Optical Network Traffic Analysis under B5G/6G RAN Operation

Shaoxuan Wang, Marc Ruiz, and Luis Velasco*
Optical Communications Group (GCO), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
e-mail: luis.velasco@upc.edu

2023 23rd International Conference on Transparent Optical Networks (ICTON) | 979-8-3503-0303-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICTON59386.2023.10207514

ABSTRACT
The advent of 6G will revolutionize the way Radio Access Networks (RAN) will be operated. Expected massive small cell deployments and features, such as an adaptive functional splitting, are expected to change not only the volume but also the requirements of the traffic to be supported by the fixed transport network. This paper presents an insight into 6G RAN operation, focusing on how such operation will impact the autonomous operation of the fixed network. As concluding remarks of such analysis, key requirements and challenges of fixed network operation for B5G/6G scenarios are identified.
Keywords: 6G, RAN, optical transport network, traffic analysis.

1. INTRODUCTION

Future radio access network (RAN) will operate with massive and heterogeneous small-cell deployments and end-to-end (e2e) connectivity in support of diverse B5G/6G use cases. The optical transmission will play a fundamental role to meet 6G requirements, in terms of capacity and latency. At the same time, energy efficiency and consumption will be a major design criterion in 6G along with other metrics such as capacity, peak data rate, latency, and reliability in addition, cost-effective networks require solutions providing high adaptivity that allow providing just the right capacity, thus eliminating overprovisioning and wasting. This requires near-real-time control that can be supported through closed control loops exploiting zero-touch and intent-based networking paradigms [1]. In fact, a key operational objective in dense and heterogeneous RAN is to reduce energy consumption. This can be achieved by managing the number of active base stations (BS) that are required to support the current user traffic requirements [2]. Note that such operations change the capacity requirements of the fronthaul, and therefore, the optical layer should be able to adapt its capacity in response.

At the same time, the use of *functional splits* [3] allows distributing the processing of the 5G chain between a distributed unit (DU) and a centralized unit (CU), which can be deployed in different sites of the RAN and fixed network. With the disaggregation of the 5G RAN and the definition of different functional splits [3], the requirements for the front-haul (F-H) become stringent [4]. Recently, the adoption of *adaptive function split* is a promising solution that allows adapting dynamically to different quality of service (QoS) requirements, which substantially improves efficiency [5]. In addition, managing the operational mode (*active-sleep*) of BS as a function of current user traffic requirements reduces capacity overprovisioning and energy consumption [6]. Hence, by combining both dynamic RAN capacity management and *adaptive* functional split operation, smart 6G RAN operation can be achieved.

Access and metro optical networks play a fundamental role to meet e2e 6G requirements, in terms of both capacity and latency. Similarly, to the smart RAN operation above, optical networks can operate autonomously, e.g., to adapt optical capacity to current traffic [7]. Typically, these works assume fixed network traffic to behave according to legacy 4G scenarios, i.e., back-haul (B-H) traffic injected by BSs. Nevertheless, the foreseen B5G scenarios dramatically change this assumption, since smart 6G RAN operation generates highly *variable* and *unpredictable* traffic that mixes front-haul (F-H), mid-haul (M-H), and B-H traffic, which is also called an X-haul mechanism. This scenario fits very well with digital subcarrier multiplexing (DSCM) optical systems, due to: 1) point-to-multipoint (P2MP) connectivity can be easily implemented to connect several BSs to the access/metro network [8], which results in cost-saving by reducing transponders (TP) count; and 2) their ability to activate independently each subcarrier (SC) in near real-time [9], which also reduces energy consumption. The problem here is that sharp traffic changes coming from the activation and deactivation of BSs in the RAN could lead to temporal bottlenecks that would produce high delay and even packet loss until the optical capacity is adapted.

In view of this, in this paper, we analyze the impact of smart RAN operation on the traffic injected into the optical transport in 6G scenarios. To this aim, a flow-based traffic model is presented aimed to formally quantify the traffic contribution that each BS introduces to both access and metro segments according to the functional split and BS operational mode. X-haul traffic highly depends on both µBS management and the adopted B5G RAN functional split. RAN capacity is dynamically adjusted by switching on/off µBSs to serve user traffic.

