Measuring Service Continuity in Integrated TN/NTN for 5G-Advanced and 6G

Fanny Parzysz
Orange Innovation, Lannion, FR
email: fanny.parzysz@orange.com

Alessandro Guidotti
CNIT, Bologna, IT
email: a.guidotti@unibo.it

Alessandro Vanelli-Coralli
University of Bologna, Bologna, IT
email: alessandro.vanelli@unibo.it

***Abstract*—Service continuity has been defined in 3GPP TS 22.261 as the "uninterrupted user experience of a service that is using an active communication when a UE undergoes an access change without, as far as possible, the user noticing the change". Such definition remains open, is not associated with KPIs and thus, cannot be used to compare the many architecture and business options that can be envisaged to deploy and operate 3D networks. In this paper, we propose a generic three-phased approach, with KPIs, to quantify the seamless performance of TN / NTN switching. We also discuss the purpose, applicability and order of magnitude of each KPI, based on concrete examples.**

***Index Terms*—5G/6G-NTN, Service continuity, KPI**

I. Introduction

It is generally understood that the terrestrial network (TN) alone cannot provide the flexibility, scalability, adaptability, and coverage, required to meet the increasing need of our society for ubiquitous and continuous connectivity services. The integration of non-terrestrial networks (NTN) is thus a key enabler in the evolution of 5G-Advanced and 6G, [1]–[4].

The targeted unified 3D network is illustrated in Fig. 1 and can be defined as a network of networks, composed of several *layers*: terrestrial nodes, air-borne flying nodes (encompassing both high altitude platforms and aerial base stations), and space-borne nodes, all acting as network access points for users. Space-borne nodes can be vLEO, LEO, MEO or GEO satellites (respectively Very Low, Low, Medium, and Geostationary Earth Orbit). Each of these layers offers different performance for users - in terms of bandwidth, latency, reliability, or edge computing capabilities - and is subject to different constraints - in particular computational capabilities on board satellites, maximum connection density, feeder link capacity, and/or (de)centralized resource management.

While NTN has been widely considered for backhauling and coverage extension, the next challenge of a unified 3D network is to offer true service continuity, including for mobility scenarios. Users may desire to enjoy a continuous connectivity, with no service interruption nor perceivable data loss, whatever the underlying network component, NTN, TN or their combination. In other words, terminals should be able to seamlessly and transparently switch from TN to NTN, from NTN to TN, or from NTN to NTN. This requirement particularly applies to critical services such as remote driving or drones identification and tracking in remote areas, etc.

5G-STARDUST project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101096573. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

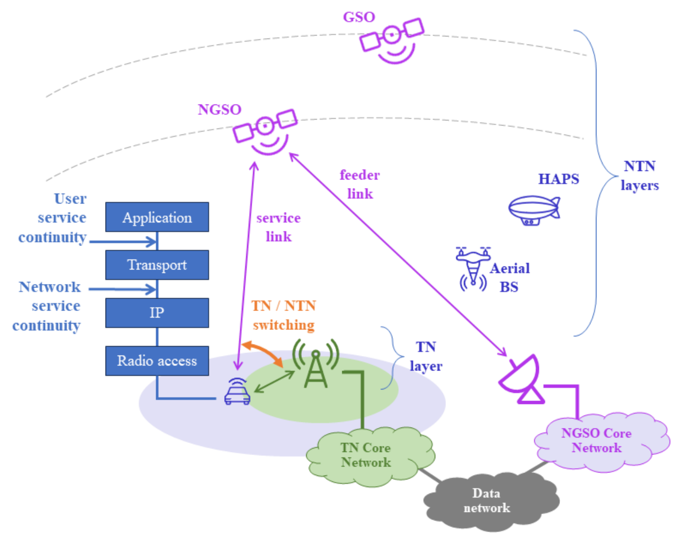

Fig. 1. Example of a 3D network: Architecture option with two core networks

A wide range of architecture options and business organizations are envisaged to operate this highly heterogeneous network of networks [5]. For example, the NTN component defined in terms of Earth-fixed or Earth-moving beams, transparent or regenerative payloads, with the latter allowing various functional split options and edge computing solutions, offers manyfold opportunities, that deeply impact mobility management. Indeed, the TN/NTN layers could work independently one from each other, or be interconnected, for more efficient resource management, enhanced mobility support and simplified business commitment for the users. However, a "one fits all use cases" design is unrealistic and novel Key Performance Indicators (KPIs) need to be designed, to compare potential options and identify the best trade-off.

