# Context-Aware Slice Negotiation for Enhanced 5G User Experience

**Mahesh Devdatta Telang**

**Abstract:** ***Traditional 5G network slicing, predominantly orchestrated from the network-side, often exhibits limited responsiveness to dynamic user application requirements. This paper investigates user-driven autonomous 5G network slicing, a paradigm empowering User Equipment (UE) to autonomously request and manage network slices facilitated by lightweight on-device Artificial Intelligence (AI). This user-centric control, informed by real-time application needs and prevailing network conditions, holds the promise of enhancing network efficiency, optimizing resource utilization, and elevating the Quality of Experience (QoE). The technical viability of this approach hinges on the 5G Standalone (SA) architecture, advancements in lightweight AI algorithms optimized for resource-constrained devices, and requisite extensions to existing UE-network interaction mechanisms. Key anticipated benefits include optimized resource consumption and improved application performance through dynamic, context-aware adaptation. Nevertheless, significant implementation challenges persist, encompassing the deployment and lifecycle management of AI models across diverse UEs, potential signaling overhead, the imperative for robust security in a UE-controlled environment, and the necessity for standardized communication protocols and Application Programming Interfaces (APIs). This paper provides a comprehensive review of this emerging concept, meticulously analyzing its feasibility, potential benefits, inherent challenges, and outlining future research and standardization trajectories essential for its successful realization.***

**Keywords:** Network slicing, slice negotiation, 5G Standalone mode, Quality of Experience (QoE), Artificial intelligence.

## 1. Introduction

Network slicing is a foundational feature of Fifth Generation (5G) mobile networks, enabling the creation of multiple virtualized, logically isolated, end-to-end networks upon a shared physical infrastructure [1]. This architectural innovation allows operators to customize network characteristics—such as bandwidth, latency, reliability, and connection density—to satisfy the diverse requirements of distinct service categories, including enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC) [2]. As network technologies progress towards Sixth Generation (6G) systems, network slicing is anticipated to assume an even more critical role in supporting hyper-flexible, intelligent frameworks capable of accommodating the stringent demands of future applications, such as extended reality (XR) and autonomous systems [3].

However, the management and orchestration of these network slices have, to date, predominantly relied on network-centric mechanisms. Mobile Network Operators (MNOs) typically manage slices from the network core, often based on predefined Service Level Agreements (SLAs) or relatively static operational policies. This network-managed approach can manifest limitations in its capacity to react instantaneously to the highly dynamic needs of individual users, specific applications, or rapidly fluctuating network conditions experienced at the device level. This inherent latency in adaptation can precipitate suboptimal resource utilization and potentially degrade the Quality of Experience (QoE) for end-users [4].

To surmount these limitations, user-driven autonomous network slicing emerges as a transformative paradigm. This approach advocates a fundamental shift, empowering the User Equipment (UE) to actively participate in managing its network connectivity by autonomously requesting, modifying, and releasing network slices based on its immediate and predicted requirements. The core tenet involves embedding lightweight Artificial Intelligence (AI) capabilities directly onto the UE. These on-device AI models would continuously monitor application behavior, assess real-time performance needs, perceive current network conditions, and make intelligent, localized decisions regarding the optimal network slice configuration. This user-centric model directly confronts the responsiveness challenge inherent in centralized, network-managed slicing. The anticipated advantages are substantial, encompassing more efficient network resource utilization, potential operational cost savings, and a markedly improved end-user QoE. Furthermore, situating slice management intelligence at the UE signifies a progression towards a more distributed and agile network control architecture.

This paper aims to furnish a comprehensive technical review of the concept of user-driven autonomous 5G network slicing. It analyzes the underlying technical feasibility, explores the potential benefits, identifies the significant implementation challenges, and discusses the necessary advancements in technology and standardization required for its widespread adoption.

