4-6 October 2023, Universidad Carlos III de Madrid, Spain

Book of Abstracts

Steering Committee

Vincent Blondel (University of Louvain)
Leo Ferres (IDSUDD, ISI Foundation & Telefónica)
Marco Fiore (IMDEA Networks Institute & Net AI)
Vanessa Frias-Martinez (University of Maryland)
Renaud Lambiotte (University of Oxford)
Esteban Moro (MIT and Universidad Carlos III de Madrid)
Alex 'Sandy' Pentland (MIT)

Data challenge

Talks

Traffic and Resource Optimization in 5G Multi-Layer Edge Networks

Marcello Pietri, Natalia Selini Hadjidimitriou, Marco Mamei, Marco Picone, Enrico Rossini
University of Modena and Reggio Emilia, Italy
{marcello.pietri,selini.hadjidimitriou,marco.mamei,marco.picone,enrico.rossini}@unimore.it

The next generation of mobile networks - 5G and beyond - will need to support services and applications with a wide range of different requirements in terms of Quality of Service (QoS). The key technology to meet such requirements is the combination of: (i) data-driven solutions and machine learning models to forecast mobile demand and throughput (ii) the virtualization of networks functions and servers to provide operators with unprecedented flexibility on how to allocate resources, re-route traffic and slice the network dynamically

In this research, we propose automating network management through data-driven intelligence, with a particular focus on anomalies and network traffic during specific events or periods (e.g., gatherings, sporting events, concerts, etc.). We analyze the NetMob23 dataset [1] with the goal of forecasting mobile demand for different classes of services, and we provide algorithms to optimize the resources allocated to network slices and optimize traffic distribution (load balancing) within the operator's network to match the demand.

Building upon existing work by [2], [3], [4], the contribution of this paper is to provide a "full-stack" solution dealing with:

- Modeling the mobile network. We create a network model based on a vertical partitioning base on network slices, and an horizontal partitioning based on multiple layers of core and edge computing nodes.
- Demand forecasting. We trained a deep-learning model - based on LSTM - to forecast network demand for different classes of services. We focus on predicting spikes that can saturate network resources and capacity.
- Network optimization algorithms. We developed high-level optimization algorithms to improve network performance by sharing/re-balancing resource among slices, and sharing/re-balancing traffic among network nodes.

In our model the mobile network consists of a set of network nodes / data-centers comprising VNFs and application servers (see Figure 1). Nodes are organized in hierarchical layers. At the bottom there are the *edge* layers. Such network nodes are associated to a subset of BSs and are in charge of managing traffic to/from such BSs. Top layers represent the *core* network, such network nodes are in charge of processing data from lower layers and interacting with cloud services and resources outside the mobile operator network. Network traffic runs from BSs to edge nodes, to core nodes and vice versa.

We assume that the mobile traffic generated by BSs is divided into different classes of services. The resources in each network node are divided between multiple network slices. There is a network slice for each class of service, and each slice is in charge of processing the traffic of that class of service (for example, the slice associated to video-services will handle all the traffic generated by videos). This architecture is general enough to accommodate a wide range of modern network deployments. Finally, in order to link the traffic demand on a node with a measure of performance – following [3], [4] – we introduce, for each node n and slice s, a capacity $C_{n,s}$ representing the amount of traffic that the resources associated to that node in the slice can handle. This is a strong simplification as $C_{n,s}$ subsumes different resources in terms of: networking, memory, storage and computation. The capacity should effectively handle normal traffic but it might be saturated during peak events. Referring to the traffic in a node n in slice s at time t as $T_{n,s,t}$, we assume that the network performance degrades if $T_{n,s,t} > C_{n,s}$. In order to optimize the network operations we will try to minimize the times in which $T_{n,s,t} > C_{n,s}$ and the amount of the excess $T_{n,s,t} - C_{n,s}$.

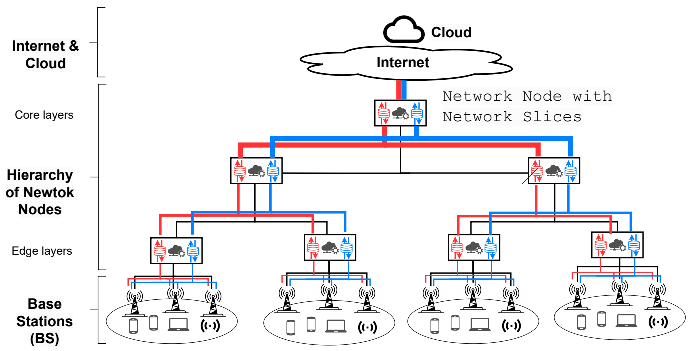

Fig. 1: A schematic representation of the hierarchical and sliced architecture associated with the 5G edge networks.

Given the above scenario, we devised a methodology to optimize network operations on the basis of forecast traffic on the different slices.

Our base hypothesis is that the network allocates resources fairly statically according to the modeled capacities $C_{n,s}$. If traffic is manageable $T_{n,s,t} < C_{n,s}$ the network does not modify its resources. Vice versa, if the traffic is predicted to grow beyond capacity, the network can temporarily adjust resources to meet the surge in demand. Once traffic is predicted to diminish, the network will return to its base allocation $C_{n,s}$.