Problem formulation: Service continuity has been defined in 3GPP TS 22.261 as the "uninterrupted user experience of a service that is using an active communication when a

UE undergoes an access change without, as far as possible, the user noticing the change," [6]. Because such definition is not linked to quantifiable KPIs, it does not allow to establish performance targets or Service Level Agreements (SLA), nor says if a given architecture option can meet user requirements and be suitable for a given use case in a given geographical area. More details are also required to specify what is an "access change", and to which extent a user experience can be considered as "uninterrupted". As far of our knowledge, no framework has been described so far to characterize and quantify service continuity in integrated TN/NTN systems.

Proposed solution: To this end, we propose a generic methodology which extends the approach developed for cross-border car mobility in [7]. It consists of three phases, with three sets of KPIs. Orders of magnitude are discussed for different examples, and insights are provided about how to measure and optimize these KPIs, for better service continuity. Without loss of generality, this work primarily focuses on TN/NTN switching. The user device is assumed to support both TN and NTN frequency bands.

Paper organization: The main architecture and business options for 3D networks are presented in Section II and discussed with the perspective of service continuity. The three-phased methodology is presented in Section III and detailed in Sections IV, V and VI. Section VII concludes this paper.

II. Service Continuity in Main Architecture and Business Options for 3D Networks

Today, also building on the success of 3GPP Rel. 17, switching to NTN in areas where 5G/6G connectivity experiences gaps is often taken as granted. Yet, there exist many options (summarized in Fig. 2) to deploy and operate 3D networks, and not all of them can guarantee service continuity.

A. Non-cooperating TN / NTN layers

In its simplest architecture option, a 3D network can be composed of independent and non-interconnected layers. In this case, the user can switch between TN and NTN through indirect access, via Customer Premise Equipment (CPE) or Very Small Aperture Terminals (VSAT), or through direct access, if provided with two radio access front-ends and related credentials. NTN may or may not be 3GPP-compliant. This also includes the various implementations for Dual SIM (Dual Active, Dual Standby) and Single SIM-Multi-IMSI [8].

Layer switching decision is generally made by a multi-IMSI applet or a dedicated algorithm at application or transport level. It is triggered by the user (or CPE), *e.g.,* because it detects being out of coverage of its primary network layer. More advanced mechanisms can leverage link failure or signal degradation prediction, to steer traffic to the best Radio Access Technology (RAT) (see Example 1 in Section III-C).

These solutions have already proven fair service continuity for specific use cases (for example, military drones) and can be implemented directly by device manufacturers, regardless of network operators. Yet, they usually focus on scenarios with overlapping network coverage and good service continuity

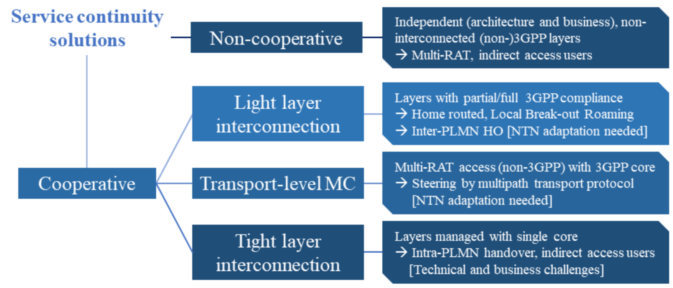

Fig. 2. Summary of the considered service continuity options.

depends on the decision-making algorithm, potentially AI-enhanced. The lack of network information (e.g. traffic load, available resource) can lead to poor switching decision and ping pong effects could occur if links are unstable.