## 2. Background

Current commercial deployments of 5G network slicing primarily target the enterprise sector through Business-to-Business (B2B) arrangements. MNOs offer customized network slices with guaranteed performance characteristics to various industries, including manufacturing, logistics, healthcare, and transportation. Business-to-Consumer (B2C) models, while less prevalent, are gradually emerging. Pricing strategies encompass subscription-based, policy-based, and application-specific slicing tiers. SLAs are pivotal in these commercial offerings, guaranteeing specific network performance metrics. Existing commercial models predominantly depend on static or semi-static provisioning, which contrasts sharply with the user-driven autonomous

Paper ID: SR25516122237 DOI: https://dx.doi.org/10.21275/SR25516122237

slicing concept that envisions highly dynamic, fine-grained slice management initiated by individual UEs based on real-time application exigencies.

Research into autonomous network management is advancing, with a significant concentration on network-side autonomy where intelligence resides within the network infrastructure itself. AI and Machine Learning (ML) are integral to these endeavors, employed to automate and optimize diverse aspects of slice lifecycle management [5]. Federated Learning (FL) has surfaced as a particularly pertinent technique, enabling the collaborative training of AI/ML models across distributed network entities without necessitating the centralization of raw data, thereby preserving data privacy [6]. While FL offers considerable promise for enhancing model accuracy and privacy, it introduces its own set of challenges, such as managing non-Independent and Identically Distributed (non-IID) data from disparate sources, controlling communication overhead during model updates, and ensuring model convergence and robustness in dynamic environments [7]. The concept of fully autonomous, real-time 5G slice management, dynamically driven by the UE based on application needs and perceived network conditions, remains relatively nascent in the literature. Table I provides a comparative overview of network-side versus user-driven autonomous slicing approaches.

**Table I:** Comparison of Network-Side Vs. User-Driven Autonomous Slicing

| Feature | Network-Side Autonomous Slicing | User-Driven Autonomous Slicing |
|---|---|---|
| Control | Network (Core/RAN/Orchestrator) | User Equipment (UE) with Network Support |
| Primary Trigger | Network policies, analytics, operator intent | UE application needs, perceived network conditions |
| Granularity | Typically Service-level or Group-level | Application-level, potentially User-specific |
| Responsiveness | Near real-time to minutes (depending on loop) | Potentially real-time (limited by signaling/AI) |
| Standardization | More mature (3GPP MANO, ETSI ZSM network focus) | Emerging (requires extensions to UE/Core specs, APIs) |
| Key Challenges | Scalability, Inter-domain coordination, Complexity | On-device AI, Signaling overhead, Security, UE diversity |
| Primary Focus | Operator efficiency, Enterprise services | End-user QoE, Dynamic consumer applications |

The 3rd Generation Partnership Project (3GPP) standards form the bedrock for 5G network slicing, with the 5G Standalone (SA) architecture being a fundamental prerequisite [8]. 3GPP specifications delineate procedures for UE involvement in slice selection processes. Ongoing work within 3GPP Releases 18 and 19 aims to further enhance slicing and automation capabilities, including the integration of AI/ML functionalities and the fortification of security mechanisms. Concurrently, the European Telecommunications Standards Institute's (ETSI) Industry Specification Group on Zero-touch network and Service Management (ISG ZSM) is pivotal in defining architectures for end-to-end network and service automation, with the goal of achieving fully autonomous networks [9]. The realization of seamless, interoperable End-to-End (E2E) network slicing, particularly a user-driven model, necessitates robust collaboration and alignment across various Standards Development Organizations (SDOs) and open-source communities. User-driven slicing, specifically, will require targeted extensions to existing standardization efforts to accommodate UE-initiated dynamic slice control.

## 3. Enabling Technologies

The materialization of user-driven autonomous network slicing is contingent upon the convergence and maturation of several pivotal technologies.

The 5G Standalone (SA) architecture is an indispensable prerequisite for advanced network slicing capabilities, including the proposed user-driven model. Its cloud-native, Service-Based Architecture (SBA) and Control and User Plane Separation (CUPS) furnish the requisite flexibility to instantiate and manage distinct network slices dynamically [10]. Key 5G Core (5GC) Network Functions (NFs)—such as the Access and Mobility Management Function (AMF), Session Management Function (SMF), User Plane Function (UPF), Network Slice Selection Function (NSSF), and Unified Data Management (UDM)—are intrinsically involved in network slice management. For user-driven autonomous slicing, these NFs would necessitate enhanced logic and standardized interfaces to proficiently process dynamic slice management requests originating from the UE, based on real-time needs signaled by on-device AI.

