Measuring 5G and Satellite Backhaul Integration

Mohammad Rajiullah[*], Delia Rico[†], Anna Brunstrom[*†], Giuseppe Caso[*]
[*]*Department of Mathematics and Computer Science*,
Karlstad University, Karlstad, Sweden
{firstname.lastname}@kau.se
[†]*University of Malaga*, Malaga, Spain
deliarico@uma.es

***Abstract*—The convergence of terrestrial and non-terrestrial networks (TNs and NTNs) is a key trend in 5G/6G, yet the performance of their integration remains underexplored. Using a real-world testbed deployment, we evaluate a 5G architecture in which a LEO satellite system provides backhaul between the RAN and core, and compare it against fiber-backed 5G and standalone satellite access. We focus on end-to-end latency and standardized real-time interactive traffic using the e-Gaming profile. Our results show that satellite exit-point placement and core-network location strongly shape latency, and that interactive traffic remains highly sensitive to delay variation and packet loss. Overall, the tested 5G-NTN configuration is promising for coverage extension, but challenging for stringent real-time services.**

I. Introduction

The convergence between terrestrial networks (TNs), e.g., mobile cellular systems, and non-terrestrial networks (NTNs), e.g., satellite systems, is emerging as an important trend across stakeholders, researchers, and standardization fora [1]. From the TN perspective, this is key for enhancing coverage and resilience, particularly in scenarios where TNs might be underdeveloped due to costs (e.g., rural regions) or unavailable due to outages (e.g., in disaster-hit regions).

Within the 3rd Generation Partnership Project (3GPP), the integration between a 5G TN and an NTN has been discussed since Rel-17, with enhancements addressed in Rel-18 and expected in Rel-19, thus moving towards 6G-NTN systems. In general, the NTN is expected to be (part of) the access network and/or (part of) the backhaul network, i.e., to connect the 5G/6G access network with the 5G/6G core network [2].

NTNs include airborne platforms and satellite systems such as Geostationary Earth Orbit (GEO), Medium Earth Orbit (MEO), and Low Earth Orbit (LEO) constellations. Among these, LEO satellites are particularly relevant to 5G/6G integration due to their comparatively low latency and increasing capacity, enabling both broadband and Internet of Things (IoT) services [3, 4].

Starlink, as the currently largest LEO-based Internet service provider, has been widely studied from the end-user perspective, including comparisons with GEO systems, 4G, 5G, and wired access [5–10]. However, most prior studies evaluate satellite and terrestrial systems independently. Empirical analyses of integrated 5G-Starlink architectures remain limited. An initial study [11] explored Starlink backhauling for 5G, but relied on an early-stage setup with SISO gNB antennas. In contrast, we consider a more representative 5G deployment with MIMO and evaluate multiple server locations with different distances from Starlink exit points.

978-3-903176-83-6 ©2026 IFIP

Given the growing interest in satellite-integrated 5G/6G networks, stronger empirical evidence on their practical performance is needed. In this paper, we present a measurement study of a 5G TN using a Starlink backhaul and compare it against two baselines: 5G with terrestrial fiber backhaul and standalone Starlink access. We focus on end-to-end latency and apply a methodology standardized by European Telecommunications Standards Institute (ETSI) [12] and International Telecommunication Union Telecommunication Standardization Sector (ITU-T) [13] to evaluate the QoS/QoE of standardized real-time interactive traffic using the e-Gaming profile [14, 15].

Our main findings can be summarized as follows. First, service latency, in terms of Round Trip Time (RTT), substantially increases when a satellite segment is introduced, and long satellite paths—often caused by the absence of a nearby satellite exit point—can affect latency more strongly than the physical location of the server. Second, baseline RTT alone is insufficient to characterize service performance, since standardized interactive traffic experiences higher delay variation than suggested by ICMP measurements. Third, standardized real-time interactive traffic faces significant challenges over satellite-based paths, often failing to meet service requirements due to increased delay variation and packet loss. Overall, these results highlight the fundamental tradeoff in integrating satellite backhaul into 5G systems: while it enables connectivity extension beyond the reach of terrestrial infrastructure, it also introduces latency and variability that can substantially impair delay-sensitive service performance.

II. Setup, Methodology & Dataset

In this section, we describe the experimental setup, the measurement methodology, and the collected dataset.