B. Transport-level multi-connectivity support

Different levels of Multi-Connectivity (MC) have been standardised for TNs, but not yet for NTNs; these include Multi-TRP at PHY/MAC layer, [9] , Carrier Aggregation (CA) at MAC layer, [10], and also solutions at Packet Data Convergence Protocol (PDCP) layer, [11]. MC can also be achieved at core level, via ATSSS (Access Traffic Steering, Switching and Splitting), which opens to the possibility of connecting a UE to a 5G Core network (5GC) via different RATs, [12], *e.g.*, 3GPP TN and non-3GPP NTN. Several steering modes have been defined since 3GPP Rel. 16, allowing a wide range of link management and traffic splitting options, *e.g.*, Load Balancing, Smallest delay, or Priority-based.

Although ATSSS enables the simultaneous use of both TN and NTN radio accesses, service continuity remains to be validated, in particular in case of link failure or Quality of Service (QoS) degradation over one of these links. Indeed, access change would be here performed at the UE and/or the User Plane Function (UPF), by the transport protocol (generally Multipath-TCP or Multipath-QUIC). It is subject to their congestion control and failure detection mechanisms, which can be severely impacted by long delays, especially if the NTN gateway is not close to the 5GC. In addition, core-based MC might be challenging when needing fast 5GC adaptations to the dynamic radio link conditions.

C. Light interconnection with several core networks

This category encompasses architecture options where the various network layers are at least partially 3GPP-compliant, enabling roaming and inter-PLMN handovers (HO).

First, the support of a "minimum" 5GC, able to manage subscribers and provide basic packet data service, allows a NTN operator (visited PLMN, V-PLMN) to serve UEs that have subscriptions with a TN operator (home PLMN, H-PLMN), or vice versa. It requires only light network interconnection and a roaming agreement between the two operators (home-routed or local-break out). In this case, TN/NTN switching corresponds to a PLMN reselection procedure. From a practical perspective, it is triggered by the UE, with no fine-grained management of the network, as illustrated in Example 2 of Section III-C. Note that this option might be a short-/mid-term solution, but a partial NTN-5GC could prevent the support of some devices, *e.g.*, IoT using DRX or power saving modes.

Second, the tight user requirements of the automotive sector has spurred the improvement of inter-PLMN HO for cross-border mobility [14]. This solution prevents UEs to remain connected to a H-PLMN gNB till connection is lost due to too-weak signal, and allows to switch to a V-PLMN gNB beforehand, nearly as for an intra-PLMN HO. This implies that gNBs in a country are configured with information (Cell IDs, frequency bands, etc.) about gNBs of other countries, managed by other PLMNs. To this end, the N14 interface is used to connect the H-PLMN's AMF to the V-PLMN's AMF. Such scenario has particularly raised the challenges of anchor UPF selection, choice of Session and Service Continuity (SSC) mode, and management of PDU sessions and IP addresses.

Although it offers enhanced service continuity, inter-PLMN HO remains a technical challenge and a matter of trust, security and legal obligations, even for TN / TN switching. In addition to QoS heterogeneity and extended space-ground delays, some practical constraints may prevent implementing this solution for TN / NTN switching, in particular the actual localization of the NTN-UPF, NTN-AMF and NTN gateway (same country as TN-5GC or not, on the ground or in space).

D. Tight interconnection with a single core network

Finally, architecture options for 3D networks can be based on a single core network (e.g. the TN-5GC), as explored in the 5G-STARDUST project [5]. For direct-to-cell, the TN / NTN switching of a UE is an intra-PLMN HO and optimal service continuity can be offered. For indirect access, Integrated Access and Backhaul (IAB) can be envisaged, as in Example 3 of Section III-C. Nevertheless, both options raise a number of technical and business challenges.

From a purely theoretical perspective, a single stakeholder would operate the whole 3D network and its different components. It would have full knowledge and full control of network configurations, satellite beam management, radio access capabilities, traffic load balancing, etc. Hence, switching decisions could be optimized, with limited HO service interruption.