Existing 3GPP UE-Network Interaction Mechanisms provide foundational procedures for UEs to participate in network slice selection during registration and Protocol Data Unit (PDU) session establishment [8]. However, these mechanisms are primarily designed for static or semi-static selection and are not inherently suited for the dynamic, real-time, autonomous slice management envisioned. Actualizing user-driven autonomy mandates the definition of new or significantly extended signaling procedures within 3GPP standards. These extensions must allow the UE to efficiently signal dynamic requests for slice activation, modification, or deactivation, with paramount importance placed on minimizing the resultant signaling overhead to maintain network stability and efficiency.

The core intelligence underpinning user-driven autonomous slicing resides in Lightweight On-Device AI. Processing data and rendering decisions locally on the UE offers distinct advantages, including real-time responsiveness, enhanced context awareness, and inherent privacy preservation [11]. Nevertheless, deploying AI on UEs presents considerable challenges attributable to their inherent resource constraints (e.g., processing power, memory, energy). Addressing these constraints necessitates the application of specialized techniques such as model pruning, quantization, knowledge distillation, and Neural Architecture Search (NAS). Frameworks like TensorFlow Lite, PyTorch Mobile, and Core ML are instrumental in facilitating the deployment of optimized models on mobile and embedded devices. On-device AI models would be tasked with predicting application performance requirements, optimizing slice requests accordingly, and continuously monitoring and adapting slice usage. Federated Learning (FL) offers a compelling paradigm for training these on-device models collaboratively while upholding user data privacy [6, 12]. However, FL introduces complexities related to communication overhead for model aggregation, managing statistical heterogeneity across devices (non-IID data), ensuring security against adversarial attacks, and guaranteeing model convergence and performance. The feasibility of user-driven slicing is critically dependent on the synergistic advancement of lightweight AI techniques, including FL, and the development of efficient, standardized UE-network signaling protocols.

For the on-device AI to make efficacious decisions, it must accurately Correlate Network Parameters with Application QoE. The AI needs to comprehend how measurable network quality parameters translate into perceived application performance and, ultimately, user QoE. Key 5G radio signal quality parameters include Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and Signal-to-Interference-plus-Noise Ratio (SINR). The correlation between these radio parameters and application-level performance is intricate and highly contingent on the specific application type (e.g., video streaming, interactive gaming, file transfer) [13]. Establishing a clear, accurate, and dynamically adaptable mapping between network parameters and perceived performance is essential for the on-device AI to reliably trigger autonomous slicing actions. This mapping might itself need to be learned and continuously refined by the AI model through experience.

## 4. Framework and Challenges

Implementing user-driven autonomous network slicing necessitates a well-defined conceptual framework. A potential architecture involves intricate interactions across multiple layers: the User Application Layer, the UE Middleware/Operating System (OS) Layer (which hosts the on-device AI engine), the UE Modem/Radio Access Network (RAN) Interface, and the Network (RAN and Core). The operational workflow would typically involve applications signaling their dynamic performance needs, the on-device AI assessing current network conditions and historical data to determine optimal slice requirements, the UE subsequently signaling these slice requests to the network, the network processing these requests and configuring or modifying the slice accordingly, followed by continuous monitoring and adaptation by both the UE and the network.

A critical enabler for this framework is the ability of user applications to effectively communicate their real-time performance needs to the UE's AI-driven slice management layer. Existing 5G QoS frameworks provide a foundational basis, but a standardized API is deemed essential to bridge the communication gap between applications and the UE's slice management logic. Such an API would empower applications to declaratively express their dynamic QoS requirements (e.g., minimum bandwidth, maximum latency, reliability targets) and receive feedback on slice availability and performance.

