ALLIANCE project: Architecting a knowledge-defined 5G-enabled network infrastructure

Universitat Politècnica de Catalunya, Jordi Girona 31, 08034, Barcelona, Spain

ABSTRACT

Leaving the current 4th generation of mobile communications behind, 5G will represent a disruptive paradigm shift integrating access networks, ultra-high capacity access/metro/core optical networks and intra-datacentre network and computational resources into a single converged 5G network infrastructure. Thanks to an extensive deployment of network virtualization techniques leveraged by Software-Defined Networking (SDN) and Network Function Virtualization (NFV) technologies, such a 5G network infrastructure will have to be capable of inter-connecting anything anywhere and over a set of network services truly meeting their diverse communication requirements. In this paper, we overview the ALLIANCE project research activities. ALLIANCE ambitiously aims at architecting a converged 5G-enabled network infrastructure investigating several networking solutions such as SDN/NFV, OpenOverlayRouter (OOR) and the clean-slate Recursive Inter-Network Architecture (RINA) covering access, metro, core and datacentre segments. ALLIANCE relies on a Knowledge-Defined Networking (KDN) orchestrator which take advantage of state-of-the-art machine learning techniques to deploy, operate, monitor and troubleshoot networks automatically.

Keywords: 5G, SDN/NFV, KDN, Open Overlay Router, RINA, orchestration.

1. Introduction

5G paradigm has been gaining momentum in very recent years with many global R&D initiatives launched in major economies worldwide. Referred in particular to Europe, 5G has been recognised as the key enabler for the digitalisation of the European economy [1]. While the current and future 5G applications are countless, many of them share one thing in common. In order to provide a good user experience is no longer enough to provide high bandwidth alone but there is a strong requirement to increase other eight technical parameters [2], ranging from area traffic capacity (an increase of 100x per squared-km) to network energy efficiency (100x more efficient).

In this 5G era, network infrastructure will go well beyond the evolution of today's transport networks and have to be properly designed to support end-to-end 5G-enabled sophisticated vertical applications as envisaged by 5G PPP [3]. In this direction, 5G promises to deliver not only a better performing network, but also one that can become an infrastructure capable of supporting ubiquitous services, while at the same time meeting the performance and commercial requirements of multiple stakeholders.

In this paper, we overview the ALLIANCE project research activities. ALLIANCE is a Spanish project started in 2018, which aims at architecting, from top to bottom, a converged 5G-enabled network infrastructure satisfying those needs to effectively realize the envisioned upcoming Digital Society. In particular, an ambitious goal of the ALLIANCE proposal is to design and implement a Knowledge-Defined Networking (KDN)-based orchestration layer, implementing machine learning techniques toward optimal end-to-end service provisioning. Under the umbrella of KDN, ALLIANCE will investigate the appropriateness of several networking solutions for 5G, such as i) SDN/NFV on top of an ultra-high capacity spatially and spectrally flexible all-optical network infrastructure, and the ii) OOR and iii) RINA over packet networks, including access, metro, core and datacentre networks. Finally, a novel network-monitoring element will be designed to collect relevant information from the networks and the users. This information will be used by the orchestrator in order to make dynamic decisions based on the current (or future) status of the network. Evaluation activities will not only consist of theoretical and simulation-based results, but also experimental activities over representative network test-beds implementing the aforementioned networking solutions for 5G, as a way to completely assess their performance in real network scenarios.

2. ALLIANCE architecture

Figure 1 shows an overview containing a sketch of the general ALLIANCE reference network architecture. It consists of different main blocks. On the top hand side, there is the KDN-based orchestration layer, which will be empowered with cognition techniques (e.g., machine learning tools) to increase the efficiency of the management of the overall resources. On the right-hand side, a management and monitoring element will be in charge of collecting users, networks and IT resources information (i.e., network telemetry) and manage the overall infrastructure. On the bottom side, we consider any network infrastructures including access, metro, core and data centre network segments. Finally, we can observe three different control domains in the centre of the figure: 1) an SDN/NFV domain, 2) an OOR domain, and 3) a RINA domain. It is worth mentioning that these domains will be investigated separately throughout the project and compared at the end. A final outcome of the ALLIANCE project and a significant added value is an exhaustive benchmarking analysis among these technologies/domains to identify their respective advantageous and disadvantageous.