2. 6G REFERENCE SCENARIO UNDER SMART RAN OPERATION

The e2e B5G/6G scenario assumed in this work is depicted in Fig. 1a, where three different network segments, i.e., macro BS (MBS), access-metro, and metro-core, are sketched. First, a RAN cell consists of a single MBS, working at the sub-6 GHz band, and a number of micro BSs (µBS) configured in the mmWave band. MBSs provide full coverage within their RAN cells and provide a minimum capacity to absorb users' traffic, whereas µBSs complement the capacity of the MBS within a limited area. Without loss of generality, we consider that the radio units (RU) of both MBS and µBSs are endpoints of e2e traffic flows. Regarding µBSs, we assume that they

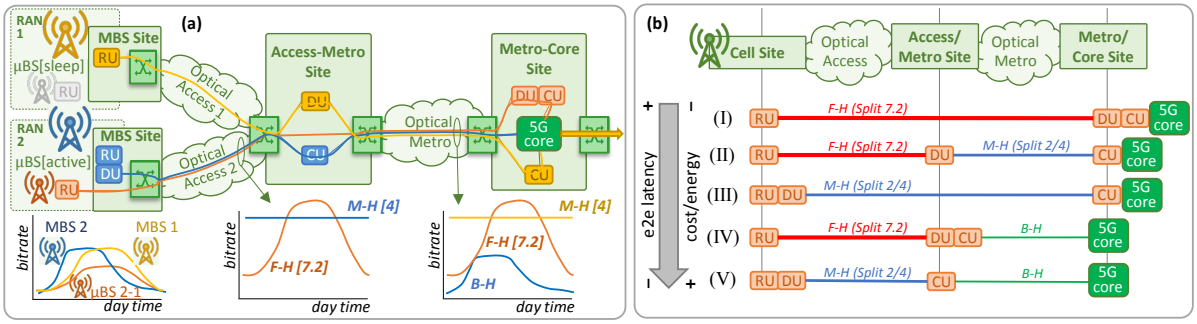

Figure 1. Reference B5G scenario (a) and considered options for functional split and DU/CU placement for flexible split (b).

provide two operational modes: *active*, where the µBS is switched on and fully operative, and *sleep*, where it is switched off. The traffic generated by the MBS (always *active*) and each active µBSs in a RAN cell is injected into the fixed network through an access optical network that connects cell sites with the reference access-metro site. Typically, the distance between both RAN cell and access-metro sites is short, i.e., from a few to tens of km. Besides optical transport and switching capabilities, access-metro sites are equipped with computing resources that enable the deployment of virtualized DU/CU functions. Finally, traffic injected by access-metro sites traverses the optical metro network segment and reaches the reference metro-core site, where the 5G core endpoint is deployed.

As introduced in Section 1, smart RAN operation is built upon two main pillars: *i*) *dynamic capacity management* by means of switching on/off the µBSs in a cell with the objective of reducing energy consumption while ensuring the minimum RAN capacity needed to support users' traffic; and *ii*) *adaptive functional split operation*, where both functional split and DU/CU placement are adapted to match the requirements of every BS in a cell. We consider the five functional split options (*I* to *V*) represented in Fig. 1b, where the e2e latency is reduced, which is opposite to that where DU/CU computing cost (including energy consumption) is increased. Therefore, when user services demand strict latency requirements, higher options will be configured and DU/CU functions are placed close to users, which increases the cost by replicating DU/CU functions. On the contrary, when latency requirements can be relaxed, lower splits will reduce cost by concentrating functions in fewer sites.

Let us now illustrate the impact of using different functional splits on the traffic to be transported by access and metro segments. Specifically, two functional splits are considered, namely 7.2 and 2/4. According to [3], under split 7.2, F-H bitrate is highly correlated with user traffic, as it mainly depends on the actual capacity used in the cell. However, the bitrate of split 2/4 depends on the actual configuration of MBSs and µBSs, so it sharply increases (decreases) when µBSs are switched on (off).

In the example of Fig. 1a, three active BSs are configured with different functional splits emulating different QoS needs: option *II* for MBS 1 (moderately low latency), option *V* for MBS 2 (ultra-low latency), and the option *I* for µBS 2-1 (no strict latency requirements). Assuming a simple static RAN configuration (no changes in time), the three inset graphs (from left to right) show the input traffic in one day served by each BS, the traffic injected in the fixed access network, and the traffic injected in the metro network. It can be observed that different BSs generate completely different traffic patterns in access and metro. For instance, µBS 2-1 injects input-dependent F-H traffic in both access and metro domains (no intermediate function at the access-metro site), whereas MBS 2 injects constant M-H traffic into the access network and, after passing CU function at the access-metro site, generates B-H traffic into the metro network.