But from a more realistic point of view, a Satellite Network Operator (SNO) would operate the constellation while several Mobile Network Operators (MNO) or Mobile Virtual Network Operators (MVNO) would use the SNO's capabilities to operate 5G/6G-NTN services (Radio-as-a-Service). Several business organizations can be envisaged: SNO acting as Towerco with passive / active assets, as Neutral Host, as Infraco with RAN sharing agreement, etc. The choice of the business organization will have deep impact on:

- the type of network resource that is shared between the different stakeholders, e.g. a static bandwidth reservation vs. a dynamic cognitive resource allocation,
- the level of control of the MNO on the NTN features,
- the interfaces or interconnections between the SNO and MNO, for network information exchange and billing,

so that TN / NTN switching decisions, and thus service continuity, are made with at least a partial knowledge and control of network capabilities.

III. Overview of the proposed methodology

We now describe the generic three-phased approach proposed to quantify service continuity for the various architecture options presented above. It is illustrated with several examples.

A. Considered terminology

The term "Handover" can be somehow misleading to designate TN / NTN switching, as it may hint at some particular technical procedures (e.g. 3GPP X2 HO) or specific business organization (e.g. RAN sharing). Therefore, we rather consider the following terminology.

Network Layer Switching (NLS): In the investigated 3D network context, NLS refers to a change of radio access for an *active* user, i.e. user in communications, from one network layer to another one, whatever the cause of this change (user mobility, planned or unplanned event, network-triggered load-balancing decision, etc.). This umbrella term encompasses a wide range of 3D network architectures (transparent / regenerative payload, single- / multi-core, etc.), switching procedures (e.g. network reselection, handovers, SIM card switching, etc.) and business relationships (Roaming, RAN sharing etc.).

Service continuity: It is both a matter of service interruption and QoS variability. As illustrated in Fig. 1, it is important to distinguish between what is under the control of networks and what is not, i.e. to distinguish between 1) the *network service continuity*, which considers management of network attachment parameters (UE registration state, bearers, PDU sessions, IP sessions, etc.), and 2) the *user service continuity*, which reflects the ability of a user application to absorb part of underlying network service failures and QoS fluctuations, due to the heterogeneity of link capabilities in 3D networks.

B. A three-phased approach to evaluate Service Continuity

In practice, KPIs to assess service continuity of NLS should capture the effects of related network events. To this end, we extend the approach of [7] and propose to isolate the three phases of such "access change", that is NLS preparation (Phase 1), NLS execution (Phase 2) and NLS evaluation (Phase 3). We define the following time line, illustrated in Fig. 3:

- t_0: Start of measurements (passive monitoring), to evaluate pre-NLS performance.
- t_1: Event E, which implies to switch N Users from one network layer to another one.
- t_2: Start of NLS for a first user (e.g. HO procedure), i.e.
  $$t_2 = \min_{i \in [1,N]} (t_{2,i})$$
  with $t_{2,i}$: Start of the NLS procedure of User $i \in [1, N]$.
- t_3: End of the NLS of the last user and back to passive monitoring, to evaluate post-NLS performance. That is
  $$t_3 = \max_{i \in [1,N]} (t_{3,i})$$
  with $t_{3,i}$: End of the NLS procedure of User $i \in [1, N]$.
- t_4: End of measurements.

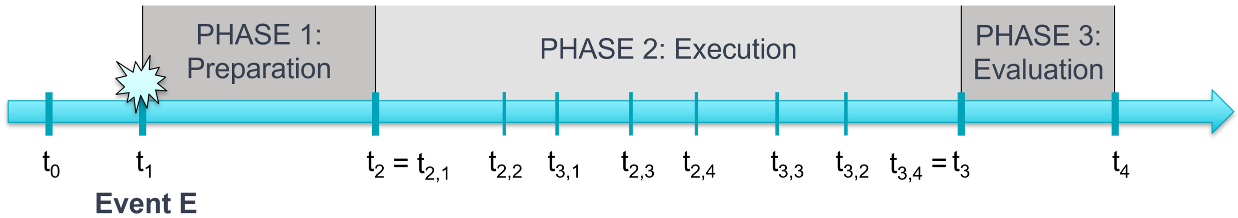

Fig. 3. Timeline for Network Layer Switching (NLS), with N = 4 Users as an example

Event E may be initiated by the serving PLMN or detected by the user. It may be unplanned, predictable, or fully deterministic, and may be on-demand or based on long-term forecast. For example, it corresponds to a ship coming from the open seas (with NTN coverage only) and approaching a port (with TN coverage). It could also be a load balancing decision between LEO and GEO, to optimize resource utilization.