Despite the compelling potential benefits, the practical implementation of user-driven autonomous network slicing faces substantial challenges:

1) On-Device AI Deployment and Management: Significant difficulties are anticipated in training, deploying, managing, and updating accurate and efficient lightweight AI models across a vast and heterogeneous ecosystem of UEs, each with varying capabilities and resource constraints. Ensuring model consistency, robustness, and security across this diverse landscape is a formidable task.
2) Signaling Overhead: The prospect of potentially millions of UEs frequently transmitting signaling messages to request, modify, or release network slices raises legitimate concerns about the potential load on the network control plane. Efficient signaling protocols and intelligent aggregation or prioritization schemes are imperative to mitigate this risk.
3) Network Security: Granting UEs a greater degree of autonomy in slice management inherently introduces new security vulnerabilities. Malicious or compromised UEs could potentially disrupt network operations or attempt to gain unauthorized access to resources. Robust authentication, authorization mechanisms, and strict isolation between slices and UEs are paramount to maintaining network integrity.
4) Coordination and Standardization: The current lack of mature and universally accepted standards for dynamic UE-initiated slice management, as well as for the crucial application-to-UE API, represents a major impediment to interoperability and widespread adoption.
5) Resource Contention and Scalability: In scenarios with high user density or during peak demand periods, multiple simultaneous UE requests for specific slice resources can lead to contention. Intelligent network-side admission control policies and scalable system designs are crucial to manage resource allocation effectively and ensure fair access [14].

## 5. Potential Benefits and Use Cases

The successful implementation of user-driven autonomous network slicing promises significant and tangible benefits for both end-users and network operators.

By transitioning to a dynamic, on-demand model driven by actual UE and application needs, this approach promotes a more optimized resource utilization within the network. Resources are allocated precisely when an application requires them and can be released promptly when the need subsides, minimizing wastage and improving overall spectral and infrastructure efficiency. This enhanced efficiency can, in turn, translate into cost savings for both users (e.g., through more granular service tiers) and MNOs (e.g., through reduced over-provisioning and operational expenditures).

Perhaps the most compelling benefit lies in the potential for a significantly Enhanced End-User QoE. By empowering the UE to autonomously ensure its applications receive the necessary network QoS, users can enjoy smoother, more reliable, and more responsive application performance. This is particularly critical for emerging and demanding applications. Illustrative use cases include:

1) Adaptive Video Streaming: The UE AI can dynamically request a slice with a guaranteed bit rate and low jitter when high-resolution video streaming is initiated and

then modify or release the slice as viewing habits change, ensuring an uninterrupted and high-quality viewing experience [15].

2) Online Gaming: For competitive online gaming, the UE can request an URLLC-type slice to ensure minimal latency and packet loss, crucial for real-time responsiveness and a fair gaming experience [16].
3) Extended Reality (XR) Applications: AR/VR applications demand both high bandwidth and extremely low latency. User-driven slicing can ensure that these stringent requirements are met on-demand, enabling immersive and seamless XR experiences [3, 17].

## 6. Conclusion

User-driven autonomous 5G network slicing presents a compelling and logical evolution beyond traditional network-centric management paradigms. By empowering UEs with on-device AI to facilitate dynamic slice management, this approach offers the potential for significant improvements in network resource efficiency, operational cost optimization, and substantial enhancements to end-user QoE. The technical feasibility of this vision is underpinned by the architectural flexibility of the 5G Standalone (SA) core, continuous advancements in lightweight AI and on-device machine learning, and the existing, albeit
foundational, UE-network interaction mechanisms.

However, the path to realizing this user-centric vision is laden with significant challenges that necessitate concerted and collaborative research, development, and standardization efforts. Key hurdles include the practical complexities of deploying, managing, and securing AI models on a diverse and extensive range of UEs, the potential for unsustainable signaling overhead on network control planes, the critical imperative to address new and evolving security vulnerabilities introduced by increased UE autonomy, the current absence of universally adopted standardized signaling protocols and APIs for UE-driven slice control, and the intricate problem of managing resource contention in a highly dynamic environment.

Addressing these multifaceted challenges will require focused and sustained efforts across several domains: advancements in AI/ML research (particularly in areas of efficient on-device learning and federated learning), proactive and comprehensive standardization activities within bodies like 3GPP and ETSI, the development of robust and adaptive security frameworks, rigorous system integration and optimization, and extensive prototyping and real-world trials to validate performance and identify unforeseen issues.

