Benefits of Programmability in 5G Transport Networks

M. R. Raza[1], M. Fiorani[2], A. Rostami[2], P. Öhlen[2], L. Wosinska[1], P. Monti[1]
[1]*KTH Royal Institute of Technology, Electrum 229, SE-164 40 Kista, Sweden*
[2]*Ericsson AB, Färögatan 6, SE-164 80 Kista, Sweden*
e-mail address: pmonti@kth.se

Abstract: This paper shows how programmability can improve operators' revenues and it presents a dynamic resource slicing policy that leads to more than one order of magnitude better resource utilization levels than convectional (static) allocation strategies.

OCIS codes: (060.0060) Fiber optics and optical communications; (060.4250) Networks; (060.4256) Networks, network optimization.

1. Introduction

The 5th generation of mobile communication systems (5G) will provide a common platform for offering a variety of networking services (e.g., enhanced mobile broadband, media delivery, industrial applications). An efficient realization of such a platform requires, among other things, a programmable multi-purpose transport network, which flexibly supports the requirements of various services in an end-to-end orchestration with other infrastructure resources, i.e., radio and cloud [1]. Software defined networking (SDN) is a promising technology for introducing the required programmability features into transport networks, and it enables operators to efficiently share their transport network infrastructure resources among different services through *network slicing* [2][3]. In network slicing, several virtual networks (VNs) are created on top of a physical infrastructure and assigned to services (i.e., usually one VN per service). Each VN is created based on the specific needs of the corresponding service and it is allocated a slice of the overall physical network (PN) resources. The allocation of resources to a network slice (i.e., a VN) can be performed either statically or dynamically. In *static slicing*, a fixed set of resources is allocated to a VN for its entire lifetime. This implies that the assigned resources should support the peak service requirements. In the *dynamic slicing*, however, the amount of resources allocated to a VN varies according to the actual service needs. This allows a network provider to leverage on the temporal variation of the resource requirements of the various VNs for improving the overall network resource utilization and potentially its revenues. Adopting dynamic slicing in a network requires more flexible control mechanisms for dynamic reconfiguration of the infrastructure, which can be achieved through SDN. The aim of this paper is to evaluate the benefits of dynamic slicing in comparison to a static slicing approach. For this purpose, dynamic slicing is formulated as a mixed integer linear programming (MILP) problem. The numerical results obtained after solving the MILP formulation for practical scenarios demonstrate that dynamic slicing can significantly reduce the VN rejection probability as compared to static slicing, which in turn can help network providers to accept more services into their infrastructure.

2. Network Architecture

Figure 1 presents the scenario under exam. An orchestrator is responsible for *(i)* assigning resource slices on demand according to the client's requirements, and for *(ii)* reconfiguring each resource slice according to the temporal variation of such requirements. The transport network is assumed to be multi-layer (i.e., it has both a packet and an optical layer as the network presented in [4]) in order to allow for traffic grooming between the VNs.

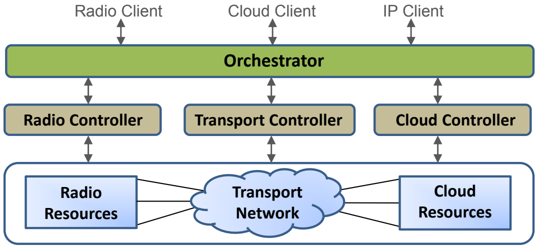


Fig. 1. Programmable multi-purpose transport network architecture with different types of clients.

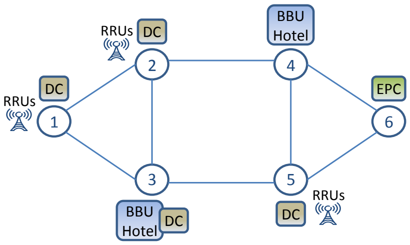


Fig. 2. The 6-node physical network.

Three client types are considered, i.e., *radio*, *cloud*, and *IP*. A radio client asks for connectivity (i.e., capacity) between remote radio units (RRUs) and the evolved packet core (EPC) while traversing a baseband unit (BBU) hotel, following the centralized radio access network architecture. A cloud client needs the allocation of cloud resources (i.e., computes/storage) at selected data centers (DCs) that in turn need to be able to communicate among them. An IP client asks for connectivity between internet service provider (ISP) islands. The radio and IP clients are assumed to have peaks in their requirements during the daytime, while the cloud client experiences them during the night time. Client services are modeled in terms of VN embedding requests, consisting of: *(i)* topology information (i.e., VN nodes/links), *(ii)* resource requirements (i.e., link capacity and node resources), and *(iii)* how such requirements vary over time. This work assumes that only the virtual links capacity requirements change over time. The extension to the case in which both node and link requirements vary is straightforward. If the orchestrator realizes that the fiber link or node resources in the PN are not sufficient to embed a new VN request, the corresponding client request is rejected.

