International Journal of Computer Networks and Applications (IJCNA)
DOI: 10.22247/ijcna/2023/220736 Volume 10, Issue 2, March – April (2023)

SURVEY ARTICLE

Resource Provisioning and Utilization in 5G Network Slicing: A Survey of Recent Advances, Challenges, and Open Issues

Simon Atuah Asakipaam
Department of Telecommunication Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
simonasakipaam@gmail.com

Jerry John Kponyo
Department of Telecommunication Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
jjkponyo.soe@knust.edu.gh

Kwame Oteng Gyasi
Department of Telecommunication Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
kermaabar@yahoo.com

Received: 17 December 2022 / Revised: 03 February 2023 / Accepted: 10 February 2023 / Published: 29 April 2023

Abstract – The increasing demands for higher bandwidth and lower latency in modern telecommunications networks have led to the exploration of network slicing as a means to meet these requirements more efficiently in next-generation 5G networks. Despite substantial academic interest in resource allocation and management in network slicing, existing research is dispersed and fragmented. This study presents a categorization and assessment of the latest research on resource allocation and optimization techniques in 5G network slicing. It also shows how advanced machine learning techniques can support resource management in sliced wireless networks. The present paper offers a complete overview and analysis of current solutions for resource allocation and management in 5G network slicing, outlines open research challenges, and suggests future research directions for researchers and engineers in this field.

Index Terms – Network Slicing, Resource Allocation, 5G Network, Management, Optimization, SDN, NFV.

1. INTRODUCTION

The fifth-generation network (5G) has been designed to fulfill diverse and complex requirements of various industries [1], such as healthcare, agriculture, industry, and entertainment [2]. A key aspect of 5G, Network Slicing (NS), enables multiple virtual segments of the network, called network slices, to run on a shared physical network infrastructure [3]. NS is made possible by the combination of Software-Defined Networking (SDN) and Network Function Virtualization (NFV), which decouple network computation and networking resources into logical networks that are not limited by physical resources [4]. According to the 3GPP [5], a network slice is a customized, logically isolated virtual network that is dynamically created by the assignment of network Service Function Chains (SFCs) and physical network resources to meet the demands of vertical businesses. Each SFC consists of interconnected Virtual Network Functions (VNFs) that define the processing of data flows associated with a particular service.

However, the diverse and competing quality of service requirements of vertical market applications [6], along with the massive amounts of data generated [7], present significant challenges for efficient resource management in 5G NS. For example, machine-type communication applications require constant availability, but may not communicate frequently and may use duty cycling to connect to the network [8]. Mission-critical applications demand dedicated network slices for optimal performance, while a massive industrial IoT slice may need a lightweight 5G core and a high number of connections, but no handover. On the other hand, a mobile broadband slice may demand a high-capacity core, full mobility support, and low latency. As the number of slices increases, the challenge of end-to-end slice management and orchestration becomes more complex. The existing research on resource allocation and management in 5G network slicing is dispersed, with various solutions proposed in the literature.

SURVEY ARTICLE

Some studies focus on the use of machine learning [9], while others use traditional optimization algorithms [10]. Previous survey studies have provided evaluations of 5G network slicing with SDN and NFV [11], collections of recent open-source software and frameworks for 5G cellular networks [12], advances in slice admission control and optimization algorithms [13][14], and the use of machine learning algorithms in 5G networks for energy efficiency [15][16].

However, these studies have limitations and gaps, which this survey paper aims to address by reviewing and evaluating the most recent research on resource allocation and optimization strategies in 5G NS and highlighting open research challenges and future research directions. A comparison of this survey paper with previous survey studies is presented in Table 1.

