Data Plane and Control Architectures for 5G Transport Networks

Peter Öhlén, Björn Skubic, Ahmad Rostami, Matteo Fiorani, Paolo Monti, Zere Ghebretensaé, Jonas Mårtensson, Kun Wang, and Lena Wosinska

(Invited Paper)

***Abstract*—Next generation 5G mobile system will support the vision of connecting all devices that benefit from a connection, and support a wide range of services. Consequently, 5G transport networks need to provide the required capacity, latency, and flexibility in order to integrate the different technology domains of radio, transport, and cloud. This paper outlines the main challenges, which the 5G transport networks are facing and discusses in more detail data plane, control architectures, and the tradeoff between different network abstraction models.**

***Index Terms*—5G mobile communication, 5G transport, optical transport networks, software defined networking (SDN).**

I. Introduction

THE next generation mobile network (NGMN), 5G, is expected to be introduced around 2020. With 3G and 4G, mobile traffic shifted from traditional telephony services to data, and with the evolution to 5G [1]–[3], it is expected that everything that benefits from a connection will be connected. We will see a massive growth in both traffic and the number of connected devices. As a consequence, there will be a range of devices supporting both user-services and machine-type communications (MTC). The network therefore needs to support requirements from very different services including, e.g., sensor networks, industrial applications, media services, and evolved mobile broadband. Along with this evolution the traffic rates continue to increase and there is an increased focus on low latency for specific applications. At the same time, these services need to be provided in a cost-effective manner.

To address these challenges, densification of macro cells, and new small cells are used on the mobile side. To provide transport connectivity, re-use of infrastructure is important. Several research projects have investigated converged infrastructure [4]–[6] supporting both traditional residential and business access services as well as Ethernet and IP backhaul and fronthaul based on Common Public Radio Interface (CPRI) [7]. Extending the use of dense wavelength division multiplexing (DWDM) technologies provide the needed capacity, while interworking with IP is important for control and management of end-to-end services as well as efficient use of fiber infrastructure because of the aggregation capabilities at the IP layer.

It is also important to provide services in a flexible and programmable way by means of control and orchestration across different technology domains. With end-to-end services deployed in a distributed cloud environment, coordination is needed across the different domains of radio, transport and cloud. In our research, we adopt a hierarchical control and orchestration architecture based on software defined networking [8] (SDN). This makes the network programmable through software by control applications and services. Several efforts are underway, both in standardization and research to enable a fully programmable network [8], [9].

In this paper we extend the work in [10] mainly in the areas of requirements for 5G transport, data plane & deployment architectures. Moreover, three abstraction models for the transport network are presented and compared. In the next section, service requirement in future networks are outlined, followed by the challenges that 5G transport networks [11] will face—new radio requirements, abstraction and programmability, migration of legacy, and best use of new technologies to enable cost-efficient and sustainable deployment. The focus of 5G transport is how these building blocks will provide support needed for NGMNs. Specifically, requirements from emerging 5G radio technology and related use cases are identified and put in the transport network context, both from a data plane and a control plane perspective.

II. Services in Future Networks

Services like mobile broadband and media distribution will continue to evolve in line with our growing dependence on connectivity. Networks will experience huge increases in traffic and will need to service an ever-expanding number of connected devices. These new traffic requirements are driven by evolved and new services: 1) Media—the majority of traffic already in today's network is video, which drives the average traffic and system capacity; 2) Critical MTC—applications with high requirements on avail-ability, bandwidth, and latency, e.g. industrial applications and intelligent transportation systems; 3) Massive MTC—devices and appliances which are deployed in

Manuscript received October 12, 2015; revised December 21, 2015; accepted January 25, 2016. Date of publication February 2, 2016; date of current version March 3, 2016. This work was supported in part by the VINNOVA project "Kista 5G Transport Lab (K5)" and European FP7 projects UNIFY and COMBO.

P. Öhlén, B. Skubic, A. Rostami, and Z. Ghebretensaé are with Ericsson Research, 164 40 Kista, Sweden (e-mail: peter.ohlen@ericsson.com; bjorn.skubic@ericsson.com; ahmad.rostami@ericsson.com; zere.ghebretensae@ericsson.com).

J. Mårtensson and K. Wang are with the Acreo Swedish ICT, 164 25 Kista, Sweden (e-mail: jonas.martensson@acreo.se; kun.wang@acreo.se).

M. Fiorani, P. Monti, and L. Wosinska are with the School of Information and Communication Technology, KTH Royal Institute of Technology, 164 40 Kista, Sweden (e-mail: matteof@kth.se; pmonti@kth.se; wosinska@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2016.2524209

very large numbers, but where each individual unit has lower requirements on, e.g., bandwidth and reliability, e.g., sensor networks; 4) Evolved mobile broadband supporting higher peak rates and overall capacity.

The digital and mobile transformations currently sweeping through industries worldwide are giving rise to innovative cross-sector applications that are demanding in terms of network resources. And so, 5G networks will not only need to meet a wide range of requirements derived from user demand and device development; they will also need to support advanced services—including those yet to be developed. There is also a shift from a model that is operator steered to one that is user driven. Flexibility and operational scalability are key enablers for rapid innovation, short time to market for deployment of services, and speedy adaptation to the changing requirements that modern industry demands.