3. MILP Formulation

The proposed dynamic slicing strategy works in two phases: *(i)* mapping of VN requests into the PN, and *(ii)* VN reconfiguration to adapt to the changes in the VN resource requirements over time. Two MILP formulations are presented, i.e., MILP_{arr} for optimal mapping, and $\text{MILP}_{\text{reconf}}$ for optimal VN reconfiguration. MILP_{arr} minimizes the wavelength resource usage in the network. It is derived from the MILP formulations presented in [4][5] and it maps each new VN request without removing or modifying the embedding of any of the VNs already mapped into the PN. MILP_{arr} is not presented in this paper due to space limitation, while $\text{MILP}_{\text{reconf}}$ is introduced next.

Input Parameters and Variables

R^{v}_{sb}	Node mapping (i.e., virtual node s is mapped to substrate node b) of VN v	V	Set of all VNs currently mapped in PN
$h^{v,sd}_{be}$	1 if end points of virtual link s-d of VN v are mapped to substrate nodes b-e	N_s	Set of nodes in the PN
$f^{v,be}_{ij}$	Traffic flowing from node b to e through lightpath(s) i-j for VN v	Nb_m	Set of neighboring nodes of node m in PN
l^{v}_{be}	Traffic that needs to be routed from node b to e corresponding to VN v	SD_v	Set of source-destination pairs of VN v
TA_{ij}	Capacity used on lightpaths between nodes i and j after reconfiguration	$g^{v,sd}_{be}$	XOR of R^{v}_{sb} and R^{v}_{de}
u_{ij}	No. of reconfigurations of lightpaths between nodes i and j	x_{ij}	No. of lightpaths between nodes i and j
d^{v}_{be}	Degradation of virtual link of VN v that is mapped to substrate nodes b-e	ρ^{v}_{sd}	Traffic demand for virtual link s-d of VN v
zq^{ij}_{mn}	Difference between z^{ij}_{mn} and q^{ij}_{mn}	C	Capacity of each wavelength
δ^{ij}_{mn}	1 if z^{ij}_{mn} is greater than or equal to q^{ij}_{mn}	W	No. of wavelengths per fiber link in PN
q^{ij}_{mn}/z^{ij}_{mn}	No. of lightpaths between nodes i and j passing through fiber link (m,n) before/after reconfiguration		

$$\text{minimize}\left(\alpha\sum_{v\,\epsilon\,V}\sum_{\substack{b,e\,\epsilon\,N_s\\ b\neq e}} d^{v}_{be} + \beta\sum_{\substack{i,j\,\epsilon\,N_s\\ i\neq j}} u_{ij} + \gamma\sum_{\substack{i,j\,\epsilon\,N_s\\ i\neq j}}\sum_{\substack{m\,\epsilon\,N_s,\\ n\,\epsilon\,Nb_m}} z^{ij}_{mn}\right)\quad(\mathbf{1}),\qquad \text{subject to}$$

$$R^{v}_{sb}+R^{v}_{de}=g^{v,sd}_{be}+2.h^{v,sd}_{be},\forall v\epsilon V,\forall(s,d)\epsilon SD_v,\forall b,e\epsilon N_s: b\neq e\ (\mathbf{2})$$

$$\textstyle\sum_{(s,d)\,\epsilon\,SD_v}\left(\rho^{v}_{sd}\times h^{v,sd}_{be}\right)=l^{v}_{be},\forall v\epsilon V,\forall b,e\epsilon N_s: b\neq e\ (\mathbf{3})$$

$$\textstyle\sum_{\substack{j\,\epsilon\,N_s\\ j\neq i}} f^{v,be}_{ij}-\sum_{\substack{j\,\epsilon\,N_s\\ j\neq i}} f^{v,be}_{ji}=\begin{cases}(l^{v}_{be}-d^{v}_{be}), & \text{if } i=b\\ -(l^{v}_{be}-d^{v}_{be}), & \text{if } i=e\\ 0, & \text{otherwise}\end{cases},\forall v\epsilon V,\forall b,e,i\epsilon N_s: b\neq e\ (\mathbf{4})$$

$$\textstyle\sum_{v\,\epsilon\,V}\sum_{\substack{b,e\,\epsilon\,N_s\\ b\neq e}} f^{v,be}_{ij}=TA_{ij},\forall i,j\epsilon N_s: i\neq j\ (\mathbf{5})$$