Table 1 Comparison between This Survey Paper and the Existing Surveys Papers

Reference	5G architecture	SDN and NFV	Standardization efforts	NS resource management			Research challenges
				Resource allocation	Traditional resource optimization	ML-based resource optimization	
[11], 2020	☒	✓	✓	✓	×	×	✓
[12], 2020	✓	✓	✓	×	×	✓	✓
[13], 2020	×	×	×	×	✓	✓	×
[14], 2020	✓	×	×	✓	✓	×	✓
[15], 2020	×	✓	×	×	×	✓	✓
[16], 2020	☒	☒	×	✓	×	×	✓
This paper	✓	✓	✓	✓	✓	✓	✓
Key	✓ : Full			☒: Partial			×: None

This paper's contributions can be summarized as follows:

- Provides a thorough survey of current literature (2017 to 2022) on 5G network slicing resource allocation and optimization.
- Provides commentary on the softwarized 5G network's architecture, and maps the reviewed solutions into this architecture.
- Outlines each solution's significance and functionality and identifies gaps in the existing research
- Identifies open research challenges and future research directions in 5G network slicing resource allocation and optimization.

The remainder of the paper is structured as follows: Section 2 discusses the design of 5G networks, enabling technologies for 5G NS, and challenges of dynamic network slicing. Section 3 reviews the latest literature on network slicing frameworks, resource provisioning, and network slicing optimization. Section 4 discusses open issues and possible research areas, while Section 5 concludes the paper.

2. BACKGROUND

2.1. Anatomy of 5g Cellular Network

Since the introduction of the cellular network, it has comprised a Radio Access Network (RAN), a Transport Network (TN), and a Core Network (CN). The RAN uses radio access technology to connect user equipment to the core network. The TN establishes and maintains connections for data transmission, while the CN offers services to consumers [17]. The 5G network architecture has been updated to meet industry demand while remaining consistent with these requirements [18]. The 5G RAN has a wider spectrum of carrier frequencies, including millimeter waves and a more flexible frame structure [19], and comes in two flavors: non-standalone and standalone [20]. This design is a significant advancement for 4G LTE, particularly its protocol stack, features, and capabilities. The Next Generation Node Bases (gNBs) or NR base stations enable decentralized deployment by dividing the protocol stack into separate hardware components, making it easier to implement virtualization [21]. In response to the need for openness and flexibility, the 5G Core (5GC) has adopted a service-based approach and implemented several network functions to decouple the control and user planes. These functions include User Plane Function (UPF), Access and Mobility Management Function (AMF), and Session Management Function (SMF), which manages IP allocation and user plane services [22]. The UPF serves as a public Internet user plane gateway, anchoring mobility and classifying incoming flow QoS. By using softwarization, 5G networks can separate services and functions from hardware and offer customized network services to multiple users [23]. Figure 1 illustrates the architectural design of a 5G network.