To ensure that networks will be able to cope with the varied landscape of future services, a variety of forums like NGMN [12], ITU-R, and 5G Infrastructure public private partnership [13] are working on the definition of performance targets for 5G systems. In comparison with 2015 levels, the performance projections for the most demanding radio scenarios that will have most impact on transport networks are:

1) 1000 × mobile data volume per geographical area, reaching target levels of the order of Tbps/km^2
2) 1000 × the number of connected devices, reaching a density of over a million terminals per km^2
3) 5 × improvement in end-to-end latency, reaching as low as 5 ms—as is required by the tactile internet.

However, the maximum levels of performance will not all apply at the same time for every application or service. Instead, 5G systems will be built to meet a range of performance targets, so that different services with widely varying demands can be deployed on a single infrastructure.

Getting networks to provide such different types of connectivity, however, requires flexibility in system architecture. Aside from meeting the stringent requirements for capacity, synchronization, timing, delay, and jitter, transport networks will also need to meet highly flexible flow and connectivity demands between sites—and in some cases even for individual user terminals [12].

Emerging 5G radio capabilities and the convergence of radio access and wireless backhaul have triggered an uptake of fixed wireless technologies as a complement to fixed broadband. As such, 5G radio will increasingly complement and overlap with traditional fixed-broadband accesses.

III. 5G Transport Network

As described in the previous section, new and evolved services are driving requirements on the network performance in terms of, e.g., capacity, latency and flexibility in how services can be deployed and scaled across radio, transport networks and distributed cloud infrastructure. The 5G architecture [14] will support separate network slices, each supporting requirements of different services, and providing isolation between services. As illustrated in Fig. 1, evolved packet core (EPC) functions

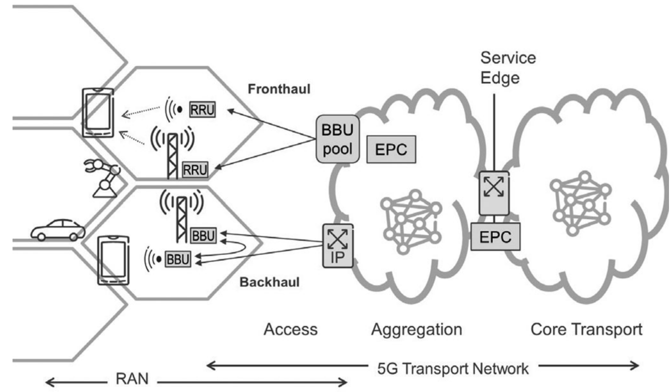

Fig. 1. Simplified 5G reference architecture, showing deployment options (fronthaul and backhaul), and possible locations for EPC for two different network slices.

can be specific to a slices, and placed at different locations depending on the requirements on, e.g., delay constraints. In the radio access network (RAN), there are currently two main deployment options. The lower part of Fig. 1 shows a traditional packet-based backhaul deployment where the baseband processing is performed at the antenna site. There are also lateral connections for RAN functions which require coordination between base stations. The upper part of Fig. 1 shows a centralized RAN (C-RAN) deploy-ment where baseband processing units (BBUs) are centralized in a BBU pool, and digitized radio signals are transmitted to remote radio units (RRUs) over a fronthaul network using CPRI. There are also proposals to introduce a new split of the radio functions between BBU and RRU, leading to a new type of transport interface, often denoted midhaul. In the following, we will detail what requirements transport networks will face and how to adapt to a new and challenging landscape.

A. 5G Radio Requirements

The 5G radio solution is expected to comprise of new radio access technologies (RATs) and evolution of existing RATs exploiting existing spectrum as well as new spectrum (e.g., sub-6 GHz, mm-wave RAT) [2]. Both wider spectrum and technologies such as beam-forming and massive multiple-input and multiple-output (MIMO) [3] will be put in place, resulting in peak rates up to 10 Gbps. As a consequence, transport networks will need to support higher peak rates at the cells sites, and traffic that is burstier than in typical network of today. Multi-site and multi-RAT connectivity provides greater flexibility in how user equipment (UEs) connect to the network and how end-to-end services are setup across radio and transport. For example, allowing for efficient load balancing of UEs among base stations improves user experience and connection performance. A UE that is simultaneously connected to a number of sites may also be connected to several different access technologies. The device may be connected to a macro over long-term evolution (LTE), and to a small cell using a new 5G radio-access technology.

Higher peak rates and traffic volumes will drive transport capacities required at each site. Radio network densification will drive the number of radio sites that need to be connected. Furthermore, radio coordination requirements and new architectural splits of RAN functionality (fronthaul/midhaul) will bring new requirements on the transport network in terms of capacity, latency and jitter as well as node placement. The RAN deployment architecture determines what type of mobile signals needs to be transported, e.g., packet, CPRI, or low latency lateral connections for coordinated multi-point. Available transport capabilities and infra-structure may in turn constrict possible RAN deployment alternatives and radio interference coordination opportunities. For these reasons, the 5G radio deployment determines the requirements on transport, and which level of performance that can be achieved.