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. DOI: 10.1109/ICTON.2018.8473626. https://ieeexplore.ieee.org/document/8473626
3. KDN-based orchestration

The Knowledge Plane (KP) for the Internet proposed in [4] is a new construct that relies on Machine Learning (ML) and cognitive techniques to operate the network. A KP would bring many advantages to networking, such as automation (recognize-act) and recommendation (recognize-explain-suggest), and it has the potential to represent a paradigm shift on the way we operate, optimize and troubleshoot data networks. However, at the time of this writing, we are yet to see the KP prototyped or deployed. Why?

One of the biggest challenges when applying ML for network operation and control is that networks are inherently distributed systems, where each node (i.e., switch, router) has only a partial view and control over the complete system. Learning from nodes that can only view and act over a small portion of the system is very complex, particularly if the end goal is to exercise control beyond the local domain. The emerging trend towards logical centralization of control will ease the complexity of learning in an inherently distributed environment. In particular, the SDN paradigm decouples control from the data plane and provides a logically centralized control plane, i.e., a logical single point in the network with knowledge of the whole.

The KDN paradigm advocates that the centralized control offered by SDN, combined with a rich centralized view of the network provided by network analytics, enable the deployment of the KP concept. In this context, the heart of the knowledge plane is its ability to integrate behavioural models and reasoning processes oriented to decision making into an SDN network. In the KDN paradigm, the KP takes advantage of the control and management planes to obtain a rich view and control over the network. It is responsible for learning the behaviour of the network and, in some cases, automatically operate the network accordingly. Fundamentally, the KP processes the network analytics collected by the management plane, either preprocessed data or raw data, transforms them into knowledge via ML, and uses that knowledge to make decisions (either automatically or through human intervention).

3.1 Machine learning techniques applicable to ALLIANCE

ML algorithms (such as Deep Learning techniques) are the heart of the KP, which are able to learn from the network behaviour. The current and historical data provided by the analytics platform are used to feed learning algorithms that learn from the network and generate knowledge (e.g., a model of the network). We consider three approaches: supervised learning, unsupervised learning and reinforcement learning.

In supervised learning, the KP learns a model that describes the behaviour of the network, i.e., a function that relates relevant network variables to the operation of the network (e.g., the performance of the network as a function of the traffic load and network configuration). It requires labelled training data and feature engineering to represent network data.

Unsupervised learning is a data-driven knowledge discovery approach that can automatically infer a function that describes the structure of the analysed data or can highlight correlations in the data that the network operator may be unaware of. As an example, the KP may be able to discover how the local weather affects the link's utilization.

Finally, in reinforcement learning approach, a software agent aims to discover which actions lead to an optimal configuration. As an example, the network administrator can set a target policy, for instance the delay of a set of flows, then the agent acts on the SDN controller by changing the configuration and for each action receives a reward, which increases as the in-place policy gets closer to the target policy. Ultimately, the agent will learn the set of configuration updates (actions) that result in such target policy. Recently, deep reinforcement learning techniques have provided important breakthroughs in the AI field that are being applied in many network-related fields (e.g., [5]).

3.2 Network orchestrated strategies based on ML techniques

The KP eases the transition between telemetry data [6] collected by the analytics platform (see Section 4) and control specific actions. Traditionally, a network operator had to examine the metrics collected from network
measurements and make a decision on how to act on the network. In KDN, this process is partially offloaded to the KP, which is able to make -or recommend- control decisions taking advantage of ML techniques.

The northbound controller API offers a common interface to, human, software-based network applications and policy makers to control the network elements. The API offered by the SDN controller can be either a traditional imperative language or a declarative one [7]. In the latter case, the users of the API express their intentions towards the network, which then are translated into specific control directives.

The KP can operate both on top of imperative or declarative languages as long as it is trained accordingly. At the time of this writing however, developing truly expressive and high-level declarative northbound APIs is an open research question. Such intent-based declarative languages provide automation and intelligence capabilities to the system. In this context, we advocate that the KP represents an opportunity to help on their development, rather than an additional level of intelligence. As a result, we envision the KP operating on top of imperative languages, while helping on the translation of the intentions stated by the policy makers into network directives.

