Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

5G Non-Standalone (NSA) Network Slicing: Trial Evaluation and Performance Analysis

Abdul Quader Syed, Syed Naveed Maqdoom, Hasan Omair Mohammed, Mohammed Yahiya Pasha Gulam
Nokia Solution & Networks

Abstract - The coexistence of heterogeneous device classes—latency-sensitive smartphones and throughput-intensive fixed-wireless routers—within a shared 5G New Radio (NR) carrier introduces a resource-contention problem that uncontrolled scheduling cannot resolve fairly. This paper presents a live-network trial of Subscriber Profile Identifier (SPID)-based Network Slicing in 5G Non-Standalone (NSA) mode, in which Physical Resource Block (PRB) quotas are enforced per Resource Group (RG) over FR1 bands. Two Resource Groups were defined—RG#1 for smartphone user equipment (UE) and RG#2 for router/CPE UE—and three cell-load-adaptive quota profiles were configured. The trial was conducted across two urban deployment clusters using a comparative PRE/POST methodology spanning both network-side counters (N-KPI) and user-perceived speed-measurement results (S-KPI). Results show downlink (DL) throughput improvements of 2.5% to 5.2%, uplink (UL) throughput gains of up to 10%, and a measurable rise in average NSA user count, while the gNB scheduler was confirmed to correctly prioritize smartphone UE through higher PRB allocation and lower scheduling delay. Per-band analysis demonstrated that the joint deployment of primary/secondary cell load balancing and NSA slicing redistributes router UE away from congested mid-band layers, relieving Physical Downlink Control Channel (PDCCH) starvation. The study concludes that NSA slicing is an effective, deployable QoS-differentiation mechanism, and that minimum/maximum quota tuning is as decisive as feature activation in realizing its benefits.

Index Terms - 5G NR, Non-Standalone (NSA), network slicing, radio resource management, PRB allocation, QoS differentiation, slice-aware scheduler, resource group, SPID-based slicing, primary/secondary cell load balancing, UE-type differentiation, KPI analysis.

I. INTRODUCTION

The transition to fifth-generation (5G) mobile networks has been driven largely through the Non-Standalone (NSA) architecture, in which a 5G New Radio (NR) carrier is anchored to an existing Long-Term Evolution (LTE) control plane. While NSA accelerates 5G capacity deployment, it inherits an important constraint: end-to-end network slicing defined by the 5G core is not available, and slice-like service differentiation must therefore be realized entirely within the Radio Access Network (RAN).

A practical consequence of this constraint emerges when device classes with fundamentally different traffic characteristics share the same carrier. Smartphones generate bursty, latency-sensitive traffic, whereas fixed-wireless routers and customer-premises equipment (CPE) generate sustained, throughput-intensive flows. When both classes are scheduled from a common resource pool without differentiation, router traffic can dominate Physical Resource Block (PRB) occupancy, degrading the quality of experience for smartphone users.

This paper addresses that problem through *SPID-based NSA slicing*, a RAN-only mechanism that classifies UE by Subscriber Profile Identifier (SPID), assigns each class to a Resource Group (RG), and enforces minimum and maximum PRB quotas per RG. The mechanism was trialed on a live commercial network across two urban deployment clusters over FR1 bands.

The contributions of this work are fourfold. First, we present a practical implementation and parameter design for SPID-based NSA slicing over FR1 multi-band deployments. Second, we provide a quantitative PRE/POST comparison spanning both network-side key performance indicators (N-KPI) and user-perceived service KPI (S-KPI). Third, we present a cross-layer analysis of per-slice UE behavior, including throughput, scheduling delay, and PRB utilization. Fourth, we derive operational recommendations for minimum/maximum quota tuning and outline a roadmap of future slicing enhancements.

II. BACKGROUND AND RELATED WORK

Network slicing in 5G is specified by the 3rd Generation Partnership Project (3GPP) primarily for the Standalone (SA) architecture, where a Single Network Slice Selection Assistance Information (S-NSSAI) identifier is propagated end-to-end through the core network and RAN. In NSA deployments this propagation is unavailable, so slice differentiation is constrained to RAN scheduling decisions.