$$TA_{ij}\leq C\times x_{ij},\forall i,j\epsilon N_s: i\neq j\ (\mathbf{6})$$

$$\textstyle\sum_{n\,\epsilon\,Nb_m} z^{ij}_{mn}-\sum_{n\,\epsilon\,Nb_m} z^{ij}_{nm}=\begin{cases}x_{ij}, & \text{if } m=i\\ -x_{ij}, & \text{if } m=j\\ 0, & \text{otherwise}\end{cases},\forall i,j,m\epsilon N_s: i\neq j\ (\mathbf{7})$$

$$\textstyle\sum_{\substack{i,j\,\epsilon\,N_s\\ i\neq j}}\left(z^{ij}_{mn}+z^{ij}_{nm}\right)\leq W,\forall m\epsilon N_s, n\epsilon Nb_m\ (\mathbf{8})$$

$$zq^{ij}_{mn}=z^{ij}_{mn}-q^{ij}_{mn},\forall i,j\epsilon N_s: i\neq j,\forall m\epsilon N_s, n\epsilon Nb_m\ (\mathbf{9})$$

$$\delta^{ij}_{mn}=\begin{cases}0, & \text{if } zq^{ij}_{mn}<0\\ 1, & \text{if } zq^{ij}_{mn}\geq 0\end{cases},\forall i,j\,\epsilon\,N_s: i\neq j,\forall m\epsilon N_s, n\epsilon Nb_m\ (\mathbf{10})$$

$$u_{ij}=\textstyle\sum_{\substack{m\,\epsilon\,N_s,\\ n\,\epsilon\,Nb_m}} zq^{ij}_{mn}\times\delta^{ij}_{mn},\forall i,j\epsilon N_s: i\neq j\ (\mathbf{11})$$

$\text{MILP}_{\text{reconf}}$ dynamically adapts an existing VN mapping to the changes in the link capacity requirements over time. In case of a capacity increase, $\text{MILP}_{\text{reconf}}$ first tries to re-size the existing VN mapping. If this is not possible, it tries to re-map the VN over the existing lightpaths in the PN or over newly established ones. If both re-sizing and re-mapping are not successful, the VN is degraded. The degradation value (*D*) is computed as follows:

$$D=\frac{\int_{t_1}^{t_1+T} C_{req}(t)-\int_{t_1}^{t_1+T} C_{prov}(t)}{\int_{t_1}^{t_1+T} C_{req}\,(t)}.\qquad(\mathbf{12})$$

$C_{req}(t)$ is the total capacity required over time by a VN provisioned at time t_1 with a holding time of T time units. $C_{prov}(t)$ is the total capacity provided over time, i.e., the sum of the capacity provided over each one of the VN virtual links. $\text{MILP}_{\text{reconf}}$ minimizes the number of reconfigurations, the degradation of the VNs mapped in the PN, and the wavelength resource usage (1). α, β, and γ are weighting factors, with α>>β>>γ. Constraints (2)-(3) compute the total traffic demand between two nodes. Constraint (4) is the flow conservation in the IP layer. Constraint (5) computes the total capacity used on each lightpath after reconfiguration. Constraint (6) ensures that the traffic between two nodes is less than the capacity provided by the lightpath(s) between them. Constraint (7) is the flow conservation in the optical layer. Constraint (8) ensures that the number of lightpaths through a fiber link is less than that of the available wavelengths. Constraints (9)-(11) are used for computing the number of reconfigurations. Note that the constraints (10)-(11) can be linearized by using a set of simple linear constraints.

4. Performance Evaluation

The PN is a 6-node network [4]. Each fiber link has 80 wavelengths, each with 100 Gbps capacity. Two BBU hotels, one EPC, and four DCs are randomly distributed in the PN (Fig. 2). It is assumed that BBU hotels and DCs have enough BBU ports and compute/storage resources to guarantee that no VNs are rejected during the node mapping process. Three different VN types are considered for each client (Fig. 3). Their requirements are presented in Table 1. RRUs are randomly placed at the RRU nodes (Fig. 2). ISP nodes are randomly chosen among the network nodes.

Table 1. Radio, cloud, and IP clients' requirements.

Radio	**Cloud**	**IP**
Number of RRUs per node ~ Uniform(5,15)	link capacity for DC-DC connection (Day) ~ Uniform(50,100) Gbps	link capacity for ISP-ISP connection (Day) ~ Uniform(1200,1500) Gbps
Fronthaul (RRUs-BBU) link capacity (Day) = 10 Gbps		
Backhaul (BBU-EPC) link capacity (Day) = 10% of fronthaul	Night traffic variation factor = 25 [7]	Night traffic variation factor = 1/8 [7]
Night traffic variation factor = 1/8 [6]		