A. Setup

The measurements were conducted on a research testbed at the University of Malaga in Spain, where an integrated 5G–Starlink deployment is available. The 5G testbed consists of a standalone 5G network comprising a Nokia AirScale Pico RRH 5G RAN and an Open5GS-based 5G core compliant with

3GPP Release 16. The RAN operates in standalone mode on the n78 TDD band with a channel bandwidth of 100 MHz. We consider three network configurations, shown in Figure 1: two baselines and one integrated 5G–NTN architecture.

i. **5G with wired backhaul (5G-F):** A baseline configuration in which the User Equipment (UE) accesses the Internet through a 5G TN whose core is connected via fiber.
ii. **Satellite-only (NTN):** A baseline configuration in which the UE accesses the Internet through an end-to-end Starlink connection.
iii. **5G with satellite backhaul (5G-NTN):** An integrated configuration in which the Starlink system provides the backhaul between the 5G access network and the 5G core network. Since this configuration combines the 5G access segment with a satellite backhaul, its end-to-end performance is expected to be constrained by the satellite segment and may therefore be worse than NTN.

This placement is relevant to operator deployments in which NTN is used to extend the reach of a terrestrial 5G network while retaining integration with the 5G core.

To study how end-to-end performance depends on architecture and server location, we use three servers: *Server-A*, located locally within the testbed; *Server-B*, located at Telefónica in Madrid, about 500 km away; and *Server-C*, located at Karlstad University in Sweden, about 3500 km away. These servers are evaluated across the three configurations: 5G-F, NTN, and 5G-NTN.

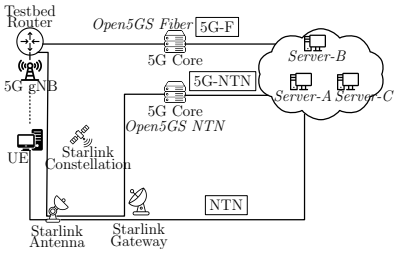

Fig. 1: The three analyzed network configurations: 5G with wired backhaul (5G-F), satellite-only (NTN), and 5G with satellite backhaul (5G-NTN).

B. Methodology

We evaluate network-level latency and path properties using ICMP `ping` and `traceroute`, and service-level QoS/QoE using the ITU-T methodology for real-time interactive applications.

1) End-to-end connectivity and initial latency tests: We use ICMP echo requests to obtain baseline RTT measurements to *Server-A*, *Server-B*, and *Server-C* under the three considered configurations. These measurements provide a simple reference for end-to-end responsiveness, but are not treated as a proxy for application-level latency.

2) Network mapping and per-path latency tests: We use `traceroute` to inspect the paths traversed between the UE and the application servers.

3) Application-specific performance tests: To evaluate latency-sensitive service performance, we adopt the methodology proposed by ETSI and standardized by ITU-T for real-time interactive traffic. We use the standardized *e-Gaming real-time* traffic pattern, with symmetric DL/UL UDP rates between 0.1 and 1 Mbps, to assess bidirectional interactive traffic under strict delay, variation, and loss constraints [13].

To avoid retransmission effects and expose raw network behavior, the ITU-T methodology mandates UDP as transport. On top of UDP, we use the Two-Way Active Measurement Protocol (TWAMP), standardized in IETF RFC 5357 [16]. Once the traffic flow is initialized, service *latency*, *stability*, and *continuity* are evaluated in terms of RTT, Packet Delay Variation (PDV), and Packet Loss Rate (PLR). RTT is computed from packets successfully returned to the UE, while PDV is measured according to IETF RFC 5481 [17]. In addition to packets that are not received, PLR also includes packets exceeding the target delay budget of 100 ms, as recommended by 3GPP for this service class.

To summarize QoE, the methodology defines the *i-score*, which captures client-perceived service responsiveness as a multiplicative combination of the QoS components:

$$\textit{i-score} = score_{\text{RTT}} \times score_{\text{PDV}} \times score_{\text{PLR}} \times 100\% \quad (1)$$

This formulation intentionally penalizes instability: severe degradation in any one component drives the overall score towards zero. We refer the interested reader to [15] for details on the computation of the coefficients in Equation (1).

C. Dataset

Our measurement campaign was conducted in November and December 2024. In total, we collected more than 15,000 ICMP ping samples, 1,500 traceroute samples, and 3,779 interactive sessions across the three servers and configurations.

III. Results

In this section, we analyze the measurements collected across the three network configurations. We begin with end-to-end latency and path-level observations, then examine service-level performance under the standardized *e-Gaming real-time* traffic profile.

A. ICMP `ping`

Figure 2 reports the RTT values obtained with ICMP `ping` toward *Server-A*, *Server-B*, and *Server-C*. These measurements provide a baseline view of end-to-end responsiveness across the three architectures.