4. Management and Monitoring technology

4.1 Introduction

The monitoring component of the ALLIANCE architecture is responsible of collecting cross-layer network measurements and providing them to the KDN-based orchestration layer, using a combination of passive and active probing techniques. Network measurements should include relevant performance metrics, such as latency or throughput, but also other information from users, networks and IT resources, including service-specific measurements of voice, video and data traffic.

Such measurements will primarily be used to estimate the Quality of Experience (QoE) of different types of services as perceived by the end users, which is the main metric that will drive decisions in the ALLIANCE control plane. The KDN orchestrator will use the information provided by the Management and Monitoring component as its main input to optimize the end-to-end service quality as described in Section 3. The monitoring technology to be developed should support the three network architectures evaluated in ALLIANCE (i.e., SDN, OOR and RINA). Proposed monitoring methods will exploit the features available in each network technology and consider their particular intricacies to provide a solution that is accurate and cost-effective.

4.2 Monitoring technologies and QoE estimation

Within the framework of the ALLIANCE project, we already developed a first prototype of the monitoring component for SDN/OpenFlow. Our solution (described in [8]) aggregates network packets into flows directly in the OpenFlow devices (leveraging the OpenFlow support to maintain flow measurements in the flow tables) and asynchronously sends traffic reports to the management and monitoring component of ALLIANCE.

The main challenge when implementing such a flow measurement solution for OpenFlow is that the number of entries in the flow tables is limited and small. Consequently, the simple solution of installing an entry per flow is not scalable due to the limited hardware resources available in current OpenFlow devices (i.e., number of TCAM entries and processing power). In [9], we proposed two sampling schemes to reduce this overhead that can be implemented in current off-the-shelf OpenFlow devices in a scalable way, and without interfering with other rules already installed in the device (e.g., forwarding rules). Traffic sampling allows us to control the processing and memory usage in the OpenFlow devices to obtain a certain degree of measurement accuracy.

Given that QoS requirements and QoE expectations for different applications are significantly different, flow-level measurements have to be combined with information about the application that generated each traffic flow. Our system combines Deep Packet Inspection (DPI) and ML to identify different types of network applications and services. In particular, we analyse the HTTP headers and certificates of encrypted (SSL/TLS) connections in order to identify web-based services and encrypted traffic. Moreover, we process the DNS traffic as a complementary source of information to discover the domain names associated to the different flows. For the rest of the traffic, we use a C5.0 decision tree using the information present in the flow-level records as features (e.g., source and destination ports, and IP protocol) together with the size of the first few packets (6 packets at most). We found this combination between DPI and ML to offer the best trade-off between accuracy and overhead in our evaluation scenarios. Details about our traffic monitoring and classification system, and their evaluation, are provided in [9].

In order to estimate the QoE perceived by the end-users from the flow-level measurements and application reports described above, we plan to use machine learning methods similar to those we proposed in our previous works [10][11] to estimate the QoE of video streaming traffic from passive network measurements. The development of these methods is part of our future work. We implemented our flow monitoring prototype with the OpenDaylight controller and evaluated its accuracy in a testbed with Open vSwitch [12]. Our experimental results using real-world traffic traces show that the proposed monitoring and traffic classification methods are accurate and can effectively reduce the resource requirements of flow measurements in SDN environments, while providing enough information to estimate the QoE of different services.
5. SDN/NFV technology
SDN and NFV will be key enabler technologies of 5G networks deployment. On one hand, SDN abstracts the complexity of the underlying network infrastructure by deploying a logically centralized software-based controller. Complementary to SDN, NFV aims to further increase the softwarization of the network. In fact, network functions are removed from specific-hardware and deployed in compute resources as software appliances, named Virtual Network Functions (VNFs). In this context, both SDN and NFV represent a paradigm shift in the networking field towards flexible software implementation.

5.1 Control and Orchestration plane in support of 5G network slices
A key use case for 5G infrastructure is the delivery of network slices customized to the requirements of the vertical services and applications. Generally speaking, a network slice provisioned to a vertical (e.g., automotive industry, IoT, smart city, etc.) is composed by a set of orchestrated network, computing and service functions resources that may encompass multiple technologies/segments and domains. Indeed, network slicing is the enabler to support the verticals on a single shared infrastructure keeping the SLA agreements between the digital service providers and the verticals.