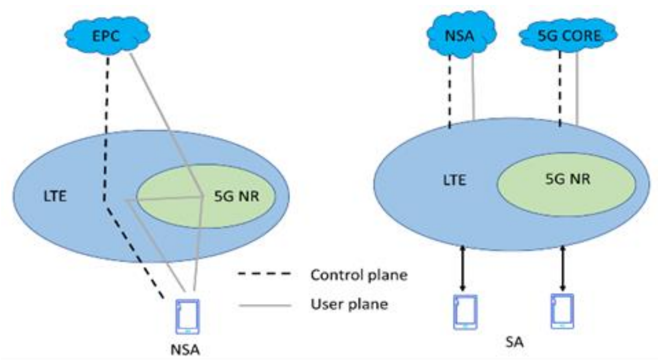

Figure 1 5G Network Architecture

2.2. Enabling Technologies for Softwarized 5G Cellular Networks

2.2.1. Software-Defined Networking (SDN)

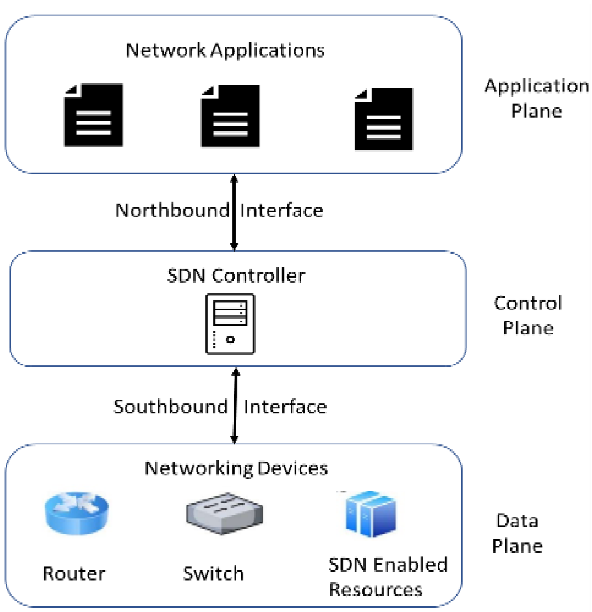

Figure 2 SDN Architecture

Traditional IP networks consist of three interlinked planes: control, data, and management [24]. The management plane sets new policies, which the control plane implements, while the data plane enables traffic flow by enforcing the policies set by the control plane. This architecture leads to difficulties in maintaining the networks, handling misconfigurations, limited room for innovation, and high costs when adding new network capabilities [25]. To overcome these challenges, SDN employs softwarization, as depicted in Figure 2, to separate network control from the forwarding layer and network programmability to make the network more flexible and controllable [26]. By isolating control plane functions, SDN simplifies the data plane hardware, resulting in lighter and more straightforward networking forwarding devices compared to traditional routers/switches. The SDN controller is responsible for network intelligence, maintaining a wider network view, and making policy decisions for automatic network optimization and management [27]. 5G networks have applied this technique to separate NR and 5GC hardware components from their networking and service capabilities. SDN has enabled the disaggregation of NR, with radio units serving as simple transceivers and control and processing done via open standards in software.

2.2.2. Network Function Virtualization (NFV)

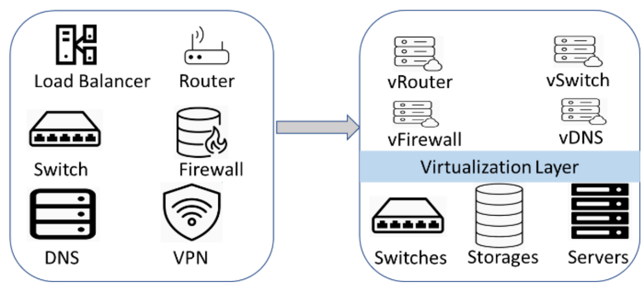

Figure 3 Virtualization of Network Functions

Traditionally, service providers have utilized dedicated hardware middleboxes to execute network functions such as firewalls, intrusion detection systems, and network optimizers [28]. However, these dedicated hardware middleboxes are expensive and limit deployment and administrative flexibility, particularly with the massive scale connection of 5G networks and diverse traffic requirements [29]. Network Function Virtualization (NFV) addresses these limitations by delivering network functions in a virtual environment as a service, improving resource consumption, application performance, and network resource utilization. NFV also enables flexible administration and orchestration of networks [30].

This is accomplished by implementing each service in software using VNFs that run on VMs built on general-purpose hardware and controlled by hypervisors [31]. Several VNFs can be joined to create robust and tailored network services and can be hosted in both large data centers and smaller facilities at the network's edge [32]. Figure 3 illustrates the concept of network function virtualization.

SURVEY ARTICLE

2.3. Network Slicing for 5G Mobile Networks with NFV / SDN

Due to the vast scale of connectivity in 5G networks and the diversity of traffic needs, the "one size fits all" approach is no longer feasible [33]. To address this, Next Generation Mobile Network Alliance [34] introduced the notion of network slicing. This technique uses virtualization technology to separate computing and networking resources into logical networks, known as network slices, allowing for the creation of on-demand networks that can be adapted to provide optimal services in various market circumstances [35], [36] [37] as depicted in Figure 4.