The impact of interference may favor deployment models where coordination can be handled more effectively. In small-cell deployments, UEs are often within reach of a number of base stations, which increases the level of interference, and at times requires radio coordination capabilities for mitigation. However, the method used for handling interference depends on how transport connectivity is deployed. In a centralized baseband deployment, tight coordination features can be implemented such as joint processing. Whereas in traditional Ethernet or IP-based backhaul, tight coordination requires low latency lateral connections between participating base stations.

Centralized baseband processing enables enhanced coordination, as radio signals from different antenna sites can be processed jointly. Currently, there is also a growing interest in this approach, as it tends to result in lower operational costs. However, centralized baseband comes at the cost of typically high CPRI bandwidths in the transport network. The high bandwidth, together with stringent delay and jitter requirements, makes dedicated optical connectivity a preferred solution for fronthaul.

In 5G networks, the bandwidth requirements for fronthaul could be very high. The demand will be created by, for example, antennas for multi user MIMO and beamforming—which could use in the order of 100 antenna elements at each location. In combination with dense deployments and wider frequency bands, in the 100 MHz range, traditional CPRI capacity requirements can quickly reach levels of several Tb/s. A new split of RAN functionality is under investigation to satisfy requirements for cost-effective deployments and radio performance, while keeping capacity requirements on transport within a manageable range.

But some primary networking principles remain valid, such as timing and synchronization. Defining new packet-based fronthaul and midhaul interfaces requires that the underlying network includes protocols and functions for time-sensitive transport. Related standardization efforts are also underway [15].

B. Abstraction and Programmability

Abstracting network resources and functionality, and managing services on-the-fly through programmatic APIs, are the pillars of SDN, and its promise to reduce network complexity, and increase flexibility.

With a new split in the RAN, some functions can be deployed on general-purpose hardware, while others, those closer to the air interface with strict real-time characteristics, should continue to be deployed on specialized hardware. Most of the functions of the EPC will be deployed as software—following the concept of network functions virtualization. Deploying network functions in this way makes it possible to build end-to-end network slices that are customized for specific services and applications. Each layer of the network slice, including the transport layer, will be designed to meet a specific set of performance characteristics.

The significance of network slices is best illustrated by comparing applications with different requirements. A network of sensors, for example, requires the capability to capture data from a vast amount of devices. In this instance, the need for capacity and mobility is not significant. Media distribution, on the other hand, is challenged by large capacity requirements, which can be eased through distributed caching. Whereas the network characteristics for remote-control applications based on real-time video are high bandwidth and low latency.

From a 5G-transport perspective, there is a need to provide efficient methods for network sharing, so that applications like these, each with their individual requirements, including mechanisms to satisfy traffic isolation and service-level agreement fulfillment, can be supported for several clients. In addition, distributed network functions need to be connected over links that fulfill set performance levels for bandwidth, delay, and availability.

Transport networks will need to exhibit a high degree of flexibility to support new services. To this end, key features are abstraction and programmability in all aspects of networking—not just connectivity but also storage and processing.

C. Legacy, Migration, and New Technologies

5G transport will be a mix of legacy as well as new technologies. Long-term, network evolution plans tend to include fiber-to-the-endpoint. In practice, however, providing small-cell connectivity needs to take local conditions into consideration, which results in the need for several technologies, such as copper, wireless links, self-backhauling, and free-space optics, to be included in the connectivity solution. Re-use of existing fixed access infrastructure [6] and systems will be important, and new technologies and systems may in turn provide more efficient use of available infrastructure. For example, additional capacity can be provided by extending the use of coarse WDM and DWDM closer to the access segment of the network. Existing infrastructure, together with operator preferences, determines the necessary evolution steps, and how the migration process from legacy to desired architecture should proceed.

The design of 5G transport networks will need to continue to be affordable and sustainable, keeping the cost per bit transported contained. Handling legacy in a smart way, and integrating sustainable advances in technology into packet and optical networks will help to keep a lid on costs.

IV. Data Plane and Deployment Architectures

As described in the previous section there will be different ways to deploy 5G and this will determine the requirements on the transport network. For a centralized baseband deployment