Compared to [7] where switching is performed between peers, the proposed framework is designed and discussed with respect to NTN specificities. It accounts in particular for multi-user scenarios and national vs. international coverage. Insights on how to measure KPIs are provided for several contexts (simulations, Proof-of-Concept, commercial operations).

C. *Examples of Network Layer Switching*

Based on the Use Cases proposed within the 5G-Stardust project [15], we propose the following examples, which will be used throughout Sections IV, V and VI.

Example 1 - Automotive Use Case with Multi-SIM: A connected car needs seamless TN / NTN connectivity to benefit from safety and cloud services, such as NG eCalls, dynamic HD maps for traffic efficiency, whatever its localization. To this end, it is provided with a Dual SIM (or equivalently a Single SIM - Multi-IMSI) terminal, with an application-level algorithm that steers the traffic to the best radio access. Both TN / NTN radio accesses are active at the same time and NLS leverages prediction of link failure or signal degradation. For this example, N = 1 and Event E = the car reaches TN coverage edge and a signal strength threshold is met.

Example 2 - PPDR Use Case with Roaming agreement: A natural disaster occurs and part of TN infrastructure is damaged. All UEs in the area are disconnected at the same time, but the MNO has a Home-routed roaming agreement with a SNO for 5G/6G-NTN services. Here, Event E = Natural disaster, N = several thousand UEs.

Example 3 - Moving platforms with Integrated Access and Backhaul (IAB): A moving platform (aircraft, ship, train) is equipped with an IAB-node that connects to the 5GC via the TN or NTN component to serve the on-board terminals. For long-range international transports, network services may rather be deployed on the move, as a function of the country. Terminals (N = up to thousands on a cruise ship) all move together and may have different requirements (*e.g.*, sensor data for cargo ships/trains or broadband for passengers). In this scenario, Event E corresponds a network-triggered switching decision when the platform approaches TN coverage.

IV. PHASE 1 - PREPARATION OF NETWORK LAYER SWITCHING AND ASSOCIATED KPIS

This phase corresponds to the time elapsed between t_1 (event E, triggering the need for some UEs to move from one network layer to another one) and t_2 (start of NLS procedures).

Phase 1 - KPIs: Three KPIs related to the NLS preparation phase are proposed to assess Service Continuity:

- KPI_0: Service state of UEs until successful NLS.
- KPI_1: Time duration of the preparation phase, i.e. $t_2 - t_1$.
- KPI_2: Number of expected NLS attempts.

A. *KPI_0 - Service state of UEs after Event E*

Depending on the considered architecture option, business organization and type of Event E, the UEs concerned by NLS may remain registered and RRC connected / RRC inactive after Event E till successful NLS procedure (i.e. $\text{KPI}_0 = 1$, as in Examples 1 and 3) or undergo service interruption, i.e. are deregistered / RRC Idle ($\text{KIP}_0 = 0$, as in Example 2). Typically, service interruption can be due to a sudden unplanned event E (e.g. natural disaster), to NTN link unavailability (e.g. cloudy day, constellation build-up phase, car driving in a tunnel) or to insufficient network capacity (e.g. exceptional sport events). In an experimental setup, this KPI is monitored at the UE, using tools like Keysight Nemo Handy or ADB logs (not available for all chipsets though).

B. *KPI_1 - Time duration of Phase 1*

The time elapsed between t_1 (event E) and t_2 (start of NLS procedure of the first user) mainly depends on the time needed for event detection and for UE synchronization with the new network layer, but also on who triggers the NLS procedure.

In Example 1 (Multi-SIM), both radio accesses are simultaneously active and KPI_1 corresponds the sampling period of the signal strength indicator used for switching decision. Depending on the chipset configuration, sampling is done a few times per seconds (ex: ADB logs). Longer refresh period is expected if data from external sources is required. Prediction of event E could however significantly reduce KPI_1 and avoid any potential service interruption due to late detection. For example, the "Network Data Service" has been defined for drones to allow the sharing of Network KPI monitoring and coverage information [16]. This would allow fast NLS for critical real-time use cases, such as remote piloting and identification & tracking.