To realize differentiation within the RAN, the SPID, an operator-assigned subscriber attribute already used for idle-mode frequency steering, is repurposed as a UE-class label. UEs sharing a SPID are grouped into a Resource Group, and the scheduler applies slice-aware weighting and PRB quota enforcement to each group. This approach trades the rich, end-to-end isolation of SA slicing for an immediately deployable mechanism compatible with the installed NSA base.

Prior studies of 5G RAN slicing have largely concentrated on SA architectures and on simulation-based evaluation of PRB partitioning and slice-aware scheduling. Comparatively little has been published on live-network trial results for NSA-mode SPID-based slicing across multiple FR1 bands, particularly under realistic mixed smartphone/router traffic. This paper helps to close that gap by reporting measured outcomes from a commercial deployment.

III. NSA NETWORK SLICING ARCHITECTURE AND RESOURCE MANAGEMENT FRAMEWORK

A. System Architecture

Because NSA precludes core-driven slicing, the architecture is entirely gNB-centric. UE classification is performed via SPID: a smartphone profile and a router profile are defined and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

mapped to SPID groups through the SPID-group and SPID-profile managed objects. Each SPID group is bound to a Resource Group: RG#1 carries smartphone UE and RG#2 carries router/CPE UE.

Quota assignment is governed by a Resource-Group profile object, which associates each RG with its minimum and maximum PRB quota and is selected according to the prevailing cell-load condition. This separation of UE classification (SPID), grouping (RG), and quota policy (RG profile) allows the same physical carrier to apply different differentiation policies as load varies.

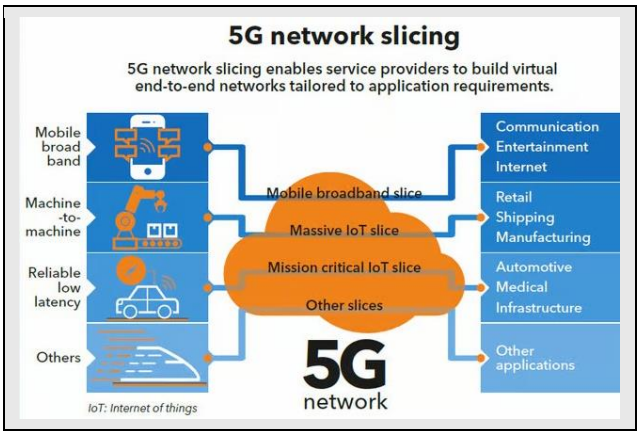

Fig. 1. NSA SPID-based slicing architecture: UE classification by SPID, mapping to Resource Groups, and quota assignment through cell-load-adaptive profiles.

B. PRB Quota Allocation Model

Resource sharing is organized around a Common Resource Pool (CRP): a reserved fraction of cell PRBs shared across Resource Groups. In the trial the CRP was set to 30% of cell PRBs. Within this pool, each RG is governed by two parameters: the minimum quota (minQuota), a guaranteed PRB share in the range 0–100%, and the maximum quota (maxQuota), an upper bound in the range 1–100%.

The target share allocated to a Resource Group within the Common Resource Pool is derived from the relative minimum quotas of the competing groups. For two Resource Groups it can be expressed as:

$$Share(RG_i) = minQuota_i / (minQuota_1 + minQuota_2) \times CRP \quad (1)$$

As a worked example, with the CRP shared between RG#1 and RG#2, a configuration of minQuota = 50 / maxQuota = 100 for RG#1 and minQuota = 20 / maxQuota = 50 for RG#2 yields target shares of approximately 68.75% for RG#1 and 31.25% for RG#2. The smartphone group is thereby guaranteed roughly twice the protected resource share of the router group.