User-driven autonomous network slicing represents a promising and potentially disruptive evolutionary step, aligning congruently with the broader industry trends towards increased automation, embedded intelligence, and enhanced user-centricity envisioned for 5G Advanced and the forthcoming 6G networks. While the implementation challenges are undeniably substantial, requiring significant innovation and cross-ecosystem collaboration, the potential rewards in terms of network operational efficiency and transformative user experiences render user-driven autonomous network slicing a critical area for continued exploration and development.

## References

[1] Afolabi, I., Taleb, T., Samdanis, K., Ksentini, A., & Flinck, H. (2018). Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions. IEEE Communications Surveys & Tutorials, 20(3), 2429–2553.
[2] Ordonez-Lucena, J., et al. (2017). Network Slicing for 5G with LA/RA/SLA: A Service-Oriented Framework. IEEE Communications Magazine, 55(5), 78–85.
[3] Saad, W., Bennis, M., & Chen, M. (2019). A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems. IEEE Network, 34(3), 134–142. (Also available as a preprint on arXiv: arXiv)
[4] Taleb, T., et al. (2017). On Multi-Access Edge Computing: A Survey of the Emerging Approaches and Relevant Areas of Research. IEEE Communications Surveys & Tutorials, 19(3), 1657–1681.
[5] Boutaba, R., et al. (2018). A Comprehensive Survey on Machine Learning for Networking: Evolution, Applications and Research Challenges. IEEE Communications Surveys & Tutorials, 21(2), 1747–1782.
[6] McMahan, B., et al. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS). (also available on arXiv: arXiv)
[7] Kairouz, P., et al. (2021). Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning, 14(1–2), 1–210. (Preprint available on arXiv: arXiv)
[8] 3GPP TS 23.501. (2023). System Architecture for the 5G System (5GS). (Official specification available via ETSI: iTeh Standards)
[9] ETSI GS ZSM 002. (2019). Zero-touch Network and Service Management (ZSM); Reference Architecture. (Published by ETSI: iTeh Standards)
[10] Rost, P., et al. (2017). Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks. IEEE Communications Magazine, 55(5), 72–79. (Preprint available on arXiv: arXiv)
[11] Shi, W., et al. (2016). Edge Computing: Vision and Challenges. IEEE Internet of Things Journal, 3(5), 637–646.
[12] Du, Z., et al. (2020). Federated Learning for Vehicular Internet of Things: Recent Advances and Open Issues. IEEE Open Journal of the Computer Society, 1, 75–90.
[13] Ksentini, A., & Nikaein, N. (2017). Toward Enforcing Network Slicing in 5G: A Survey on Challenges and Enabling Technologies. IEEE Communications Surveys & Tutorials, 20(1), 480–508.
[14] Jiang, M., et al. (2020). Network Slicing Resource Allocation: A Survey. IEEE Communications Surveys & Tutorials, 22(4), 2836–2861.
[15] Khan, L. U., et al. (2020). 5G Network Slicing: Resource Allocation and Management, Mobility, and Security. IEEE Communications Magazine, 58(9), 30–36.

[16] Parvez, I., et al. (2018). A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions. IEEE Communications Surveys & Tutorials, 20(4), 3098–3130.

[17] Mach, P., et al. (2017). In-band Control Plane for Network Slicing in 5G: A Survey. IEEE Communications Surveys & Tutorials, 20(1), 424–448.

## Author Profile

**Mahesh Devdatta Telang** received his B.E. in Electronics and Communication Engineering from Visvesvaraya Technological University in 2011 and an M.S. in Computer Engineering from the University of Texas at Dallas in 2013. From 2013 to 2019, he was a Senior Engineer at Qualcomm Technologies Inc., contributing to the development of 4G and 5G wireless communication technologies. Subsequently, he worked at Google from 2019 to 2022, where his focus was on baseband and connectivity for Android and Pixel devices. Mr. Telang currently works at Meta Platforms Inc., where he is involved in designing and developing innovative connectivity solutions for AR/VR/XR devices.