In a general view, a control plane framework manages heterogeneous physical and virtualized resources to deploy network slice instances as a composition of different types of network functions and VNFs. The main role of the control plane is to enable the enforcement of specific configurations at the resources level to support the slice provisioning, tailored to the specific needs of the vertical. Among others, one of the mission of control plane is the maintenance of the slice QoE in order to deliver an optimized service to the vertical user. To this purpose, the QoE performance metrics are monitored and analysed. If necessary, the control plane has to take the most optimal actions needed to re-establish desired quality levels and trigger the necessary actions at the underlying infrastructure level (physical and/or virtual). In support of the runtime QoE optimization, machine learning techniques can be employed to infer future QoE levels so as to proactively act against poor quality situations and determine suitable actions to be taken for QoE levels maintenance.

Although several proposals are already available, a deep consensus on the design of a network slice control and orchestration framework to automatically and dynamically provision and maintain slices still has to be achieved.

5.2 Optical nodes
5G network infrastructure will be required to inter-connecting anything anywhere, over a set of network infrastructures and meeting their diverse communication requirements. This requires flexible network nodes, main responsible of resource allocation, in the sense of frequency slots, and spectral bandwidth in general, for the diverse set of network infrastructures. Previous metro and access networks will be merged by adequate network nodes, in order to provide traditional high capacity residential services and also, providing support to standard Front-hauling solutions and providing access to Cloud-Radio Access Network (C-RAN) resources. On the one hand, Reconfigurable Optical Add&Drop Multiplexer (ROADM) provides flexible support to the C-RAN resources, very sensible to optical transport impairments. On the other hand, Optical Cross Connects (OXC) will provide a flexible reallocation of resources, augmenting the flexibility and capacity of network virtualization technologies and implementing physical layer solutions proposed and analysed by deep learning technologies. Therefore, network nodes will be managed by SDN orchestrator, integrating these key network elements with KDN control and management planes. Cost and energy consumption are key for any future deployment. Hence, a design of the network nodes as simple as possible is crucial. A first design for this DWDM ROADM has been already published, with an implementation cost of one order of magnitude lower than a typical Colourless and Directionless (CD) ROADM based on commercial Wavelength-Selective Switches (WSSs) [13]-[16]. Hardware implementation of the design DWDM ROADM has shown very good IL and no generation of transport impairments to the very sensitive 5G modulation formats [16]. For an optimal design of the cost-effective DWDM ROADMs, an iterative process has been developed to determine the characteristics of the DWDM ROADMs, showing that the DWDM ROADMs with optimal design allocate 80% of the total traffic of a network using WSS-based ROADMs with Frequency Slot (FS) granularity [13]. Finally, ALLIANCE’s nodes will be extended to future designs supporting the 5G-supported autonomous-vehicle market. A first analysis of traffic requirement will be presented in [17].

6. OpenOverlayRouter technology in support of 5G
OOR is an open-source software router to deploy programmable overlay networks [18]. OOR leverages the Locator/ID Separation Protocol (LISP) [19] to map overlay identifiers to underlay locators, and to dynamically tunnel overlay traffic through the underlay network. OOR runs in user-space and is available on multiple platforms, such as Linux, Android or OpenWRT home routers [20]. OOR offers a flexible, portable, and extensible overlay solution, making it easy for researchers and networking enthusiasts to develop and test new solutions based on overlay networking.

Identifier-Localtion separation technologies are a key building block for 5G. At the present moment, within 3GPP’s 5G standardization effort, LISP is a potential candidate to replace the GTP tunnelling protocol in N9.
Interface [21]. In this context, OOR is an excellent platform to prototype mobility features for 5G, thanks to its LISP Mobile Node and Tunnelling Router implementations.

OOR also focuses on SDN architectures for end nodes, i.e., enabling network programmability at the edge. In such scenario, it is possible to push network policies directly to the very origin of the traffic, thus enabling interesting applications such as load balancing or bandwidth aggregation for each individual node in the network. OOR is currently available on Android phones and an iOS implementation is underway [22]. These mobile flavours are key enablers of the aforementioned SDN architecture, paving the way for rapid development and testing of new features.