For 5G-F, *Server-A* provides the lowest median RTT (13.3 ms), and RTT increases progressively for *Server-B* and *Server-C* as the geographic distance between the UE and the server increases. In contrast, under NTN the lowest median RTT is observed for *Server-B* (30.8 ms), whereas *Server-A* reaches 48.6 ms. This indicates that, in the satellite-only configuration, latency is influenced more by the location of the Starlink exit point than by the physical proximity of the server itself. Under 5G-NTN, *Server-A* again provides the

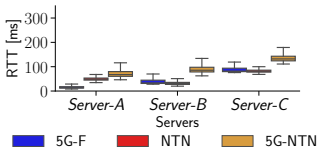

Fig. 2: RTT results obtained using ICMP ping across different servers and network configurations.

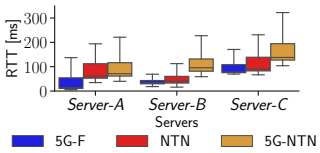

Fig. 3: RTT values collected from all interactive sessions, grouped by server and network configuration.

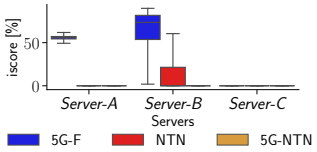

Fig. 4: *i-score* measured for all interactive sessions, grouped by server and network configuration.

lowest median RTT (67.6 ms), followed by *Server-B* (85.8 ms), because the 5G core is located near *Server-A* and traffic is routed through that point. As expected, *Server-C* yields the highest RTT in 5G-NTN (132 ms), reflecting both the satellite segment and the longer terrestrial path.

Overall, once a satellite segment is introduced, latency is shaped by exit-point placement and satellite–terrestrial routing, not server distance alone.

B. ICMP traceroute

We use `traceroute` to inspect the visible UE–server paths. Since parts of the 5G and satellite infrastructure are tunnelled, these results provide only partial path observations. For *Server-A*, 5G-F and 5G-NTN show only 3 visible hops, while NTN shows 15; in 5G-NTN, the satellite infrastructure is hidden inside the GTP tunnel between the gNB and the core. For *Server-B*, NTN has 10 visible hops, compared with 15 for 5G-F and 14 for 5G-NTN, while *Server-C* yields the highest counts: 20 for NTN, 30 for 5G-F, and 29 for 5G-NTN.

While hop count generally increases with server distance, NTN is an exception: *Server-B* requires fewer hops than *Server-A*, consistent with its lower RTT in Section III-A. This supports the interpretation that Starlink routing and exit-point placement can dominate latency behavior. Across all Starlink traceroutes, exposed router names consistently contained `mdrd/es`, suggesting a Madrid-area Internet PoP.

C. e-Gaming real-time

We next evaluate service-level performance using the standardized *e-Gaming real-time* traffic profile. We first compare the RTT observed during the interactive sessions against the ICMP-based RTTs, and then analyze the resulting QoE through the *i-score*.

1) Comparison between ICMP ping and interactivity tests: Figure 3 reports the RTT values measured during the *e-Gaming real-time* sessions. Compared with the ICMP RTTs in Figure 2, interactive traffic generally increases both RTT and variability, especially for NTN and 5G-NTN. For 5G-F, median RTT values remain broadly similar, although variability increases for *Server-A*. For NTN and 5G-NTN, RTTs during interactive sessions are consistently higher and more variable than the ICMP RTTs.

More specifically, the median RTT increases by about 10–13 ms in NTN and by about 3–11 ms in 5G-NTN, depending on the server. This indicates that baseline ICMP measurements underestimate the latency experienced by real interactive traffic. This likely reflects the traffic pattern: unlike sparse ICMP probes, continuous bidirectional UDP traffic is more exposed to queueing, multiplexing, and scheduling effects. In line with prior work [15, 18], our results show that `ping` is useful for basic reachability testing, but not for inferring service latency under interactive traffic.

2) i-score: Figure 4 shows the *i-score* values measured during the *e-Gaming real-time* sessions across servers and configurations. For NTN and 5G-NTN, the *i-score* collapses to 0% in nearly all cases, with a limited number of non-zero values observed for *Server-B* under NTN. Here, PLR represents effective service-level loss, including both missing packets and packets arriving after the 100 ms delay budget.

Since the *i-score* combines RTT, PDV, and PLR, degradation in any component can drive the score toward zero; thus, even moderate RTT can still yield poor service quality when variability or loss is high.

An instructive comparison appears within 5G-F. Although *Server-A* has a lower median RTT than *Server-B*, it exhibits greater variability and therefore worse *i-score* behavior. This confirms that stability, rather than median latency alone, is the dominant factor for interactive service performance.