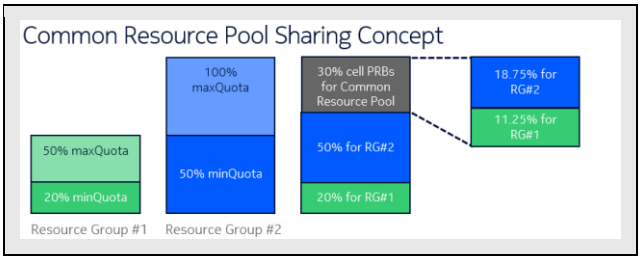

Fig. 2. Common Resource Pool concept. A reserved 30% of cell PRBs is divided between RG#1 and RG#2 according to their relative minimum quotas, producing the computed target shares.

C. Cell-Load-Adaptive Profile Selection

Three Resource-Group profiles were trialed so that quota policy could adapt to instantaneous cell occupancy. The high-load profile applies under heavy cell loading; the low-to-medium profile relaxes the router ceiling under lighter loading; and an extended-range profile permits the router group to reach the full pool when smartphone demand is low. The resulting target shares are summarized in Table I.

TABLE I
Cell-Load-Adaptive Quota Profiles and Computed Target Shares

Profile	Condition	RG#1 Min/Max	RG#2 Min/Max	RG#1 Target	RG#2 Target
P-0	High (≥40 UE)	50 / 100	20 / 50	68.75%	31.25%
P-1	Low–Med (≤40 UE)	60 / 100	20 / 60	63.64%	36.36%
P-2	Extended	60 / 100	20 / 100	61.54%	38.46%

D. Scheduler Activation Parameters

Enabling the mechanism requires activation of a set of RAN scheduler features at the base-station level. The trial enabled SPID-based NSA slicing, the slice-aware scheduler and its uplink enhancement, and a high slice-weight factor that strengthens the priority differential between groups. The per-Resource-Group cell-utilization measurement was configured with a 15-minute interval to drive profile selection.

IV. PARAMETER CONFIGURATION AND TRIAL METHODOLOGY

A. SPID Group Configuration

Two SPID groups were defined to separate the device classes, as shown in Table II. The smartphone group was bound to the lower SPID value and the router group to the higher value, each under the configured PLMN.

TABLE II
SPID Group to UE-Class Mapping

SPID Group	Parameter	Value	UE Type
Group-1	spid	1	Smartphone
Group-2	spid	16	Router / CPE

B. Trial Sites and Deployment Context

The mechanism was evaluated across two urban deployment clusters, designated Cluster A (urban dense) and Cluster B (high-density urban). Both clusters used an FR1 multi-band layout with a primary capacity band (n78), a secondary capacity band (n77), a mid-band coverage layer (n40), and low-band coverage extension (n28/n71). The measured user profile in Cluster B was approximately 70% smartphone and 30% router.

Cluster A was assessed through a sequence of incremental parameter changes producing three successive POST phases, allowing the contribution of each change to be isolated. Cluster

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

B was assessed using a fixed PRE window and a fixed POST window of comparable duration. Two KPI families were collected: N-KPI, comprising network-side counters such as Medium Access Control (MAC) throughput, PRB utilization, average NSA user count, control-channel starvation, and scheduling delay; and S-KPI, comprising user-perceived downlink and uplink throughput obtained from a commercial crowd-sourced speed-measurement platform.

C. Baseline Conditions

In the PRE phase, primary/secondary cell load balancing was disabled and NSA slicing was inactive, so all UE were scheduled from the common pool without differentiation. Under these conditions router UE camped predominantly on the mid-band coverage layer, concentrating load and contributing to control-channel congestion on that layer.

V. PERFORMANCE EVALUATION — CLUSTER A (URBAN DENSE)

A. Downlink Throughput

MAC-layer user downlink throughput improved progressively across the three POST phases, with the final configuration delivering the largest gain of 5.2% relative to the PRE baseline. Each successive parameter change contributed incrementally, and the last change produced the strongest single improvement, indicating that the fully tuned configuration—not merely feature activation—was responsible for the realized benefit.

B. Downlink Traffic Volume

Downlink MAC traffic volume increased by 1.2% following implementation. This rise is consistent with the throughput improvement: higher achievable rates enable greater data consumption within the same observation window.

C. NSA User Count

The average downlink NSA user count increased by 1.2%. The improvement is attributed to load balancing, which distributes users more evenly across layers and allows additional users to be admitted without degrading per-user performance.