7. RINA technology

Over the last several years, large research attention has been given to clean-slate network architectures for the Future Internet, capable of efficiently and effectively solving the well-known limitations of the current TCP/IP-based Internet architecture, e.g., in terms of routing scalability, application-specific Quality of Service (QoS) delivery or built-in security. In this context, RINA [23] has emerged as a very promising architectural solution to address these challenges.

7.1 Architecture

RINA is a back to basics approach learning from the experience with TCP/IP and other technologies in the past, which reminded us that from the earliest days, networking was viewed as Inter-Process Communication (IPC). Thus, RINA starts from the premise that networking is IPC and only IPC. In particular, networking provides the means by which application processes on separate systems communicate, generalizing the model of local IPC.

RINA is based on a single type of layer, which is repeated as many times as required by the network designer. The layer is called a Distributed IPC Facility (DIF) and the application processes instantiated in a node to enrol a certain DIF is called IPC Processes (IPCPs), i.e., that IPCP represents that node in that DIF. A DIF is a distributed application that provides IPC services over a given scope to the distributed applications above (which can be other DIFs or regular applications). A key characteristic of RINA is its design based on the separation of all functions in mechanisms and policies, which dramatically simplifies networking. While all DIFs implement the same two protocols (called Error and Flow Control Protocol (EFCP) and Common Distributed Application Protocol (CDAP)), the specific operation of each DIF can be customized to its particular scope via programmable policies. In this way, for instance, the routing or packet forwarding policies configured in a backbone DIF can differ from those in a datacentre DIF, as their topological characteristics and dynamicity of the supported traffic can differ significantly.

7.2 RINA in ALLIANCE

Current 5G approaches still model the network as a flat collection of physical devices forwarding data between interfaces, hiding the underlying complexity via overlays. This suffers of scalability issue on the one side and still maintain the complexity of managing the network infrastructure on the other side. A completely different focus has been taken in RINA. In fact, RINA unifies the concept of distributed computing and networking and key 5G aspects like virtualization and programmability are essential part of the architecture. This flexible approach could revolutionize the rigid architecture deals in place today and would allow hundreds of service providers and enterprises to use the 5G network.

In the context of the ALLIANCE project, the objective is to investigate the potentiality of RINA as an alternative, yet compatible, solution for the legacy SDN/NFV-oriented 5G network architecture and evaluate its capability in offering digital services with isolated network slices and guaranteed QoS. Indeed, RINA and its development through ALLIANCE will allow for the dynamic commissioning of virtual, dedicated slices of the network, on-demand, for a particular customer or application, optimised for that usage with its particular QoS requirements – and the de-commissioning of those slices again when no longer required. In order to provide application-specific QoS support, RINA incorporates the configurable Quantitative Timeliness Agreement (QTA)-Mux policy in the IPCPs. This policy was proposed and initially prototyped during the FP7 PRISTINE project [24], and further refined in the H2020 ARCFIRE project [25].

Figure 2. Small-scale experimental scenario for initial tests.
During the first year of the project, the scenario depicted in Figure 2 will be experimented in order to evaluate the QoS support that RINA can deliver to heterogeneous applications with different requirements. Different levels of congestion will be considered in the central network between the IRs and the QoS experienced by all injected traffic flows will be analysed, observing if it fits the requirements in terms of packet losses and delay.

8. Summary and expected outcome

In summary, the ALLIANCE project ambitiously aims at architecting, from top to bottom, a converged 5G-enabled network infrastructure capable of supporting ubiquitous services, while at the same time meeting the performance and commercial requirements of multiple stakeholders. The expected outcome of the ALLIANCE project is to deliver three different stand-alone complete network architecture prototypes oriented towards the 5G era, namely the SDN/NFV-based, the OOR-based and the RINA-based solutions and their relative and exhaustive benchmarking analysis to identify their respective advantageous and disadvantageous. Each solution will be controlled by a specific KDN-oriented orchestration layer built according to its particular characteristics. Lastly, it is envisioned a final demonstration of an end-to-end connectivity provisioning between the SDN/NFV and RINA domains connected through the OOR technology.

ACKNOWLEDGEMENTS

This work has been partially funded by the Spanish Ministry of Economy and Competitiveness under contract FEDER TEC2017-90034-C2-R (ALLIANCE project) and by the Generalitat de Catalunya under contract 2017SGR-605 and supported but not funded under contract 2017SGR-1037.

REFERENCES