokes APIs offered by the components that are hosted in the MANO Blocks, Tenant Controller and OSS/BSS/VTN based on received policies and network state information to enforce the policies.

C. NFV Architectural Framework

The NFV architectural framework decouples the software of a given network function from the hardware or the infrastructure (compute, storage, and networking) it relies on. In the CogNet solution, the NFV framework is composed of the following key elements: NFVI, VNFs, Tenant Controller and the Management and Orchestration blocks, which is aligned with the specification of ETSI NFV working group.

- *NFV Infrastructure (NFVI)* – it includes the physical and the virtual resources where the VNFs instances will be deployed. The physical resources provide the computing, storage and networking capabilities to VNFs. The virtual resources abstract those physical resources through the virtualisation techniques (i.e. hypervisor).
- *Virtual Network functions (VNFs)* – they are the cornerstone of the NFV architecture. A VNF represents a given network function (HSS, MME, EPC, etc.) that is hosted/accommodated over virtual machines (VMs), containers, etc. Thus a Networking Service (NS), (e.g. VPN, Virtual VoIP, etc.) relies on a multiple VNFs. In its turn, a VNF could be atomic or composite. A composite VNF is a set of VNF components (VNFC) and each VNFC is deployed on a VM.
- *Tenant controller* – a tenant is a concept that could reference a person, a client/customer, an organization, etc. It is using the NFVI to fulfil or manage the NS. The Tenant Controller refers to position the SDN controller within the tenant domain (i.e. the tenant controller could be also a VNF that indicates to other VNFs to take actions on the traffic). The Tenant Controller has a direct interface

to the VNFs to provide programmability of these VNFs. It ensures then a dynamic composition of VNFs according to tenant information, profiles, constraints, etc.

- *Management & Orchestration Block (MANO Block)* – it is the block responsible to ensure continuous provisioning including configuration, lifecycle management and orchestration for NS, VNFs, VNFC and their underlying virtual and physical infrastructure. This component is responsible of all the management operations that are specific to these virtualised components. Three main functions are defined: NFVO, VNFM, and VIM.
 - *Network Function Virtualization Orchestration (NFVO)* – it is responsible to manage the NS lifecycle. The NS lifecycle management includes the registration of a NS within the catalogue and verification that all templates (descriptors) are available and complete. It also includes the instantiation of the NS while using the available artefacts. NFVO handles the scaling up and down of NS, by reducing or augmenting its capacity in response to networks changes or business objectives. In this regards, NFVO update NS templates and configurations including the changes of VNF connectivity or changing the VNFCs of a given VNF. Obviously, NFVO handles the following operations: create, delete, query, and update for NS.
 - *Virtual Network Function Manager (VNFM)* – it manages the lifecycle of VNFs. VNFM is responsible to instantiate VNFs, ensure scaling up and down by reducing or augmenting capacity, update VNFs changes with respect to configurations and terminate VNFs by releasing its NFVI resources to the resource pool.
 - *Virtual Infrastructure Manager (VIM)* – it manages the compute, storage and compute resources.

D. External Interfaces

NFV architecture is exposing multiple reference points. We cite here the most relevant ones to the Cognet architecture:

- *OSS-NFVO reference point* – Os-Ma-nfvo, a reference point between OSS/BSS and NFVO. Through this reference point, the NFVO exposes the NS lifecycle to the OSS/BSS and sends notifications if changes occur. This reference point is also serving OSS/BSS to send notifications to NFVO if faults (e.g. congestion) are occurring within the network function from an application logic point of view.
- *EM-VNFM reference point* – Ve-Vnfm-em, is a reference point between EM and VNFM. Through this reference point the VNFM exposes the VNF lifecycle management to the EM. The EM is ensuring the FCAPS functions for the VNFs.

IV. REQUIREMENT FULFILMENT

This section analyses how the presented architecture addresses and fulfils the requirements listed in Section II.

Req.1 QoS & Cost Management

To ensure QoS and improve network efficiency by reducing CAPEX and OPEX, the strategy from industrial partners, network operators, equipment vendors and standardization bodies, like ETSI, is to deploy systems for intelligent traffic steering. The shifting from operative switching and forwarding into programmable and configurable functions enable autonomous network management. The CogNet solution follows this trend by providing an additional layer of network autonomous intelligence which offloads tasks that in legacy environment are performed by manual intervention. This enables a significant reduction in service deployment time and large savings in energy efficiency. Eventually, it can be translated in OPEX reduction for the operator. More in detail the CSE provides means to facilitate rapid provisioning of new services as well as an optimization of the utilization of underlying infrastructure. This is realized by integrating the CSE and an NFV framework which drives the configuration of service instances in the operator domain.