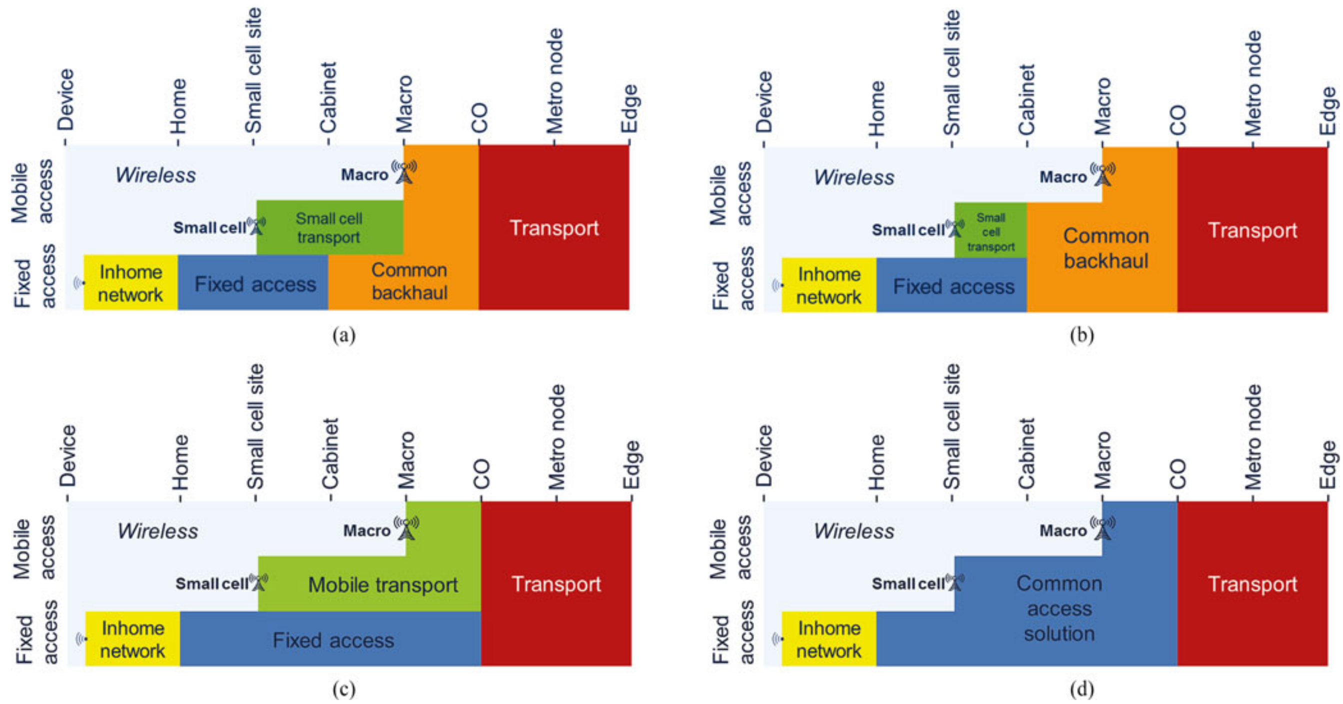

Fig. 2. Illustration of four different infrastructure scenarios: (a) common FTTH-based access solution, (b) dedicated mobile transport, (c) common FTTB/C-based backhaul solution with dedicated small cell transport, and (d) common FTTB/C-based backhaul solution with small cells served directly from the macro sites.

and 5G, the CPRI interface requires orders of magnitude increased capacity compared to backhaul. As an example, a three sector site with 100 MHz bandwidth and 200 Tx/Rx MIMO antennas (100 dual-polarization antenna elements) generates 2765 Gb/s of IQ data (excluding CPRI overhead). Capacity grows linearly with number of served cells as there is no statistical multiplexing gain in IQ transport. By shifting from time-domain to frequency-domain IQ-sampling, bandwidth for the example can be reduced to 802 Gb/s. By furthermore shifting from pure digital precoding to hybrid analog beamforming (assuming eight antennas per RF (radio frequency) chain), required fronthaul bitrate per macro site can be further reduced to 100 Gb/s. Hence, fronthaul architectures and in particular combined with 5G radio technologies push the need for 100G interfaces even further out in the access networks. Especially for the case of massive MIMO solutions investigated in 5G, bandwidth for CPRI are quite high and probably not feasible from a transport perspective.

Fig. 2 shows a number of different infrastructure scenarios. A first scenario (depicted in Fig. 2(a)) is centered around a common access solution. It assumes the mobile operator has access to an fiber-to-the-home (FTTH) infrastructure for deploying cell sites. The infrastructure could be based on passive optical networks (PON), e.g., G-PON or NG-PON2. The common last mile access solution is advantageous for simplified roll out of large numbers of small cells, while it is less suitable for serving high capacity sites or for supporting tight radio coordination where latency between cell sites or a centralized controller is critical. Furthermore, while node consolidation exploiting the reach of the common access solution may be attractive for reducing costs in transport/access, it further inhibits the radio coordination capabilities.

A second scenario (see Fig. 2(b)) shows an alternative FTTH scenario where there is a dedicated solution for mobile transport. The advantage with this scenario is that the mobile transport solution can be tailored for the specific needs of the mobile deployment in terms, i.e., capacities and low latency connectivity. The mobile transport solution can be a completely stand-alone solution or a solution permitting some degree of infrastructure sharing with the fixed access solution. A main drawback is the increased complexity for adding sites which potentially also requires deployment of new infrastructure.

A third scenario (see Fig. 2(c)) relies on a larger number of distributed nodes. To cater for increasing residential traffic volumes, fiber will need to penetrate further out in the field. However, costly in-house cabling may favor deployments such as fiber-to-the-building (FTTB), fiber-to-the-curb. The presence of large numbers of remote nodes implies large operational cost in the access/transport. However, such nodes may also facilitate certain radio performance enhancing features such as radio coordination among small cells.

A fourth related scenario (see Fig. 2(d)) is one where the macro sites are the remote nodes for providing connectivity to the small cells. This architecture would facilitate coordination among small cells and macros. However, the infrastructure differs from what is commonly deployed today. In many cases this scenario would require large investments in infrastructure for connecting small cells directly to the macros.

