Enhanced Connectivity in Wireless Mobile Programmable Networks

PhD Thesis
Author: Luca Cominardi
Advisor: Carlos J. Bernardos

Non-technical summary

The architecture of current operator infrastructures is being challenged by the non-stop growing demand of data hungry services appearing every day. While currently deployed operator networks have been able to cope with traffic demands so far, the architectures for the 5^{th} generation of mobile networks (5G) are expected to support unprecedented traffic loads while decreasing costs associated with the network deployment and operations. Indeed, the forthcoming set of 5G standards will bring programmability and flexibility to levels never seen before. This has required introducing changes in the architecture of mobile networks, enabling different features such as the split of control and data planes, as required to support rapid programming of heterogeneous data planes. Network *softwarisation* is hence seen as a key enabler to cope with such network evolution, as it permits controlling all networking functions through (re)programming, thus providing higher flexibility to meet heterogeneous requirements while keeping deployment and operational costs low. A great diversity in terms of traffic patterns, multi-tenancy, heterogeneous and stringent traffic requirements is therefore expected in 5G networks.

Software Defined Networking (SDN) and Network Function Virtualisation (NFV) have emerged as a basic tool-set for operators to manage their infrastructure with increased flexibility and reduced costs. As a result, new 5G services can now be envisioned and quickly programmed and provisioned in response of user and market necessities, imposing a paradigm shift in the services design. However, such flexibility requires the 5G transport network to undergo a profound transformation, evolving from a static connectivity substrate into a service-oriented infrastructure capable of accommodating the various 5G services, including Ultra-Reliable and Low Latency Communications (URLLC). Moreover, to achieve the desired flexibility and cost reduction, one promising approach is to leverage virtualisation technologies to dynamically host contents, services, and applications closer to the users so as to offload the core network and reduce the communication delay. This thesis tackles the above challenges which are detailed in the following.

A common characteristic of the 5G services is the ubiquity and the almost permanent connection that is required from the mobile network. This really imposes a challenge in the signalling procedures provided to get track of the users and to guarantee session continuity. The mobility management mechanisms will hence play a central role in the 5G networks because of the always-on connectivity demand. Distributed Mobility Management (DMM) helps going towards this direction, by flattening the network, hence improving its scala-

bility, and enabling local access to the Internet and other communication services, like mobile-edge clouds. Simultaneously, SDN opens up the possibility of running a multitude of intelligent and advanced applications for network optimisation purposes in a centralised network controller. The combination of DMM architectural principles with SDN management appears as a powerful tool for operators to cope with the management and data burden expected in 5G networks.

To meet the future mobile user demand at a reduced cost, operators are also looking at solutions such as C-RAN and different functional splits to decrease the cost of deploying and maintaining cell sites. The increasing stress on mobile radio access performance in a context of declining revenues for operators is hence requiring the evolution of backhaul and fronthaul transport networks, which currently work decoupled. The heterogeneity of the nodes and transmission technologies inter-connecting the fronthaul and backhaul segments makes the network quite complex, costly and inefficient to manage flexibly and dynamically. Indeed, the use of heterogeneous technologies forces operators to manage two physically separated networks, one for backhaul and one for fronthaul. In order to meet 5G requirements in a cost-effective manner, a unified 5G transport network that unifies the data, control, and management planes is hence required. Such an integrated fronthaul/backhaul transport network, denoted as crosshaul, will hence carry both fronthaul and backhaul traffic operating over heterogeneous data plane technologies, which are software-controlled so as to adapt to the fluctuating capacity demand of the 5G air interfaces.

Moreover, 5G transport networks will need to accommodate a wide spectrum of services on top of the same physical infrastructure. To that end, network slicing is seen as a suitable candidate for providing the necessary Quality of Service (QoS). Traffic differentiation is usually enforced at the border of the network in order to ensure a proper forwarding of the traffic according to its class through the backbone. With network slicing, the traffic may now traverse many slice edges where the traffic policy needs to be enforced, discriminated and ensured, according to the service and tenants needs. However, the very basic nature that makes this efficient management and operation possible in a flexible way – the logical *centralisation* – poses important challenges due to the lack of proper monitoring tools, suited for SDN-based architectures. In order to take timely and right decisions while operating a network, centralised intelligence applications need to be fed with a continuous stream of up-to-date network statistics. However, this is not feasible with current SDN solutions due to scalability and accuracy issues. Therefore, an adaptive telemetry system is required so as to support the diversity of 5G services and their stringent traffic requirements.

The path towards 5G wireless networks also presents a clear trend of carrying out computations close to end users. Indeed, pushing contents, applications, and network functions closer to end users is necessary to cope with the huge data volume and low latency required in future 5G networks. Edge and fog frameworks have emerged recently to address this challenge. Whilst the edge framework was more infrastructure-focused and more mobile operator-oriented, the fog was more pervasive and included any node (stationary or mobile), including terminal devices. By further utilising pervasive computational resources in proximity to users, edge and fog can be merged to construct a computing platform, which can also be used as a common stage for multiple radio access technologies (RATs) to share their information, hence opening a new dimension of multi-RAT integration.