The CogNet solution will provide enhanced and distributed monitoring, which collects information from active/passive probes and measured points on both network and end-user domains, capturing the metrics and signals needed for the CSE. Technologies like OpenFlow and OpenDaylight support data collection from nodes for forwarding network and service infrastructure monitoring. By this way, CogNet will retrieve performance metrics in an event basis to track network activities. Meanwhile, it will offer situational context vis its monitoring that provides a qualitative assessment of network status. This situational context will be used by CSE to provide a run-time evaluation of network status and to provide suitable policies to drive network infrastructure reconfiguration in order to dynamically meet end-user SLA requirements.

In parallel the External Interfaces will provide additional information channels to the CSE for the demand prediction and provisioning. This way, Machine Learning algorithms will process inputs such as: client metrics, forwarding network metrics, service infrastructure metrics, and business thresholds.

Furthermore, the CogNet solution provides to the Policy Manager a gateway to apply countermeasures to react to malfunction evidences and to prevent prints of degradation. OpenStack and vSwitch will apply different networking setups according to the triggered policies. For example the application of policies, integrated with OpenMano technology, will allow the network to resize and resource itself, using virtualisation, to serve predicted demand.

Req.2 Big Data Analysis & Management

Substantial amounts of data will be generated by a variety of devices and systems in 5G. These data will be collected based on different time intervals and diverse schemes. Meanwhile, the quality and accuracy of the data are less controllable. The CSE of the CogNet architecture is equipped with multiple types of adaptors that will transform different types of data into the one that can be processed by the engine directly. This addresses the variety of big data.

The sub-component named CogNet Data Pre-processing will identify suitable predictor variables and perform a number of operations, such as data transformation and normalisation, to improve the quality of collected data and tackle uncertainty of data – veracity. Meanwhile, due to the velocity of big data, the CSE may not have a priori knowledge on models suitable for given datasets, and algorithms need to be identified on-the-fly, which is covered by the Algorithm Selection component.

The CSE is equipped with two processing engines to address the challenges raised by the big data. The Batch Processing Engine is intended to offer an accurate insights on vast amounts of data but it comes with high latency whereas the (Near) Real-time Processing Engine deal with a small amount of records with minimal latency. The two engines support each other to handle massive quantities of data by taking advantage of both batch and stream-processing method.

V. CONCLUSION

This paper presents the software architecture designed in CogNet to facilitate the management of 5G networks. It turns networks into flexible, programmable platforms with intelligence to scale up and down by means of the application of Machine Learning, in particular (near) real-time learning. This is the vision of CogNet to address challenges and requirements on QoS, efficiency and data analysis from 5G networks. It is expected that the presented architecture will be implemented and applied in use cases and scenarios of the project to test and validate the solution in the near future.

ACKNOWLEDGMENT

This work was supported by the EU project CogNet, 671625 (H2020-ICT-2014-2, Research and Innovation action).

REFERENCES

- [1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternina, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the metis project," *Communications Magazine, IEEE*, vol. 52, no. 5, pp. 26–35, May 2014.
- [2] M. Centenaro and L. Vangelista, "A study on m2m traffic and its impact on cellular networks," in *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*, Dec 2015, pp. 154–159.
- [3] A. Imran and A. Zoha, "Challenges in 5G: how to empower SON with big data for enabling 5G," *Network, IEEE*, vol. 28, no. 6, pp. 27–33, Nov 2014.
- [4] T. S. Buda, H. Assem, L. Xu, D. Raz, U. Margolin, E. Rosensweig, D. R. Lopez, M.-I. Corici, M. Smirnov, R. Mullins, O. Uryupina, A. Mozo, B. Ordozgoiti, A. Martin, A. Alloush, P. O'Sullivan, and I. G. B. Yahia, "Can machine learning aid in delivering new use cases and scenarios in 5G?" in *5GMAN Workshop, 2016 IEEE/IFIP Network Operations and Management Symposium (NOMS)*, April 2016.
- [5] M. Sanchez, A. Asadi, M. Draxler, R. Gupta, V. Mancuso, A. Morelli, A. De La Oliva, and V. Sciancalepore, "Tackling the increased density of 5g networks: The crowd approach," in *Vehicle Technology Conference (VTC Spring), 2015 IEEE 81st*, May 2015, pp. 1–5.
- [6] L. Jiang, G. Feng, and S. Qin, "Cooperative content distribution for 5g systems based on distributed cloud service network," in *Communication Workshop (ICCW), 2015 IEEE International Conference on*, June 2015, pp. 1125–1130.
- [7] ETSI, "Network functions virtualisation (nfv); architectural framework," *ETSI GS NFV 002*, 2014.
- [8] IBM, "The Four V's of Big Data," *IBM Big Data & Analytics Hub*, <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>.