In conclusion, 5G radio deployment and transport infrastructure are interdependent. 5G radio may dictate requirements

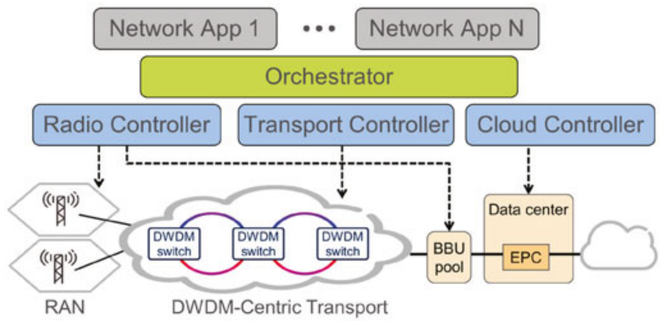

Fig. 3. Hierarchical control architecture with an orchestration layer to bring together different technology domains.

on transport, but availability of transport infrastructure may also dictate the deployment of 5G. Some transport scenarios provide simplified installation of cells at low cost of transport, while other scenarios require more costly transport motivated by increased radio performance and deployment flexibility. In terms of technologies, extending the use of DWDM closer to access enables more efficient use of fiber infrastructure, and lower power consumption. But even today, cost is limiting factor. Advances in integrated optics are important to provide more advanced functionality in a cost-effective way.

V. Control Architecture

There is a wide range of services and requirements that the 5G networks will have to support, including sharing of the network for different purposes. Therefore, flexibility and programmability are important requirements for network control. We adopt an architecture based on SDN for this purpose. In an SDN-controlled network, a logically centralized controller provides an application-programming interface (API), which exposes networking infrastructure capabilities to higher layer control applications and services, and enables them to dynamically program network resources. The impact of such an API goes beyond traditional network control, since this allows deployment of applications on top of the infrastructure to automatically optimize across heterogeneous domains and quickly instantiate new end-to-end services.

Fig. 3 shows the simplified view of how the different domains of radio, transport, and cloud are brought together by an orchestration layer, where different types of resources from several individual control domains are brought together to provide a joint view towards applications or higher-level orchestration. An orchestrator typically operates on a simplified resource view whereas specific domain controllers interface directly to either physical or virtual nodes in the infrastructure. Through the orchestration layer, different applications and services interface towards the underlying infrastructure. There are two sides of this: 1) exposing capabilities and topology from the infrastructure, and 2) providing programmable control of available resources. In exposing the infrastructure, different levels of detail can be provided depending on the requirements from services. For some services topology and delays are important to optimize performance, whereas other services only require connectivity between end-points. Different models have been investigated to balance such requirements. The details of these models are presented in the sections below along with an analysis of their performance. As multiple applications will be interfacing to one orchestrator, separation of control through slicing or virtualization is a key feature. Different applications/services will interface through the programmable API to control its allocated resources to enable flexible control, without impacting parallel services. The following main drivers for flexibility are identified:

1) Rapid and dynamic service provisioning.
2) Optimizing resource utilization.
3) Rapid scale-out of service capacity.
4) Automation of network and services.

There are several scenarios where flexibility would be beneficial. Joint orchestration between radio, cloud & transport was demonstrated [16] to show how varying traffic demands can be utilized to enable infrastructure sharing between two different clients—a CPRI-based RAN, and datacenter bulk transfer.

A. Transport Abstraction Models

The process of exposing resources to the orchestrator is referred to as network abstraction. The challenge in defining the optimal level of network abstraction lies on the fact that, on the one hand, we would like to hide as much details as possible from the orchestrator in order to limit its complexity and make the control architecture more scalable. On the other hand, the more information is available at the orchestrator the more effective can be the outcome of the resources orchestration work. The best level of resources abstraction in SDN-controlled transport networks is currently subject of extensive research. In [17] an abstraction model for packet-over-optical transport networks is proposed. It enables the establishment of connectivity services with low blocking probability, while limiting the complexity of the control infrastructure. In [18] three transport abstraction models are presented and qualitatively compared in terms of tradeoff between performance and complexity. Furthermore, the optimal level of transport resources abstraction is currently under discussion within standardization bodies, e.g., in [19].

However, the transport abstraction models proposed so far usually fail to take into account the fact that processing resources can be embedded in some nodes of the transport network. In this scenario, the network performance can be significantly affected by the orchestrator's knowledge about the reachability of the nodes connected to processing resources. An example is a C-RAN architecture in which baseband hotels (BBHs) are connected to some nodes of the optical transport infrastructure (see Fig. 3). Three transport abstraction models for this scenario have been recently proposed in [20].

The *big switch* (BS) presents the network to the orchestrator as a single node (switch). Consequently, the orchestrator does not possess any information about the internal network connectivity and resources availability.

The *virtual link with constant weight* (VLCW) presents the network to the orchestrator as a set of virtual links between

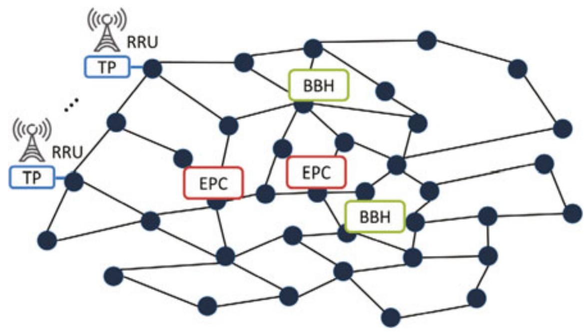

Fig. 4. Reference C-RAN transport topology.