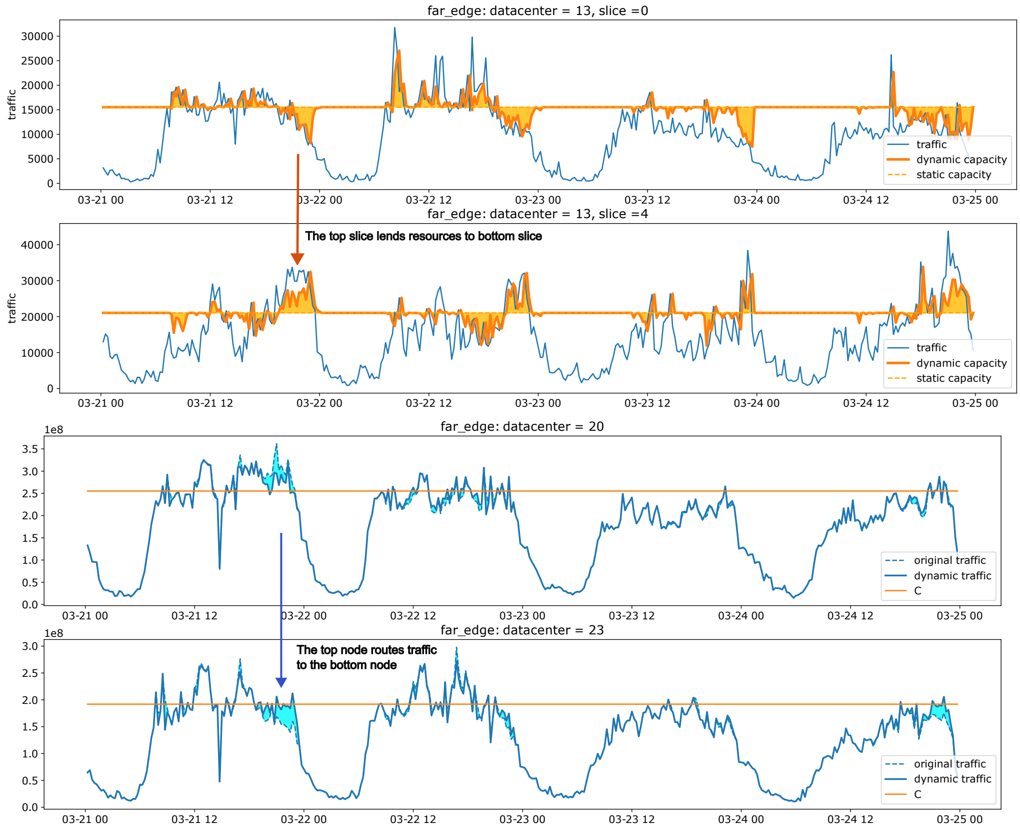

Fig. 2: Dynamic capacity optimization and Dynamic traffic optimization

Our approach is based on two complementary mechanisms: **Resource sharing/re-balancing among slices.** Our network model is based on a set of hierarchical nodes / data-centers processing network traffic. The resources in each node C are divided to handle multiple slices. C_s are the resources allocated to handle slice s ($C = \sum C_s$). If the traffic associated to a slice $s1$ at time t is greater than the capacity $T_{s1,t} > C_{s1}$ the network under-performs. If there is another slice $s2$ in the same data-center with spare capacity $T_{s2,t} < C_{s2}$, some resources could be reallocated from $s2$ to $s1$. The idea of this mechanism is to apply this pattern on the basis of forecast traffic: if the network forecasts that at future time t, $T_{s1,t} > C_{s1}$ and $T_{s2,t} < C_{s2}$, then it re-balances resources in advance, leaving the total amount C unchanged - see Fig. 2 - top. **Traffic sharing/re-balancing among network nodes/data-centers.** The dual approach is to re-balance traffic instead of resources: if the above resource-balancing-within-a-node can not accommodate all the demand ($\sum_s T_{s,t} > C$), a node can re-route part of the traffic to other nodes. This re-route can take place: *(i)* among the nodes of the same network layer, e.g. part of the traffic to a given edge node is routed to a nearby node at the same edge layer (i.e., horizontal offloading). *(ii)* among the nodes at different network layers, e.g. part of the traffic to a given edge node is routed to and processed by a node at an upper layer of the hierarchy - a core node (i.e., vertical offloading) - see Fig. 2 - bottom.

Experimental results show that dynamic reallocation of resources among slices and traffic between nodes improves performance by more than 11% on average.

Acknowledgment. We would like to thank Andrea Buldorini, Edoardo Maria Sanna, Daniela Di Rienzo and Jovanka Adzic from Telecom Italia (TIM) for useful discussion and suggestions in setting up the research and the manuscript.

This work is supported by Ecomobility project co-funded by the European Commission JU-KDT grant agreement No. 101112306 (Research and Innovation Action).

REFERENCES

[1] O. E. Martínez-Durive, S. Mishra, C. Ziemlicki, S. Rubrichi, Z. Smoreda, and M. Fiore, "The netmob23 dataset: A high-resolution multi-region service-level mobile data traffic cartography," 2023.

[2] A. Ceselli, M. Fiore, A. Furno, M. Premoli, S. Secci, and R. Stanica, "Prescriptive analytics for mec orchestration," in *2018 IFIP Networking Conference (IFIP Networking) and Workshops*, 2018.

[3] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, "How should i slice my network? a multi-service empirical evaluation of resource sharing efficiency," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18, New Delhi, India, 2018. [Online]. Available: https://doi.org/10.1145/3241539.3241567

[4] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Pérez, "Resource sharing efficiency in network slicing," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 909–923, 2019.