Assessment of the impact of limited computing resources on vRAN deployment

María Hervás-Gutiérrez, Carlos Baena, Eduardo Baena, Sergio Fortes and Raquel Barco
TELMA - Instituto de Telecomunicación Universidad de Málaga
CEI Andalucía, TECH, E.T.S. Ingeniería de Telecomunicación
Bulevar Louis Pasteur 35, 29010 Málaga (Spain)
Email: {mhervas, jcbg, emb, sfr, rbm}@ic.uma.es

***Abstract*—The new generation of mobile networks goes beyond radio communications by providing a resilient and flexible architecture. In this context, the virtualization of Radio Access Networks (vRAN) completes the Network Function Virtualization (NFV) milestone, enabling a distributed and scalable network architecture. However, this approach increases the complexity of management tasks where computing resources start to play an essential role in the network provisioning process. In this sense, this work aims to assess the impact of computational resources on network performance. The results obtained prove that inadequate resource assignment to vRAN instances leads to service degradation that may remain unnoticed by the network operator. Furthermore, the importance of the vRAN configuration in these scenarios is highlighted, as allocated compute resources can have unintended effects on the quality of service (QoS).**

I. Introduction

5G networks aim to offer a wide range of user cases, also providing novel experiences to users. Thus, several services with heterogeneous requirements are expected to be delivered over the same network infrastructure. In this context, virtualization emerges as a key enabler for mobile network deployments, where cost-efficient and energy-saving solutions are required.

Virtualization makes possible Software Defined Network (SDN) and Network Function Virtualization (NFV) architectures, which provides flexible network deployments thanks to the decoupling of hardware from software [1]. In this way, a resilient and flexible end-to-end Service-Based Architecture (SBA) is possible through the distributed creation of Virtual Network Functions (VNFs).

This approach also includes the radio network side, resulting in the virtualization of the Radio Access Network (vRAN). This means replacing the traditional and physical Base Station (BS) with a virtualized BS (vBS) running on commodity servers in the cloud, enabling the separation of RAN functions from the underlying hardware. Thus, OpenRAN concept has emerged, which aims to implement a RAN based on disaggregated and standardized hardware and software components, moving away from traditional monolithic network environments.

There are, however, several challenges to the OpenRAN paradigm for its successful implementation and operation. These include the management of computational resources for each VNF, which requires a high level of integration across multiple network domains. In this sense, some works such as [2]–[4] have focused on algorithms for dynamically adapting the vRAN split to optimize user throughput and virtualized resource allocation. The authors in [5] propose a learning framework for orchestrating vRAN, considering computational and radio resources.

Nevertheless, no previous work has experimentally demonstrated how underprovisioning affects network performance. In this sense, the present work aims to reveal the effect that a lack of computing resources has on the user's connection. It also seeks to demonstrate the fundamental relationship that exists between the vRAN configuration in terms of bandwidth and the allocation of virtualized resources.

II. Assessing Service Impact

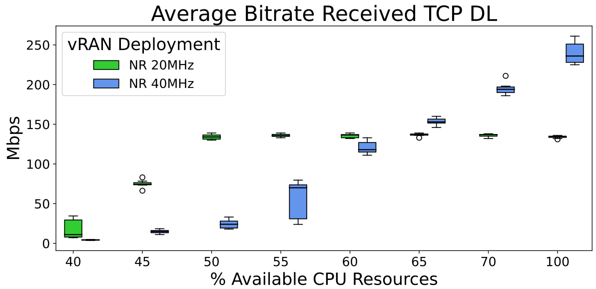
A. Test Scenario

Our vRAN testbed consists of a 4G and 5G Non-Stand Alone (NSA) software-based network deployment. It includes a fully vRAN Release 17 compliant implementation by Amarisoft and the USRP X310 as the Software Defined Radio (SDR) device. For this study, we consider Split 8 (PHY-RF Split) as defined by 3GPP in Release 15 [6]. In particular, we focused on the computational resources used by the VNF of the BS. To do so, the vBS component has been instantiated under a virtual CPU assignment emulator running on a Intel Core i7-10700k at 3.8GHz.

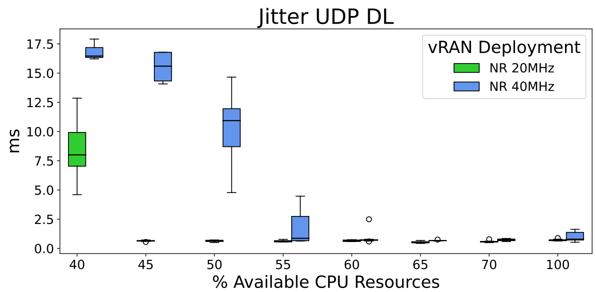
With this deployment, we developed a set of service tests and measured the influence on the QoS provided by limiting the available computing resources in the host machine. We evaluated the impact at the network layer by conducting ping and iperf tests. Service tests are performed with the network fully deployed and with a single user connected to the network through a Customer Premises Equipment (CPE). These requirements are checked each time a measurement campaign is started.