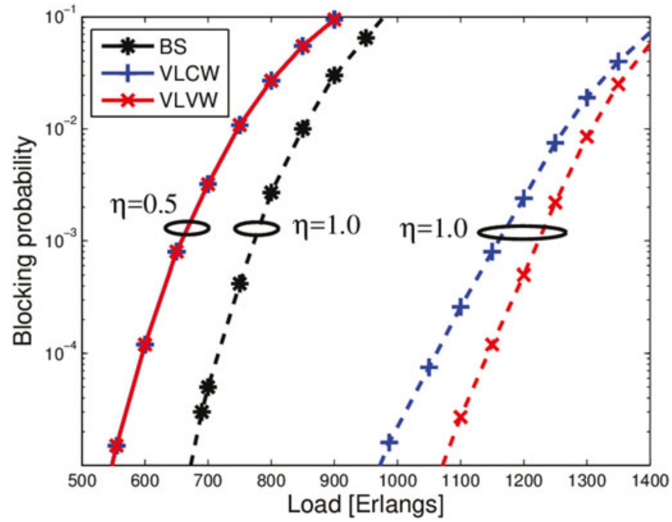

Fig. 5. Blocking probability of the proposed transport abstraction models as a function of the network load and for different values of the C-RAN dimensioning parameter η.

the ingress and the egress nodes (i.e., border nodes). Each virtual link is assigned a constant weight (e.g., representing the shortest path distance between the border nodes). When the connectivity between two border nodes is lost (e.g., it is not possible to set up a lightpath between them anymore) the corresponding virtual link is deleted from the abstract topology. As a result, the orchestrator is always aware of which egress nodes can be reached by any ingress node.

Finally, the *virtual link with variable weight* (VLVW) is a variant of the VLCW in which the weight of each virtual link is proportional to the amount of available network resources (e.g., wavelengths) between the border nodes and it is constantly updated according to the network status. The orchestrator might exploit this information to perform more advanced traffic engineering decisions (e.g., load balancing).

In the following section, we present a comparison between the performance of the BS, VLCW and VLVW using the C-RAN as reference use-case. However, the obtained conclusions can be considered generally applicable to any network scenario in which processing resources are embedded in the transport network.

B. Performance Evaluation of Abstraction Models

To compare the performance of BS, VLCW and VLVW we implemented an event-driven C++ simulator. The simulated use case is a C-RAN in which the transport traffic is carried over a flexible DWDM network with mesh topology as shown in Fig. 4. The reference network topology comprises of 38 optical nodes and 59 fiber-links. Each fiber is equipped with 256 wavelengths. RRUs are connected to the optical DWDM nodes using tunable transponders. In addition, two of the optical nodes are connected to BBHs and two other nodes act as hubs toward the mobile EPC nodes. Every time a RRU requests a connection (e.g., after being activated) the orchestrator is responsible for establishing: 1) a fronthaul connection between the RRU and a BBH with available processing resources, and 2) a backhaul connection between the selected BBH and an EPC hub. The RRU request for connection might be blocked for two reasons: 1) the selected BBH has no available baseband processing resources; 2) the selected BBH cannot be reached because there are no available network resources (i.e., wavelengths). The connection requests generated by the RRUs are supposed to have an exponentially distributed inter-arrival time (i.e., we consider Poisson arrivals), which has been varied to mimic different traffic loads. We evaluate the following performance metrics: 1) blocking probability, i.e., probability that a RRU connection request arriving at the orchestrator cannot be accommodated because the required resources are not found; 2) update complexity, i.e., average number of update messages exchanged between transport controller and orchestrator in order to maintain the abstract topology. The simulation results presented in the following have a confidence interval not exceeding 6%, with 95% confidence level.

Fig. 5 shows the blocking probability of BS, VLCW and VLVW as a function of the network load. The results are shown for two possible C-RANs dimensioning. These are reflected by the parameter η, which is the ratio between the total number of BBH resources available in the network and the maximum number of BBH resources that can be potentially reached via the available optical resources. Plausible values for the η parameter are comprised between 0.5 and 1.0. The case $\eta = 0.5$ means that BBH resources are scarce compared to optical resources, so that a connection request from a RRU to the EPC is more likely to be blocked because of the lack of available baseband processing. As it can be observed from fig. 3, with $\eta = 0.5$ the BS, VLCW and VLVW present the same blocking probability. The case $\eta = 1.0$ means that the amount of BBH resources corresponds exactly to the amount of BBH resources that can be potentially reached via the available optical network resources. In this case, a connection request from a RRU to the EPC is more likely to be blocked because of the lack of network resources. Fig. 3 illustrates that with $\eta = 1.0$ the BS presents up to three orders of magnitude higher blocking probability with respect to VLCW and VLVW. As a result, it is possible to accommodate roughly 50% more load with the same blocking requirements. By exploiting traffic engineering (i.e., load balancing), the VLVW shows the best blocking performance and achieves almost one