D. Per-Slice PRB Utilization

PRB usage per Resource Group followed the configured quota algorithm in every POST phase: the smartphone group consistently received a larger PRB share than the router group, confirming that the scheduler enforced the intended target shares rather than allocating resources proportionally to offered load.

E. User-Perceived Speed Measurements

Speed-measurement results showed a clear improvement in both downlink and uplink across the PRE and successive POST phases. Devices with modern chipsets maintained their performance throughout, and the consistency of the gains supported a recommendation for wider, city-scale cluster deployment.

F. Control-Channel Starvation

On the mid-band layer, uplink PDCCH starvation and Physical Downlink Shared Channel (PDSCH) blocking caused by control-channel congestion were both reduced after implementation, as load balancing relieved that layer. A slight increase in starvation was observed on the n78 and n77 layers as users migrated onto them; this redistribution was assessed as an acceptable trade-off given the overall improvement.

VI. PERFORMANCE EVALUATION — CLUSTER B (HIGH-DENSITY URBAN)

A. Aggregate KPI Summary

Across Cluster B, the aggregate effect of the deployment was positive on nearly every indicator. Table III summarizes the principal PRE-to-POST changes spanning both network counters and user-perceived measurements.

TABLE III
Cluster B Aggregate PRE-to-POST KPI Changes

KPI	Domain	Change
DL UE PDCP throughput	N-KPI	+2.5%
UL UE PDCP throughput	N-KPI	+0.5%
DL traffic volume	N-KPI	+0.5%
UL traffic volume	N-KPI	+1.5%
Average NSA users	N-KPI	+5.8%
DL throughput	S-KPI	+2.75%
UL throughput	S-KPI	+10%
DL competitive rank	S-KPI	Held 1st
UL competitive rank	S-KPI	Rose to 1st

B. Per-Band Analysis

Disaggregating the results by band reveals the redistribution mechanism by which slicing and load balancing jointly act. As router UE are released from the mid-band layer, traffic and users migrate to the capacity and coverage layers, changing each layer's loading and per-user performance.

n78 (Primary Capacity Layer):

Downlink traffic rose 25.2% and uplink traffic 19.9% as routers redistributed onto this layer. Downlink PDCP throughput improved 2.1% and uplink 0.7%, while the average NSA user count increased 38.2%. PRB and slot utilization rose accordingly; a minor reduction in handover success and call access reflected the higher volume of secondary-node addition attempts and was statistically expected.

n40 (Mid-Band Coverage Layer):

Downlink traffic fell 29.5% and uplink 29.1% as the layer was relieved of router load. With congestion reduced, downlink PDCP throughput improved 6% and uplink 4%, and the average NSA user count declined 9.6%. PRB and slot utilization decreased correspondingly.

n77 (Secondary Capacity Layer):

Downlink traffic rose 34.5% and uplink 30.7%. A small reduction in downlink PDCP throughput of 1.1% and a larger reduction in uplink of 13% accompanied the increased loading, with the average NSA user count rising 22%. These reductions were assessed as acceptable given the substantial traffic absorbed by the layer.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

n28/n71 (Coverage Extension):

Downlink traffic rose 17.7% and uplink 16.4% as routers became able to camp on these layers at the cell edge. Downlink PDCP throughput fell 2.2% while uplink rose 1.6%, and the average NSA user count increased 23.7%, reflecting newly enabled edge camping.

C. Traffic Distribution Evolution

Aggregating the per-band movements, the share carried by the n78 and n77 capacity layers increased while the mid-band share decreased, as the previously enforced concentration of routers on the mid-band layer was removed and routers were allowed to camp freely. For sites combining n78 and n40, the bandwidth ratio implies an ideal load split of approximately 60% on n78 to 40% on n40, and the post-trial distribution trended toward this target.