On top of the above, smart operation of B5G/6G RAN will manage both µBS operational mode and functional split as a function of the users' traffic. Therefore, a more complex traffic flow model than that for traditional 5G is needed to characterize traffic flows injected in the transport network. Next section is devoted to present such model.

3. FIXED NETWORK TRAFFIC FLOW MODEL FOR B5G/6G SCENARIOS

The proposed model aims at characterizing, for every time t, the access and metro traffic flow components (variables y_{cat} and z_{amt}, respectively), as a function of input traffic at every BS (variables x_{bt}). We assume that a given RAN cell $c \in C$ connects to one single access site $a \in A$ and metro site $m \in M$ and, consequently, all BS $b \in B$ in cell c have the same reference access and metro sites. Independently of where DU and CU functions are placed, the traffic generated from BS b will traverse the fixed access segment until reaching reference access site a and then, will traverse the fixed metro segment from a to reference metro site m. Table 1 summarizes the used notation.

The traffic flow model is defined through the following equations. Eq. (1) models the traffic that a given BS b injects into the access network (y_{bt}). The value is zero if the BS is not active; otherwise, it can be F-H, M-H, or B-H depending on the placement of DU/CU functions. Similarly, eq. (2) characterizes the traffic injected into the metro network (z_{bt}). Note that these two variables do not depend on the actual network configuration, e.g., where a given function is placed. The output is the expected F-H or M-H capacity for each cell i for the next period, $z_i(t+1)$, which depends on the functional split and $y_i(t+1)$ be the traffic monitored at time t. The generic model for split s, based on models in [4], is defined in eq. (3), and where $\eta_{ij}^{s}(t+1) \in [0,1]$ is a factor that scales the

component K_j^s that accounts for the F-H traffic that cell j injects at maximum load when split s is used. The scaling factors η for the considered splits in this work are in eq. (4) and eq. (5). At the same time, C_j is the maximum RAN capacity of cell j. From equations (4) and (5), we clearly observe how F-H traffic in split 7.2 scales proportionally to user traffic, whereas split 2/4 produces a constant F-H traffic per BS. In addition, although component K depends on multiple BS parameters, such as the number of antennas, layers, and chosen modulation format (see [4] for further details), $K_{2/4}^j > K_{7.2}^j$ for any BS j. Eq. (6) and (7) compute the target access and metro traffic flow components, respectively. For a given pair cell-access site <c, a> and pair access-metro site <a, m>, y_{cat} and z_{amt} aggregate the components of every BS that is in cell c and has assigned the access site a and metro site m as reference ones.

Table 1. Parameters and variables.

ρ_b: 1, if BS b is active	δ_{bc}: 1, if BS b is in cell c	x_{bt}: User traffic in BS b at time t [Gb/s]
k_b: Capacity of BS b [Gb/s]	δ_{ba}: 1, if BS b sends to access site a	y_{bt}: Access traffic by BS b at time t [Gb/s]
du_b: Position of DU of BS b [site type]	δ_{bm}: 1, if BS b sends to metro site m	z_{bt}: Metro traffic by BS b at time t [Gb/s]
cu_b: Position of CU of BS b [site type]	fh_b: Max F-H traffic of BS b [Gb/s]	y_{cat}: Traffic in pair <c, a> at time t [Gb/s]
	mh_b: Max M-H traffic of BS b [Gb/s]	$\underline{z_{amt}}$: Traffic in pair <a, m> at time t [Gb/s]

$$y_{bt} = \rho_b \cdot \begin{cases} x_{bt}, & if\ cu_b == "cell" \\ mh_b, & if\ cu_b \neq "cell"\ \&\ du_b == "cell" \\ x_{bt} \cdot \dfrac{fh_b}{k_b}, & otherwise \end{cases} \quad (1)$$

$$z_{bt} = \rho_b \cdot \begin{cases} x_{bt} \cdot \dfrac{fh_b}{k_b}, & if\ du_b == "metro" \\ mh_b, & if\ du_b \neq "metro"\ \&\ cu_b == "metro" \\ x_{bt}, & otherwise \end{cases} \quad (2)$$