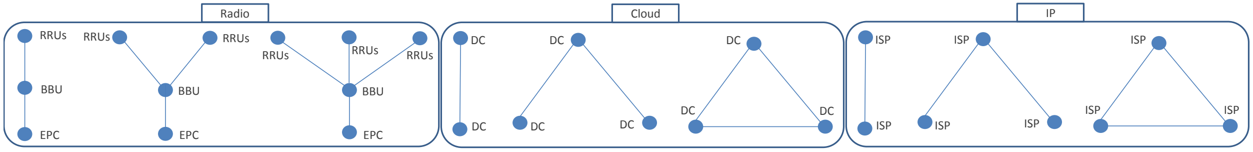
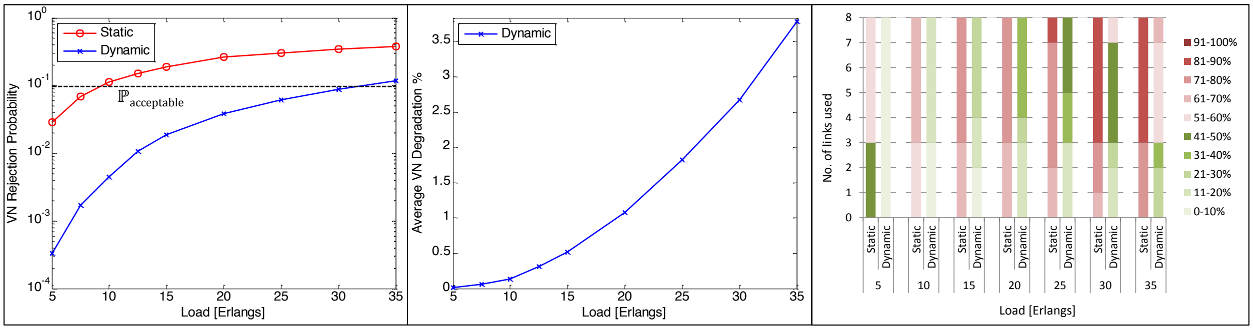


Fig. 3. Service requested by radio, cloud, and IP clients.

Figure 4 compares the performance results of dynamic vs. static slicing, obtained by averaging 50 experiments with 1500 VN requests generated in each experiment. $\text{MILP}_{\text{reconf}}$ is called and solved every 12 hours (i.e., at each day/night time variation) using IBM ILOG CPLEX. The inter-arrival time and the holding time of VN requests are exponentially distributed. The mean holding time is 50 hours, while the mean inter-arrival time is varied between 10 and 1.4 hours. α, β, and γ are set to 10000, 1, and 0.0001, respectively. Figure 4(a) shows that dynamic slicing reduces the VN rejection probability by more than one order of magnitude when the network is in medium to high load condition (i.e., rejection probability < 0.1). This gain in the VN rejection performance comes at a cost in terms of VN degradation. Figure 4(b) presents the value of *D* averaged over all the accepted VN requests during an experiment. This value is very small for low loads and tends to increase at high loads. Nonetheless, it can be noticed that the biggest gains in terms of VN rejection come at relatively low degradation values, i.e., degradation amounts at most to 0.1% at 10 Erlangs. Figure 4(c) shows how many wavelengths are used on average in each of the 8 links in the PN. As expected, by adapting the VN mapping to the actual capacity requirements, the proposed dynamic slicing approach reduces the number of congested fibers on average. This decrease in link usage may help the network providers to accept more VN requests into their physical infrastructure, and hence increase their revenues.



(a) (b) (c)

Fig. 4. (a) Average VN rejection probability, (b) average VN degradation, and (c) average link usage for different values of loads.

5. Conclusions

This paper analyzes the benefits of dynamic vs. static slicing. The results from the proposed MILP-based dynamic slicing strategy show that re-sizing and re-mapping resource slices to match the requirements of each service has the potential to considerably improve the VN rejection probability, and consequently operators' revenues.

Acknowledgments

The work described in this paper was carried out with the support of Kista 5G Transport Lab (K5) project funded by VINNOVA and Ericsson.

References

[1] P. Öhlén, et al., "Data Plane and Control Architectures for 5G Transport Networks," IEEE/OSA JLT, 2016.
[2] M. Fiorani, et al., "On the Design of 5G Transport Networks," Springer PNET, 2015.
[3] A. Mayoral, et al., "Need for a Transport API in 5G for Global Orchestration of Cloud and Networks Through a Virtualized Infrastructure Manager and Planner," JOCN, 2017.
[4] L. Nonde, et al., "Energy Efficient Virtual Network Embedding for Cloud Networks," IEEE/OSA JLT, 2015.
[5] S. Zhang, et al., "Network Virtualization over WDM and Flexible-Grid Optical Networks," J. of Optical Switching and Networking, 2013.
[6] EARTH Deliverable D2.3, "Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown," 2012.
[7] F. Morales, et al., "Virtual Network Topology Reconfiguration based on Big Data Analytics for Traffic Prediction," in Proc. OFC, 2016.