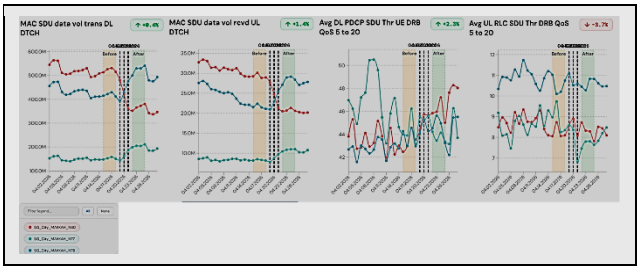

Fig. 3. Per-band traffic and user redistribution following joint deployment of cell load balancing and NSA slicing. Mid-band load decreases while capacity and coverage layers absorb the released router traffic.

VII. CROSS-LAYER ANALYSIS: PER-SLICE UE BEHAVIOR

A. Throughput Differentiation

Across both clusters, smartphone UE consistently achieved higher downlink MAC and uplink Radio Link Control (RLC) throughput than router UE. This ordering is the intended outcome of the quota policy and confirms that the gNB scheduler prioritized RG#1 over RG#2 in line with the configured minimum quotas.

B. Scheduling Delay

Despite a higher count of smartphone NSA users, the proportion of smartphone UE awaiting scheduling was lower than for routers, indicating effective prioritization. In absolute terms smartphones exhibited slightly higher scheduling delay than routers, attributable to their larger aggregate demand rather than to deprioritization; the behavior remained consistent with minimum-quota-enforced priority.

C. UE Layer Camping Behavior

Before implementation, router UE were largely confined to the mid-band layer. After implementation, routers camped freely across all available layers, including capacity and edge-coverage bands. Discontinuous Reception (DRX) for router UE was disabled during the trial, keeping routers continuously active in the scheduler; re-enabling DRX is recommended to release scheduler capacity and improve smartphone scheduling headroom.

D. User Profile Alignment

With a measured profile of 70% smartphone and 30% router in Cluster B, the high-load quota configuration—minimum quota 50 for smartphones and 20 for routers—produces a protected PRB ratio of approximately 68.75 to 31.25. This allocation aligns closely with the underlying user-population ratio, indicating that the chosen quotas were well matched to the deployment.

VIII. DISCUSSION AND OPTIMIZATION RECOMMENDATIONS

A. Quota Tuning Strategy

The trial's central operational finding is that minimum and maximum quota values must be actively monitored and tuned, not set once and left fixed. Effective tuning responds to cell-level user count, which triggers profile selection; to the smartphone-to-router traffic mix per cell; and to layer-specific PRB utilization. A semi-static approach to quota configuration was introduced in Cluster B to provide more deliberate PRB-volume control between slices, and proved effective.

B. DRX Re-enablement for Router UE

Because router UE were held continuously active, they consumed scheduler resources even when idle. Re-enabling DRX for the router group would free downlink and uplink scheduler capacity for smartphone UE, directly reducing smartphone scheduling delay. This is recommended as a near-term tuning action.

C. Interaction with Cell Load Balancing

Primary/secondary cell load balancing and NSA slicing produced better overall performance together than either feature alone. Load balancing redistributes router UE onto underutilized layers, relieving mid-band congestion and allowing slicing to operate on a more evenly loaded carrier. The two features are therefore complementary and are best deployed jointly.

D. Scalability

The consistency of the gains across both clusters supports city-wide and larger-cluster deployment. A progressive rollout with continued PRE/POST monitoring is recommended, using the combined N-KPI and S-KPI scorecard methodology demonstrated here as the standard evaluation framework.

IX. FUTURE ENHANCEMENTS AND ROADMAP

Several enhancements are planned to extend the value of RAN slicing beyond the current SPID-based scheme:

- Slice-aware scheduling with QoS and UE fairness, incorporating per-UE QoS-flow awareness within each Resource Group to provide finer differentiation than SPID grouping and to maintain fairness when UE counts are asymmetric.
- Low-latency slice optimization through Scheduling Request and proactive-grant enhancements, reducing request-processing delay for real-time applications assigned to a dedicated low-latency Resource Group; this is proposed for near-term live-network trial.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

- Dynamic quota adaptation, evolving from semi-static profile selection toward continuous quota adjustment driven by real-time PRB-utilization feedback and per-slice traffic prediction.
- A migration path toward 5G Standalone, enabling end-to-end slicing with S-NSSAI propagation through the core and removing the RAN-only constraint inherent to NSA.