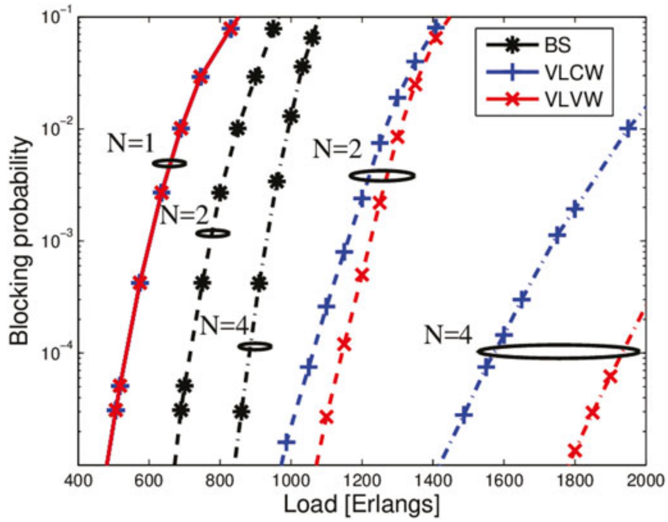

Fig. 6. Blocking probability of the proposed transport abstraction models as a function of the network load and for different numbers of BBHs (*N*).

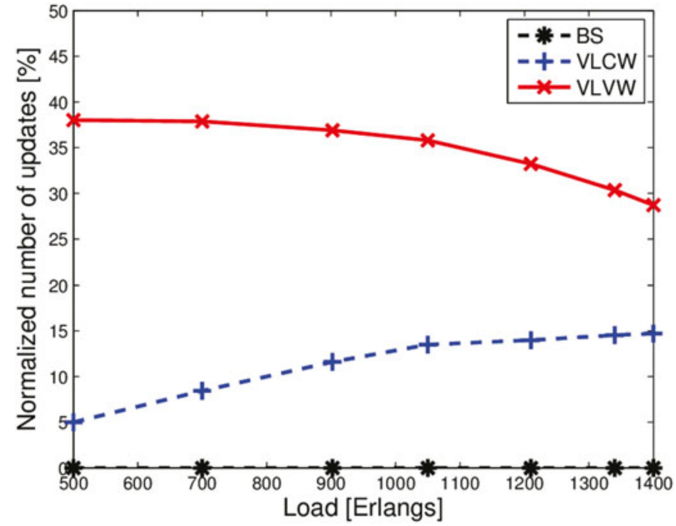

Fig. 7. Number of updates required by the proposed transport abstraction models as a function of the network load.

order of magnitude lower blocking probability at moderate/high loads compared to the VLCW. It is then possible to draw the following conclusion. If BBH resources are scarce, complex network abstraction models are not needed. Contrariwise, if BBH resources are sufficiently available, detailed transport abstraction models (such as VLVW) are required to achieve good blocking performance.

Fig. 6 shows the blocking probability of the BS, VLCW and VLVW as a function of the load and for different numbers of BBHs (*N*). We assume that each BBH is connected to a different node of the DWDM mesh network. In addition, we set $\eta = 1.0$, so that the blocking probability is dominated by the lack of network resources. It can be observed from fig. 4 that with a single BBH (i.e., $N = 1$) the three abstraction models show almost the same blocking probability. This is due to the fact that with $N = 1$ the orchestrator has limited possibility to perform a clever resources allocation. In fact, all the traffic needs always to be routed through the single BBH, which becomes the bottleneck of the network. Increasing N leads to significantly decreasing the blocking probability and increasing the relative difference among the performance of the abstraction models. With $N = 4$ the VLVW shows a blocking probability up to five orders of magnitude lower than BS, and up to three orders of magnitude lower than VLCW. In terms of load we can support roughly twice the amount of requests. The conclusion is then twofold: 1) if a single centralized BBH is used, complex abstraction models are not needed; if several distributed BBHs are employed, more sophisticated abstraction models might be required to achieve good network performance; 2) the use of a single BBH to serve a C-RAN with meshed DWDM transport might lead to high blocking probability; as a consequence, a C-RAN design with several distributed BBHs is preferable from performance perspective. From a reliability perspective a realistic deployment would also have a number of BBHs for redundancy reasons in failure situations.

Finally, in Fig. 7 we show the update complexity of the BS, VLCW and VLVW as a function of the load. We assume $\eta = 1.0$ and $N = 2$. The number of updates of the proposed models is normalized against the maximum number of updates, which would be required in case that the orchestrator is given full knowledge of all the resources in the optical transport network. The BS does not require any update to maintain the abstract topology, making it much simpler and scalable than the VLCW and VLVW. The VLCW updates the abstract topology only if the connectivity between two border nodes is lost. As a consequence, its update complexity is relatively small. This is especially true at low loads, for which it is rare that wavelength resources between nodes expire. The number of updates of the VLCW increases with the load till reaching a maximum of around 15%. On the other hand, the VLVW shows by far the highest complexity. The number of updates of VLVW is higher at low loads because very frequent updates are required to maintain the correct weights for all the virtual links in the abstract topology. The complexity decreases with the load because some virtual links are deleted from the abstract topology (i.e., once the corresponding resources are expired) and thus they do not require updates.

VI. Conclusion

There are large expectations on 5G to provide cost-effective and sustainable wireless connectivity and services to billions of things, people, enterprises, applications, and places around the world. To enable these opportunities for business and society, the architecture of 5G systems—and transport in particular—needs to be built for flexibility through programmability.