B. Results

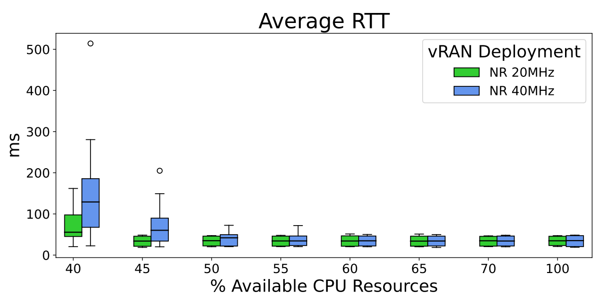
For this study, we focus on bitrate, jitter, and latency in order to narrow down the study. The downlink (DL) bitrate obtained as a result of a TCP iperf test is depicted in Figure 1a. It can be seen that for a vRAN with 20MHz bandwidth configuration, the maximum bitrate obtained is 140Mbps. At this point, it should be noted that the same throughput can be reached with half the CPU resources, and does not improve when more virtualized resources are allocated.



(a) Average Bitrate Received Variation



(b) Average Latency Variation



(c) Average Latency Variation

Fig. 1: Network Performance Impact

Conversely, it can be concluded that the only way to improve this performance is to increase the bandwidth up to 40MHz, where a value of around 240Mbps is attained. Such an improvement, though, requires a greater amount of virtualized resources due to the increased need for processed radio signals. For example, to achieve the maximum bitrate, the new configuration bandwidth requires approximately 70% of the CPU available on the host.

Although it may seem that they are directly related, results indicate that there is a non-linear correlation between bitrate, vRAN bandwidth, and CPU. This means that even if more radio resources are allocated, the network performance is degraded above a minimum CPU capacity.

Such unexpected behavior can be illustrated in Figure 1a, where at 50% CPU allocation, the maximum throughput (i.e. 140Mbps) is reached with fewer radio resources (i.e. 20MHZ). This supposes more than three times the bitrate delivered with 40MHz bandwidth for the same CPU allocation. Moreover, this latter vRAN configuration requires at least 60% of CPU resources to attain similar bitrates as the maximum obtained with NR 20MHz.

Similar behavior is observed for the jitter (Figure 1b), which has been analyzed through UDP iperf test. At 40MHz bandwidth, the jitter increases rapidly when the available CPU drops below 55%. Meanwhile, at a lower bandwidth, there is significant jitter only for the lowest CPU allocation (i.e. 40%). These values follow a similar pattern to the bitrate case (Figure 1a), indicating that more radio resources do not imply better performance.

In terms of the impact on latency, as shown in Figure 1c, there is not much difference between the two radio configurations and CPU values below 45% start to increase latency in both configurations. This effect is more noticeable at 40MHz, but the difference between the two bandwidths is not as significant as in the previous tests.

III. Conclusion

This extended abstract has presented the evaluation of computer and radio resource allocation in terms of end-user QoS. As a result, it has been found that a better allocation of radio resources does not always guarantee better network performance if the mapping of virtual resources is not jointly considered. In addition, the detection of service degradation in this context has been highlighted as a difficult task for network operators.

Future works will focus on the application of this study to adopt an OpenRAN compliant architecture that disaggregates the vCU, vDU and RU to provide intelligent management for an optimal resource allocation. Additionally, further research is needed to examine the impact of other radio parameters on network performance.

Acknowledgment

This work has been partially funded by Ministerio de Asuntos Económicos y Transformación Digital and European Union - NextGenerationEU within the framework “Recuperación, Transformación y Resiliencia (MAORI); and by Ministerio de Ciencia e Innovación (grant FPU19/04468).

References

[1] S. Sun, M. Kadoch, L. Gong, and B. Rong, “Integrating network function virtualization with SDR and SDN for 4G/5G networks,” *IEEE Network*, vol. 29, no. 3, pp. 54–59, 2015.

[2] F. W. Murti, J. A. Ayala-Romero, A. Garcia-Saavedra, X. Costa-Pérez, and G. Iosifidis, “An optimal deployment framework for multi-cloud virtualized radio access networks,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2251–2265, 2020.

[3] F. W. Murti, S. Ali, G. Iosifidis, and M. Latva-Aho, “Learning-based orchestration for dynamic functional split and resource allocation in vrans,” in *2022 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 2022, pp. 243–248.

[4] H. Gupta, A. Antony Franklin, M. Kumar, and B. R. Tamma, “Traffic-aware dynamic functional split for 5g cloud radio access networks,” in *2022 IEEE 8th International Conference on Network Softwarization (NetSoft)*, 2022, pp. 297–301.

[5] J. A. Ayala-Romero, A. Garcia-Saavedra, M. Gramaglia, X. Costa-Perez, A. Banchs, and J. J. Alcaraz, “vrain: Deep learning based orchestration for computing and radio resources in vrans,” *IEEE Transactions on Mobile Computing*, vol. 21, no. 7, pp. 2652–2670, 2020.

[6] *3GPP, "Architecture Description"*, 3rd Generation Partnership Project (3GPP), 2019, release 15.