X. CONCLUSION

This paper reported a live-network trial of SPID-based NSA network slicing with PRB-quota management across two urban deployment clusters. The mechanism proved to be an effective and immediately deployable means of differentiating quality of service between heterogeneous device classes without requiring core-network slicing support. Measured improvements reached 5.2% in downlink throughput and 10% in uplink throughput, alongside gains in user-perceived speed measurements and average NSA user count.

The gNB scheduler was confirmed to enforce Resource-Group priority correctly, granting smartphone UE a higher PRB share and lower scheduling-wait probability than router UE. The decisive operational lesson is that minimum and maximum quota tuning matters as much as feature activation: incorrectly set quotas limit the realizable gain. Finally, the joint deployment of cell load balancing and NSA slicing produced synergistic benefits unavailable from either feature in isolation. Planned enhancements targeting QoS-aware and low-latency slicing are expected to further expand the practical value of RAN slicing in NSA networks.

References

[1] 3GPP, “NR; Overall description; Stage-2,” TS 38.300, 3rd Generation Partnership Project, Tech. Spec., 2023.

[2] 3GPP, “System architecture for the 5G System (5GS),” TS 23.501, 3rd Generation Partnership Project, Tech. Spec., 2023.

[3] Mohammed Yahiya Pasha Gulam ”2L-4H Beams Antenna,” IJERT Volume 15, Issue 05 , May– 2026.

[4] 3GPP, “Telecommunication management; Study on management and orchestration of network slicing,” TR 28.801, 3rd Generation Partnership Project, Tech. Rep., 2018.

[5] 3GPP, “NR; Physical layer procedures for data,” TS 38.214, 3rd Generation Partnership Project, Tech. Spec., 2023.

[6] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network slicing in 5G: Survey and challenges,” IEEE Commun. Mag., vol. 55, no. 5, pp. 94–100, May 2017.

[7] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, “Network slicing and softwarization: A survey on principles, enabling technologies, and solutions,” IEEE Commun. Surveys Tuts., vol. 20, no. 3, pp. 2429–2453, 2018.

[8] P. Rost et al., “Network slicing to enable scalability and flexibility in 5G mobile networks,” IEEE Commun. Mag., vol. 55, no. 5, pp. 72–79, May 2017.

[9] Abdul Quader Syed, Syed Naveed Maqdoom, & Hasan Omair Mohammed. (2026). 5G Energy-Efficient Massive MIMO Evolution Using ELAA and Adaptive Beamforming Techniques. International Journal of Engineering Research & Technology, Volume 15(Issue 06 , June - 2026). https://doi.org/10.5281/zenodo.20591348.

[10] A. Ksentini and N. Nikaein, “Toward enforcing network slicing on RAN: Flexibility and resources abstraction,” IEEE Commun. Mag., vol. 55, no. 6, pp. 102–108, Jun. 2017.

[11] O. Sallent, J. Perez-Romero, R. Ferrus, and R. Agusti, “On radio access network slicing from a radio resource management perspective,” IEEE Wireless Commun., vol. 24, no. 5, pp. 166–174, Oct. 2017.

[12] Abdul Quader Syed, Hasan Omair Mohammed, & Syed Naveed Maqdoom. (2026). Prioritized Physical Resource Block Allocation for Mobile Handsets over Fixed-Wireless Routers to Enhance Throughput and Speedtest Performance in 5G NSA Networks. International Journal of Engineering Research & Technology, Volume 15(Issue 06 , June - 2026). https://doi.org/10.5281/zenodo.20538729.

[13] S. Vassilaras et al., “The algorithmic aspects of network slicing,” IEEE Commun. Mag., vol. 55, no. 8, pp. 112–119, Aug. 2017.

[14] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, “NFV and SDN—Key technology enablers for 5G networks,” IEEE J. Sel. Areas Commun., vol. 35, no. 11, pp. 2468–2478, Nov. 2017.