$$z_i(t+1) = \sum_{j=0..N} y_{ij}(t+1) \cdot \eta_{ij}^s(t+1) \cdot K_j^s \quad (3)$$

$$fh_{b_{7.2}}(t+1) = \frac{x_{ij}(t+1)}{C_j} \quad (4)$$

$$\eta_{ij}^{2/4}(t+1) = 1 \quad (5)$$

$$y_{cat} = \sum_{b \in B} \delta_{bc} \cdot \delta_{ba} \cdot y_{bt} \quad (6)$$

$$z_{amt} = \sum_{b \in B} \delta_{bc} \cdot \delta_{bm} \cdot z_{bt} \quad (7)$$

4. ILLUSTRATIVE RESULTS

For numerical evaluation purposes, we have built a Python-based flow-based simulator that reproduces the e2e B5G scenario presented in Fig. 1a. To simplify the analysis of access (y_{cat}) and metro (z_{amt}) traffic components, we configured a scenario consisting of one dense RAN cell with 1 MBS and 64 µBS, one access site, and one metro site. We consider typical configurations for MBS (2x2 MIMO, 20 MHz bandwidth) and µBSs (8×8 MIMO, 100 MHz bandwidth). The maximum F-H and M-H for every BS (fh_b and mh_b) for all splits in Fig. 1b was computed from the models in [10]. User traffic was generated following realistic daily patterns and scaled according to [11] to emulate a medium-term scenario with traffic peaks of 60 Gb/s for the whole cell.

With the configuration above, two different RAN operation policies were evaluated: *i*) *static*, where all µBS are always active and all BSs implement the same functional split option; *ii*) *dynamic*, where the split is still fixed but capacity is dynamically adapted by switching on/off µBSs according to actual traffic needs.

Figure 2 shows the performance under *static* RAN operation policy. Figure 2a shows an example of one-day total user traffic and the total capacity, i.e., aggregating MBS and active µBSs. Note that the total capacity remains constant at 96 Gb/s and the traffic usage of UEs is fluctuating with the time of day. Under this configuration, Fig. 2b and Fig. 2c show the traffic at access and metro network, respectively generated with every functional split option. We observe that this policy results in either predictable time-variant traffic (closely correlated with input traffic) or constant traffic, depending on the different functional split options. Additionally, it is worth noting that the traffic volume is dramatically affected by the chosen split, e.g., metro traffic of I and V follow a similar daily pattern but with a largely different magnitude. Moreover, access and metro traffic remain constant in options I and III, whereas drastically vary for the rest of options. In consequence, under key functional splits foreseen for B5G scenarios, metro traffic does not correspond to the aggregated access traffic, which is against the assumptions of typical traffic models and motivates the proposed ones clearly.

Regarding the *dynamic* policy (Fig. 3), we have implemented µBSs switching on/off based on simple threshold-based criteria. Specifically, two rules were implemented at every µBS b: *i*) when the load (input traffic over capacity) of b exceeds 0.6, then the closest to b inactive µBS is switched on; *ii*) when the load of b drops below 0.3 and the closest neighboring µBS is active, b is switched off. These rules provided committed users QoS (no loss) during the whole simulation. Figure 3a shows that capacity savings up to 80% can be achieved when user traffic is minimum (i.e., only 11 out of 64 µBSs are active). In addition, we observe that those options that provided constant traffic in static operation are sensitive to RAN capacity changes, i.e., options III and V in

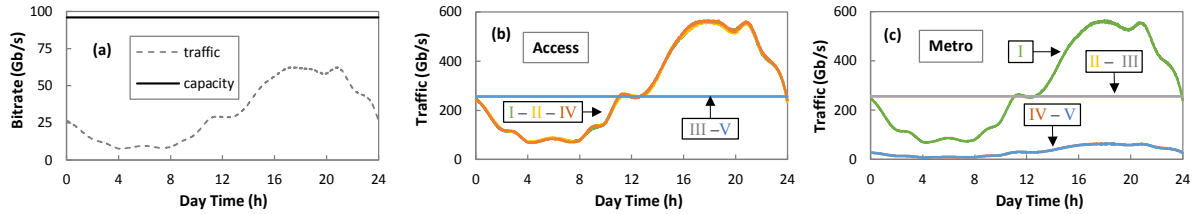

Figure 2. Static operation.