One of the biggest challenges for NS is coordinating the distribution and usage of shared resources while ensuring that each service adheres to its own set of rules and resource limits. This challenge is further complicated by the rapidly increasing network size and continually changing network environment in 5G [38]. However, advances in NFV and SDN technologies have made network slicing possible across all domains in 5G networks, including RAN, transport, core, and end-to-end [39] [40]. Table 2 provides a comparison of NFV, SDN, and NS, highlighting their differences, potential, and synergies. Further information on the standardization of NS, NFV, SDN, and 5G can be found in [11].

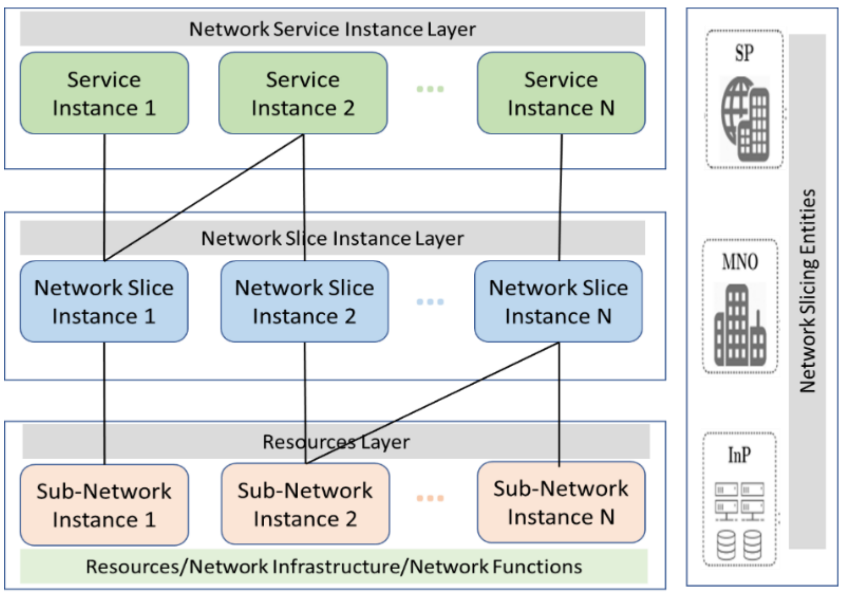

Figure 4 Network Slicing Architecture and Entities

Table 2 Distinctions, Possibilities, and Synergies of NFV, SDN, and NS

	NFV	SDN	NS
Concept	Decouples network functions from the hardware on which they are built and delivers them as VNFs	Decouples network control from the forwarding devices and makes the forwarding devices programmable	Abstracts network functionalities from hardware and software components as slices to provide specific services
Control	MANO orchestrates and manages all VNFs via a framework	SDN controller supervises traffic flow and controls the devices via application plane instructions	NSO manages the entire network service life-cycle via the support of SDN and NFV
Virtualization level	Network functions and applications	Packet flow	Network service function chaining (SFC)
Deployment in 5G	Different functions in gNBs, such as virtual radio resource [Base Band Unit (BBU) processing Pool] and 5GC (UPF, AMF, SMF)	SDN controller in RAN, Core, or both. SDN-enabled BBU pool at the RAN, user plane functions at the core, and inter-networking SDN switches.	NS can be implemented in various domains such as RAN, transport, cloud, and end-to-end.

SURVEY ARTICLE

Independent benefits	NFV enables infrastructure partitioning into multiple logical infrastructures	SDN enables flexible management and control of network operations and innovation in network design.	NS Enables the on-demand configuration of networks to provide tailored network services
	SDN controller can be implemented as VNF using NFV.	SDN enhances the efficiency and flexibility in the allocation of resources to VNFs of NS.	NS enables customization of network services and isolation of NFs
Synergy benefits	Resources can be virtualized using NFV, chained together to provide dedicated services using NS, and allocated using an SDN-optimized resource allocation policy. The trio increases network programmability and support for heterogeneous 5G applications. The synergy enables cross-layer network control, in which a unified controller monitors resource demand and consumption and triggers network reconfiguration for efficient performance. This trio simplifies operational and administrative tasks in even the most complex and diverse networks, enhances network performance, shortens the time to market for new services, and reduces operating costs.		