In Example 2 (Roaming), KPI_1 includes the time necessary for the UE to switch to the RM-deristered state, to perform a full network search and to synchronize with the Visited PLMN (using PSS / SSS). Depending on the UE configuration, it may take seconds to minutes. In this scenario, once TN is operational again, active UEs will remain attached to NTN, which is neither spectrum- nor cost-efficient. Furthermore, idle UEs may not detect TN immediately, as network scan is performed only every X*6 minutes (Expiration of 3GPP timers). This extra-time adds to KPI_1. In this example, reducing KPI_1 (and thus, service interruption) mainly implies adapting timers in 3GPP standards. The gain brought by improved RF chains and UE configuration, for accelerated scan and synchronization, would probably remain quite marginal.

In Example 3 (IAB), KPI_1 principally captures the network service deployment time (from the initiation of the service deployment until it reaches full operational capacity). Note that even if UEs remain connected in this case, KPI_1 should be preferably reduced in case of fast-moving platforms (on aircraft) and low altitude NTN layers (e.g. vLEO). Indeed, NLS has to align with regular satellite handovers and an inadequate functional split may cause significant signaling overhead or even, a non-converging NLS preparation.

Measuring time-related metrics (KPI_1 and KPI_3) can be easily envisaged in simulations (Matlab, NS3, OMNET++, etc.) but is much more challenging for experimental setups. Monitoring probes should be located at the same layer (Radio, MAC, IP, Transport, etc.) at source and destination, and synchronized at an acceptable accuracy level. Given the considered TN / NTN switching context, targeting a synchronization accuracy value of 100μs-1ms, as in [13], should be sufficient. Different synchronization methods can be considered but remain to be validated for the various TN / NTN scenarios, e.g. NTP / PTP (Network / Precision Time Protocol) or GNSS time references, which can be used for LEOs as well.

For KPI_1, logging the Control Plane is necessary to spot the start of the NLS procedure, for example a RRC Connection Request sent by the UE in a roaming scenario or a HO request sent by the Source gNB to the Target gNB (potentially via the AMF). Tools like Open RAN Studio can be envisaged.

C. KPI_2 - Number of expected NLS attempts

For simulation-based analysis and experimental setups, KPI_2 should be understood as an intermediate indicator and will be used in Phase 2, as explained in Section V-B. However, in operational or commercial networks, this metric can hardly be monitored, especially in case of user-triggered NLS. Market trends and business analysis can be used to assess it, for example to conclude a roaming agreement. A correct estimation of KPI_2 is essential for adequate network dimensioning. Key factors include the UE density and distribution, their activity factor and the TN / NTN service coverage areas.

V. PHASE 2 - EXECUTION OF NETWORK LAYER SWITCHING AND ASSOCIATED KPIS

This phase corresponds to the time elapsed between t_2 (start of NLS procedure of the first user) and t_3 (end of NLS procedure of the last user). The objective is here to measure the success of NLS from the network perspective, i.e. to evaluate how successfully NLS procedures have been executed.

Phase 2 - KPIs: We consider the following KPIs:

- $\mathrm{KPI}_3 = t_3 - t_2$: Time duration of the execution phase. We also define $\mathrm{KPI}_{3,i} = t_{3,i} - t_{2,i}$, as the time duration of the execution phase for User i.
- KPI_4: NLS attempts rate,
- KPI_5: NLS success rate.

Note that these KPIs are inspired from the ones used in practice to monitor 4G / 5G commercial networks.

A. KPI_3 - Time duration of Phase 2

A long NLS execution phase can severely impair service continuity even if UEs remain connected ($\mathrm{KPI}_0 = 1$). In fact, it may cause buffer overflows and packet loss.

First, $\mathrm{KPI}_{3,i} = t_{3,i} - t_{2,i}$ refers to the user perspective. It may range from tens of milliseconds (typical TN intra-PLMN HO), to a few seconds / minutes as in Example 2, for a full initial access procedure by the Visited PLMN, with registration, authentication, PDU session establishment, etc. In another example, the inter-PLMN handover performance has been evaluated during the European projects 5GCroCo and 5G-MOBIX [14], showing an almost imperceptible service interruption time between 120 ms to 245 ms for a TN-TN switching. Much longer delays are however expected for TN-NTN switching, depending on the type of satellite payload (transparent / regenerative with full or partial gNB) and on the number of satellite hops to reach the feeder link.