IV. Conclusion

We presented an empirical study of 5G with LEO satellite backhaul, compared against fiber-backed 5G and standalone satellite access. The results show that exit-point placement, 5G core location, and routing strongly shape latency, and that ICMP RTT is insufficient to characterize interactive service performance. With the standardized e-Gaming profile, both NTN and 5G-NTN exhibit poor performance, with the *i-score* collapsing to 0% in nearly all cases due to delay variation and packet loss. Overall, 5G-NTN is promising for coverage extension, but remains challenging for stringent real-time interaction; exact values remain deployment-specific.

Acknowledgement

This work was partially supported by the Knowledge Foundation DRIVE project, SNS-JU 6G-SANDBOX (No. 101096328), and LearnFDT (PID2022-142181OB-I00).

References

[1] M. M. Azari *et al.*, "Evolution of non-terrestrial networks from 5G to 6G: A survey," *IEEE communications surveys & tutorials*, vol. 24, no. 4, pp. 2633–2672, 2022.

[2] X. Lin *et al.*, "5G from space: An overview of 3GPP non-terrestrial networks," *IEEE Communications Standards Magazine*, vol. 5, no. 4, pp. 147–153, 2021.

[3] Sateliot, "Sateliot achieved "Hello World" on NTN 5G integrating with KSAT and AWS," https://sateliot.space/2024/02/20/sateliot-ksat-aws/, 2024, accessed: April 2026.

[4] I. Del Portillo *et al.*, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta astronautica*, vol. 159, pp. 123–135, 2019.

[5] F. Michel *et al.*, "A first look at starlink performance," in *Proceedings of the 22nd ACM Internet Measurement Conference*, 2022, pp. 130–136.

[6] B. Hu *et al.*, "LEO Satellite vs. Cellular Networks: Exploring the Potential for Synergistic Integration," in *Companion of the 19th International Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT 2023. New York, NY, USA: ACM, 2023, p. 45–51. [Online]. Available: https://doi.org/10.1145/3624354.3630588

[7] M. Rajiullah *et al.*, "CARL-W: a testbed for empirical analyses of 5G and starlink performance," in *Proceedings of the 3rd ACM Workshop on 5G and Beyond Network Measurements, Modeling, and Use Cases*, 2023, pp. 1–7.

[8] N. Mohan *et al.*, "A multifaceted look at starlink performance," in *Proceedings of the ACM Web Conference 2024*, ser. WWW '24. New York, NY, USA: ACM, 2024, p. 2723–2734. [Online]. Available: https://doi.org/10.1145/3589334.3645328

[9] A. E. Ferguson *et al.*, "Starlink performance from different perspectives," in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, ser. ACM MobiCom '24. New York, NY, USA: ACM, 2024, p. 1662–1664. [Online]. Available: https://doi.org/10.1145/3636534.3697452

[10] S. Ma *et al.*, "Network Characteristics of LEO Satellite Constellations: A Starlink-Based Measurement from End Users," *arXiv preprint arXiv:2212.13697*, 2022.

[11] S. Heine *et al.*, "Concept and evaluation of 5G backhauling via starlink," in *39th International Communications Satellite Systems Conference (ICSSC 2022)*, vol. 2022. IET, 2022, pp. 69–75.

[12] ETSI, "Speech and Multimedia Transmission Quality (STQ); QoS Parameters and Test Scenarios for Assessing Network Capabilities in 5G Performance Measurements," ETSI, Tech. Rep. TR 103 702 V1.1.1, November 2020.

[13] ITU-T, "Latency Measurement and Interactivity Scoring Under Real Application Data Traffic Patterns," International Telecommunication Union, Tech. Rep. ITU-T G.1051, March 2023.

[14] G. Caso *et al.*, "The Chronicles of 5G Non-Standalone: An Empirical Analysis of Performance and Service Evolution," *IEEE Open Journal of the Communications Society*, 2024.

[15] ——, "A Standard-compliant Assessment of Beyond-eMBB QoS/QoE in 5G Networks," in *2024 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2024, pp. 230–236.

[16] K. Hedayat *et al.*, "A Two-Way Active Measurement Protocol (TWAMP)," IETF, Tech. Rep. RFC 5357, 2008.

[17] IETF, "Packet Delay Variation Applicability Statement," IETF, Tech. Rep. RFC 5481, 2009.

[18] C. Pelsser *et al.*, "From Paris to Tokyo: On the suitability of ping to measure latency," in *Proc. of the ACM Internet Measurement Conference*, 2013, pp. 427–432.