2.4. Problem Description

The challenge with network slicing lies in the efficient allocation and utilization of resources while accommodating multiple services with different constraints. 5G network operators must efficiently distribute resources to establish network slices while also being able to adjust resources as demands change [41]. However, this becomes complicated due to varying user QoS requirements, and mobility features such as seamless handover and interference management [42], [43], and [44]. Resource sharing among slices also poses a security challenge as different slices may have different security policies [45]. On top of these challenges, finding a balance between resource allocations and sharing and meeting key performance indicators is difficult in a network environment with multiple technical and time-varying elements. Table 3 highlights the major challenges faced by 5G NS.

Table 3 Requirements of Network Slicing, Challenges, and Causes

Requirements	Challenges	Causes
Management	Creation of network granularity policies A large number of rarely used active nodes	High data count
Reliability	Seamless and high-quality connectivity Fast connection recovery The trade-off between performance, security, and flexibility	Level of slice isolation
Scalability	Optimization of resource provisioning and utilization Network robustness Acceleration of re-configurable network slicing Openness, heterogeneity, and interoperability of devices and protocols	Dynamic network characteristics
Efficiency	Reduction of capital and operational expenditures Reduction of computational complexity The trade-off between resource allocation and consumption Reduction of power usage On-demand resource allocation and optimal network resource usage	Level of service customization
Security	security attack detection and mitigation	Resource sharing among slices

SURVEY ARTICLE

3. LITERATURE REVIEW

3.1. Network Slicing Frameworks

Future networks may face an increase in user connection requests or a high number of users in a given area, which could lead to network stress and unavailability [46]. To overcome this challenge, the wireless network architecture needs to be flexible enough to allow for dynamic resource provisioning and optimization in real time. Abdulaziz et al. [47] proposed an SDN-based 5G core architecture to handle scalability and flexibility in the network. The design assumes that the gNBs are connected to an SDN-enabled switch and the SDN controller oversees the whole data plane and implements flow rules in the switches. However, the proposed architecture has not been tested for flexibility and scalability under varying data plane traffic volumes, which is a common scenario in 5G networks. Similarly, in [48], two types of SDN controllers were deployed: core SDN controllers to deploy and manage network functions and flow SDN controllers to optimize backhaul network traffic flow. Thomas et al. presented an E2E network slice architectural design and proof-of-concept implementation for future networks based on virtualized multi-region infrastructures [49]. This design enforces network QoS via SDN to deliver custom-tailored slices. However, the authors only demonstrated the time required to create and scale a network slice on the testbed and did not provide an analysis of placement and resource allocation techniques. In [50], a fine-grained network slicing architecture was introduced, consisting of physical, NFV, and slice layers to enable virtual instance monitoring, selection, and deployment. However, this framework only depicts the core network and does not fully represent the 5G network architecture specified by the 3GPP and ETSI standards. The authors in [51] created an AI-based model for cross-slice admission and congestion control to maximize operator profitability. The cross-slice admission control balances resource consumption and the dropping probability of high-priority slices, but the congestion control function limits slice request dropping by reducing resources assigned to low-priority slices during system overload to admit high-priority slices.

3.2. Dynamic Resource Provisioning For 5G Network Slicing

The resources of radio access, transport, and core network can be represented as computing and bandwidth resources. The problem of provisioning these resources can be reduced to a Virtual Network Embedding (VNE) problem, which has different objectives depending on the type of slice and user distribution. The VNE has three challenges: mapping virtual nodes (VNFs) to physical nodes, mapping virtual links between virtual and physical nodes, and updating previous mappings [52]. The VNE problem often results in a Mixed Integer Linear Programming (MILP) problem, which is known to be NP-hard [53], so various heuristics such as exact algorithms, heuristics, and meta-heuristics are used to solve it.