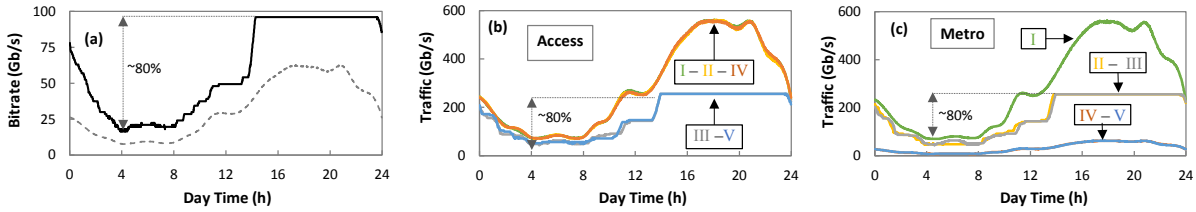

Figure 3. Dynamic operation.

access (Fig. 3b) and II and III in metro (Fig. 3c). In fact, traffic reduction in both segments is equivalent to RAN capacity reduction, which is an outstanding feature of those split options, e.g., to minimize optical capacity requirements and also reduces energy consumption. This comes at the cost of added unpredictability to access and metro traffic since constant periods are combined with varying periods, which hinders those widely used traffic forecast models based on short-term past windows predictors.

5. CONCLUSION

Smart RAN operation in B5G/6G scenarios will induce access and metro network smart operation to implement novel solutions to manage unprecedented variables and sharply changing traffic flows. Autonomous fixed network operation in tight coordination with RAN control is foreseen as a key challenge to achieve target e2e requirements. In this regard, we identify the need of RAN controller to periodically collect user traffic monitoring data gathered from different cells and perform traffic prediction to estimate the expected traffic to be required for the next time interval. Note that this prediction is necessary for deciding which µBSs need to be powered on/off. Then, RAN controller need to be extended with additional modules to perform estimation of the traffic injected in the fixed network, which will depend on both users demand and functional split implemented, as well as on RAN operation approach. That estimation is necessary to allow optical capacity setup, i.e., dynamic allocation of optical SCs based on traffic monitoring and capacity forecasting to be performed autonomously in the optical node agent. This optical capacity update needs to ensure both RAN traffic requirements and local capacity prediction forecasts.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the Smart Networks and Services Joint Undertaking under the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101096120 (SEASON), and the MICINN IBON (PID2020-114135RB-I00) projects and from the ICREA Institution.

REFERENCES

[1] L. Velasco *et al*., "End-to-end intent-based networking," *IEEE Comm. Mag*., vol. 59, pp. 106-112, 2021.
[2] Y. Zhu and S. Wang, "Joint traffic prediction and base station sleeping for energy saving in cellular networks," in *Proc. IEEE Int. Conf. on Communications (ICC)*, 2021.
[3] L. Larsen *et al*., "A survey of the functional splits proposed for 5G mobile cross-haul networks," *IEEE Comm. Surveys and Tutorials*, vol. 21, pp. 146-172, 2018.
[4] G. Perez *et al*., "Fronthaul network modeling and dimensioning meeting ultra-low latency requirements for 5G", *IEEE/OPTICA J. of Optical Communications and Networking*, vol. 10, pp. 573-581,2018.
[5] F. Morais *et al*., "Place RAN: Optimal placement of virtualized network functions in beyond 5G radio access networks," *IEEE TMC 2022*.
[6] Y. Zhu and S. Wang, "Joint traffic prediction and base station sleeping for energy saving in cellular networks," in *Proc. ICC 2021*.
[7] F. Effenberger *et al*., "Fixed 5th Generation Advanced and Beyond," ETSI White Paper, no. 50, 2022.
[8] L. Velasco *et al*., "Autonomous and energy efficient light-path operation based on digital subcarrier multiplexing," *IEEE JSAC 2021*.
[9] A. Bernal *et al*., "Near real-time estimation of end-to-end performance in converged fixed-mobile networks," *Comp. Comm. 2020*.
[10] S. Lagén et al., "Modulation compression in next generation RAN: Air interface and fronthaul trade-offs," *IEEE Comm. Mag*., vol. 59, pp. 89-95, 2021.
[11] Ericsson Mobility Report 2022. [On-line] https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/june-2022.