Next, KPI_3 corresponds to the network perspective, with $N > 1$ users. Several formulas could have been considered for this indicator, in particular the average value $\mathbb{E}\left[\mathrm{KPI}_{3,i}\right]$ or the maximum value $\max_i \left\{\mathrm{KPI}_{3,i}\right\}$, which are of course both valid to characterize service continuity of individual users in the considered framework. However, neither capture the fact that the system can handle several NLS procedures at the same time, and that, on the contrary, there may be idle time between successive NLS procedures. In addition, a UE may undergo several NLS procedures in case of failure of a first attempt.

As Phase 2 is a transition period, we propose

$$\mathrm{KPI}_3 = \max_{i \in [1,N]} (t_{3,i}) - \min_{j \in [1,N]} (t_{2,j}).$$

This metric allows that performance evaluation of Phase 3 is done in a final state, i.e. when all targeted UEs have been switched to the new network layer and share its resources. This approach particularly fits large-scale scenarios (as in Example 2, with $N =$several thousands) and fast-moving use cases (in Example 3, flying 3min at 900km/h is more or less equivalent to crossing a whole LEO satellite beam).

B. KPI_4: NLS attempts rate

The NLS attempts rate is defined as follows:

$$\mathrm{KPI}_4 = \frac{\text{\# of Users effectively starting a NLS procedure}}{\text{\# of Users expected to require NLS (i.e. } \mathrm{KPI}_2)}.$$

The target value is obviously 100%, meaning all UEs expected to switch effectively start a NLS procedure. $\mathrm{KPI}_4 <$

100% is linked to the various issues preventing UEs from initiating a NLS procedure, for example, a failure in the detection of Event E, the unavailability of the NTN link (clouds, indoors, constellation still in its deployment phase and thus, incomplete), the inability of the UE to synchronize, an error with the SIM profile (in Example 1, the NTN PLMN ID could be incorrectly listed in the prioritized networks), etc. In case $\text{KPI}_4 > 100\%$, the number of UEs expected to require NLS has been underestimated, potentially resulting in a poor network dimensioning.

C. KPI_5: NLS success rate

The NLS success rate is defined as follows:

$$\text{KPI}_5 = \frac{\#\text{ of Users successfully switched}}{\#\text{ of Users starting a NLS procedure}}$$

$\text{KPI}_5 < 100\%$ reflects the various issues preventing successful NLS procedures. Technical causes can be, for example, long signaling delays for suboptimal functional splits in regenerative payload, or an invalid subscriber identity in the target network layer. Business aspects can also lead to unsuccessful switching, e.g. a roaming failure due to TAU (Tracking Area Update) reject. Indeed, managing PLMN IDs and spectrum remains a real challenge in integrated TN / NTN networks, as TN typically have national footprint while NTN are worldwide, with different rules and legal obligations.

Measuring KPI_4 and KPI_5: When there are only a few users, advanced monitoring and Control Plane logging can be performed using the ID of each UE (e.g. phone number). However, this method can rapidly become intractable when N gets large. In this case, using OSS counters at gNBs may be a better option. Such method allows to log network data on a per-gNB basis, i.e. aggregated over served UEs. No information related to individual users (identity, localization, mobility or data consumption habits) is collected by these counters and data is usually averaged per quarter-hour, at the finest granularity. This can however provide a good basis for resource management and constellation design, as proposed in the 5G-STARDUST project [17].

VI. Phase 3 - Evaluation of network layer switching and associated KPIs

This third phase corresponds to the time elapsed between t_3 and t_4 (End of measurements) and aims to assess how successful was NLS from the user perspective. Related KPIs should reflect potential performance evolution, compared to the initial situation (before event E).

Phase 3 - KPIs: We consider the following KPIs:

- KPI_6: User rate evolution, defined as the mean user experienced data rate measured after t_3 divided by the one measured before Event E (t_1). This KPI can be computed for a single user or averaged over all users.
- KPI_7: User latency evolution, i.e. the ratio of the experienced before t_1 and after t_3.
- KPI_8: User application failure ratio.