Jalal et al. [54] proposed a new method for end-to-end resource provisioning that optimizes tenant satisfaction and InP operating expenditures while avoiding the complexity of solving the MILP problem. The authors created a distributed privacy-saving mechanism to eliminate the need for data centers or access points to share resource capacity, but this increased signaling overhead. The algorithm was implemented in distributed slice managers, which could result in a suboptimal global solution. Ruihan et al. [55] investigated robust network slicing mechanisms to address concerns about bugs in VNFs and changes in traffic demands, using a heuristic algorithm based on variable neighborhood search, but this method resulted in a high signal overhead. The authors assumed a fixed number of VNFs for each slice, but in practice, only user requirements are known. Thus, the virtual network embedding process is iteratively tested until an appropriate number of VNFs is identified. In related studies, [56]–[58] studied NS determination and embedding without prior knowledge of virtual network topology and resources, and developed a heuristic technique based on user distribution and requirements. However, the authors did not optimize physical resource sharing to meet the SLA criteria of multiple slices and did not consider reconfiguration and resource reallocation in a dynamic environment, or use 5G service-based architecture and network KPIs to evaluate the proposed solution. Afolabi et al. [46] designed a resource dimensioning heuristic algorithm to determine the computational and virtual resources allotted to the Network Slicing Orchestration System (NSOS) to maximize its response time for a particular workload. The proposed solution only allows the NSOS to modify its resources based on future workload to maintain a response time under the delay threshold. As service demand increases, the algorithm blocks excessive Slice Orchestration Requests (SORs) and when demand decreases, some resources are left idle. The authors only evaluated the delay experienced by SOR through different NSO entities. However, iterative methods for resource estimation for each slice are computationally intensive, as noted in [59].

Additionally, resource allocation across users in a slice is not necessarily equitable. Alfoudi et al [60] proposed an NS Resource Management (NSRM) technique for allocating resources in an LTE network slice, which guarantees isolation and equitable distribution of bandwidth among users. However, the approach adopted a uniform resource distribution and fixed interval physical resource redistribution and used an exponential smoothing model with only two outcomes. This model doesn't account for excess resources in a slice, and fixed interval redistribution may result in over- or under-provisioning due to uncertainty in demand. For instance, Shi et al. in [61] developed an algorithm that uses Q-learning to allocate resources based on user satisfaction. However, the algorithm does not account for unused resources and cannot adjust resources before the end of a service's lifetime. To enhance service quality and resource utilization, the authors in [62] presented a novel resource allocation strategy that considers diverse service qualities. The authors developed a computational load distribution method to balance workloads based on user association limits, and tested the algorithm on a static slicing strategy and the impact of resource allocation interval. Lowering the resource allocation period may result in less resource granularity and decreased performance. Luu et al. [57] addressed this issue by proposing a resource provisioning strategy for network slicing robustness to variable user demands. However, the proposed algorithm was unable to adapt to network and resource changes, preventing the InP from updating resources over the slice lifespan. As a result, the authors in [29], [63] proposed preemptive, data-driven, automated methods for efficient network resource consumption. The technique in [29] forecasts future capacity demands and promptly reallocates resources. However, the prediction may not accurately reflect the network state as the algorithm was implemented at the data center with substantial latency from the base station.

3.3. Traditional Resource Optimization Algorithms

In traditional business models, MNOs solely provide all network services to end users and thus have complete prior knowledge of service requests and cost/revenue models for each slice. To optimize the slices in real-time, the MNOs require low-cost computational models that are easy to solve to maximize total network utility under constrained resources using utility efficiency [64]. Borylo et al. [65] proposed a multi-objective optimization model for assigning slice resources in the cloud along with a time-efficient heuristic for finding the optimal solution with the lowest computing cost. These models jointly optimize energy and latency and were used to study various scenarios to determine the best resource provisioning methods for the 5G architecture and network slicing paradigm.