5G transport needs to support requirements from new radio technology and deployment models, as well as providing exposure capabilities and programmability for a wide variety of services. Performance requirements will be quite stringent both for some services and from the radio system itself, in terms of bandwidth, latency and reliability. An SDN-based approach is adopted to enable flexible control, and we analyzed the impact of different abstraction models for a centralized baseband deployment. We further identified that different abstraction models are more suited for certain scenarios, in particular that a scarce resource situation in the network calls for a more detailed abstraction to maintain good performance.

Delivering the required level of flexibility, needs tighter integration between 5G radio, transport networks, and cloud infrastructure. This must be carried out on the backdrop of small-cell deployment, convergence of access and backhaul, and migration of legacy equipment and technologies—while containing costs.

Here we also identify a dependence between the transport infrastructure and the radio deployment. We outlined four main architecture alternatives in the access/aggregation domain, and discussed how some of these are more suited for the requirements of 5G RAN in terms of latency and flexibility.

To meet capacity demands, increasing the use of DWDM closer to access will be feasible when flexible optics become more cost-efficient.

In short, the major challenges for 5G transport are programmability, flexibility, and finding the right balance of packet and optical technologies to provide the capacity demanded by the Networked Society.

References

[1] A. Osseiran *et al.*, "Scenarios for the 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[2] F. Boccardi *et al.*, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[4] S. Gosselin *et al.*, "Fixed and mobile convergence: Needs and solutions," in *Proc. 20th Eur. Wireless Conf.*, 2014, pp. 1–6.

[5] B. Skubic *et al.*, "Rethinking optical transport to pave the way for 5G and the networked society," *IEEE/OSA J. Lightw. Technol.*, vol. 33, no. 5, pp. 1084–1091, Mar. 2015.

[6] COMBO FP7 project: Convergence of fixed and Mobile Broadband access/aggregation networks. (2016). [Online]. Available: http://www.ict-combo.eu/

[7] CPRI Common Public Radio Interface. (2016). [Online]. Available: www.cpri.info

[8] Open Networking Foundation, "Software-defined networking: The new norm for networks," ONF White Paper, 2012.

[9] FP7 UNIFY project. (2016). [Online]. Available: https://www.fp7-unify.eu/

[10] P. Öhlén, B. Skubic, A. Rostami, Z. Ghebretensae, J. Mårtensson, K. Wang, M. Fiorani, P. Monti, and L. Wosinska, "Data plane and control architectures for 5G transport networks," presented at the European Conf. Exhibition Optical Communication, Valencia, Spain, 2015, Paper We.4.5.1.

[11] P. Öhlén, B. Skubic, A. Rostami, K. Laraqui, F. Cavaliere, B. Varga, and N. Fonseca Lindqvist, "Flexibility in 5G transport networks: The key to meeting the demand for connectivity," *Ericsson Technol. Rev.*, vol. 92, no. 8, pp. 2–15, Oct. 2015

[12] NGMN Alliance. (2015, Feb.). NGNM 5G White Paper [Online]. Available: https://www.ngmn.org/uploads/media/ NGMN_5G_White_Paper_V1_0.pdf

[13] 5G PPP. (2015, Mar.). 5G vision: The next generation of communication networks and services [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1. pdf#page = 8

[14] J. F. Monserrat, H. Droste, O. Bulakci, J. Eichinger, O. Queseth, M. Stamatelatos, H. Tullberg, V. Venkatkumar, G. Zimmermann, U. Dotsch, and A. Osseiran, "Rethinking the mobile and wireless network architecture: The METIS research into 5G," in *Proc. Eur. Conf. Netw. Commun.*, Jun. 23–26, 2014, pp. 1–5.

[15] IEEE, 802.1CM - Time-Sensitive Networking for Fronthaul, ongoing standardization effort. (2016). [Online]. Available: http://www.ieee802.org/1/pages/802.1cm.html

[16] A. Rostami *et al.*, "First experimental demonstration of orchestration of optical transport, RAN and cloud based on SDN," presented at the Optical Fiber Communication Conf. Exhibition, Los Angeles, CA, USA, 2015, Paper Th5A.7.

[17] P. Iovanna, F. Ubaldi, F. Di Michele, J. P. Fernandez-Palacios Gimenez, and V. Lopez, "E2E traffic engineering routing for transport SDN," presented at the Optical Fiber Communication Conf. Exhibition, San Francisco, CA, USA, 2014, Paper W1K.3.

[18] A. Autenrieth *et al.*, "Evaluation of virtualization models for optical connectivity service providers," in *Proc. IEEE Int. Conf. Opt. Netw. Design Model.*, May 2014, pp. 264–268.

[19] D. Cecarelli and Y. Lee. (2015, Jun.). Framework for abstraction and control of transport networks. Internet Eng. Task Force, Freemont, CA, USA. IETF draft. [Online]. Available: https://tools.ietf.org/html/draft-ceccarelli-teas-actn-framework-00

[20] M. Fiorani, A. Rostami, L. Wosinska, and P. Monti, "Transport abstraction models for a SDN-controlled centralized RAN," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1406–1409, Aug. 2015.

Authors' biography not available at the time of publication.