More KPIs can be identified depending on the use case. While KPI_6 and KPI_7 are quite obvious, we detail KPI_8.

A. KPI_8: User application failure ratio

Different from KPI_5 (NLS success rate), it is defined as the percentage of users who experienced application service failure despite a successful NLS procedure:

$$\text{KPI}_8 = \frac{\#\text{ of Users with application failure} \mid \text{ sucessful NLS}}{\#\text{ of Users with successful NLS}}.$$

Main causes of user service failure can be highlighted here. First, the rate (resp. latency) experienced after switching is too low (resp. high). For example, the ITU-T G.114 recommends that the end-to-end latency should not exceed 150 milliseconds for good voice quality. In Automotive, most services for vehicle platooning require a maximum latency of some tens of milliseconds [7]. Second, the NLS preparation and execution phases may imply service interruption (KPI_0). For a User i, if the interruption duration $t_{3,i} - t_1$ exceeds the application survival time, service undergoes failure. Another reason can be that too many packets have been lost during NLS.

VII. Conclusion

We proposed a generic method and KPIs to evaluate service continuity in integrated TN / NTN for 5G-Advanced and 6G. It is agnostic on the architecture and business options selected to deploy and operate this 3D network, and is decomposed into 3 phases - NLS preparation, execution and evaluation -, which allow to spot issues leading to poor service continuity. Orders of magnitude have been detailed through several examples.

References

[1] A. Vanelli-Coralli et al., "5G Non-Terrestrial Networks: Technologies, Standards, and System Design," Wiley-IEEE Press, 1st ed., 2024.

[2] W. Jiang et al., "The road towards 6g: A comprehensive survey," IEEE Open Journal of the Communications Society, vol. 2, pp. 334–366, 2021.

[3] M. Giordani et al., "To- ward 6g networks: Use cases and technologies," IEEE Communications Magazine, vol. 58, no. 3, pp. 55–61, 2020.

[4] A. Guidotti et al., "The path to 5G-Advanced and 6G Non-Terrestrial Network systems," 2022 11th Advanced Satellite Multimedia Systems Conference and the 17th Signal Processing for Space Communications Workshop (ASMS/SPSC), Graz, Austria, 2022.

[5] SNS JU Project 5G-STARDUST, Deliverable D3.1, "System Requirements Analysis and Specifications," July 2023.

[6] 3GPP TS 22.261, "Service requirements for the 5G system; Stage 1 (Release 19)," September 2023.

[7] EU H2020 Project 5G-MOBIX, Deliverable D2.5, "Initial evaluation KPIs and metrics," October 2019.

[8] O. Vikhrova et al., "Multi-SIM support in 5G Evolution: Challenges and Opportunities," IEEE Communications Standards Magazine, vol. 6, no. 2, pp. 64-70, 2022.

[9] 3GPP TS 38.300, "NR; NR and NG-RAN Overall Description; Stage 2 (Release 16)," September 2020.

[10] 3GPP TS 38.331, "NR; Radio Resource Control (RRC) protocol specification (Release 17)," September 2023.

[11] 3GPP TS 37.340, "NR; Multi-connectivity; Overall description; Stage-2," January 2024.

[12] 3GPP TS 22.261, "5G System; Access Traffic Steering, Switching and Splitting (ATSSS); Stage 3," January 2024.

[13] EU H2020 Project 5G-MOBIX, Deliverable D5.1, "Evaluation methodology and plans," February 2020.

[14] White Paper, 5G-MOBIX, 5G-CARMEN, 5GCroCo, "5G technologies for connected automated mobility in cross-border contexts," May 2023.

[15] SNS JU Project 5G-STARDUST, Deliverable D2.1, "Scenarios, use cases, and services," July 2023.

[16] ACJA Work Task 2, "Interface for Data Exchange between MNOs and the UAS Ecosystem," January 2023.

[17] SNS JU Project 5G-STARDUST, Deliverable D4.1, "Open data sets for ML-based RRM," January 2024.