The authors found that simpler network topologies provided closer-to-optimal solutions, but the model only addresses static optimization scenarios and cannot be applied to dynamic network environments such as 5G networks. Francesca et al. [66] proposed a unified mathematical framework for generalizing classical solutions to single and multi-resource allocation issues. This framework addresses only the linear relationship between user satisfaction and system efficiency objectives by maximizing a fairness goal function based on user satisfaction by aggregating information about user requests and available resources to various degrees of fairness.

However, these models do not apply to tenant-run slices controlled by mobile virtual network operators (MVNOs), which are essential in 5G networks. Tenants, such as utility or automotive firms and Over-The-Top (OTT) service providers are third-party service providers who do not own network infrastructure [67]. MNOs offer them networking and computing resources. In such cases, a new agreement between the tenant and MNO may be necessary to flexibly reconfigure or cancel a slice at any time, incurring additional operating costs. Alternatively, MNO can provide resources for the implementation of various types of slices and optimize its resource allocation on a broad scale by accepting or rejecting every tenant request for slice creation [68].

However, because each slice is conceptually independent of the others, a tenant cannot access other tenants' slices and can only manage its resources by either requesting additional slices or canceling existing ones. This makes it difficult for the tenants or MNO to jointly optimize all slice resources in a genuinely dynamic manner, rendering most traditional algorithms ineffective and creating a new problem for network resource management.

Yang et al. [69] investigated how tenants modify slice specifications to reduce costs while preserving service quality and proposed a data-driven vector graph and a balanced slice reconfiguration method using the random walk technique to transfer resource status to a two-dimensional vector graph. The authors only considered links as physical resources. Han et al. [70] considered the MNO's perspective and proposed an online genetic approach to optimizing the slicing strategy to maximize total income by encoding each viable slicing approach into a binary format capable of dealing with inter-slice control mechanisms based on binary decisions. However, this model has a poor convergence rate and has not been validated for slice formation delay.

Authors [71] proposed a Multi-Objective (MO) technique for optimizing resource orchestration in a 5G network's cloud-based slices. MO optimization is efficient in solving economic and engineering problems and has been applied to various 5G network problems, including energy consumption at base stations, VNF orchestration and chaining, and resource orchestration and management in 5G network slicing [72]. However, the authors used synthetic data to analyze the algorithm and the spatial distribution of 5G users was not considered. The authors [73] designed and tested a heuristic technique to reduce runtime and enable rapid decision-making, as well as a MILP method to optimize cross-domain NS deployment independent of the underlying topologies. The proposed approach did not take mobile users or real-world scenarios into account. Alaa et al. [74] created a distributed method for optimizing radio access point selection, resource assignment, and data routing to determine the best trade-off between various performance metrics and to provide an efficient way to tradeoff between complexity and optimality.

3.4. Machine Learning-Based Resource Optimization Algorithms

Machine learning (ML) is a set of techniques that enable computers to learn, automate, and optimize decisions from huge data sets [75]. Recently, various ML techniques have been developed for optimizing resource utilization in 5G network slicing [76]–[78]. These techniques can be divided into three: supervised learning, unsupervised learning, and Reinforcement Learning (RL), along with variants such as Q-learning, Markov models, and deep learning [79]–[82].

Deep Learning (DL) is capable of providing accurate predictions for resource scheduling but requires massive datasets. Reinforcement learning can quickly adapt to dynamic environments but is not effective at the beginning of a learning process. These two algorithms can be used together to optimize resource allocation in RAN slicing, with DL handling large-scale resource allocation and RL handling small-scale network dynamics through online resource scheduling [83]. However, this approach may result in periodic traffic prediction and significant signaling overhead because DL is used to estimate traffic volume in each prediction window and RL is used to accomplish online resource scheduling.

The development of low-complexity traffic prediction ML models is ongoing. Chergui et al. [84] described a low-complexity network slice traffic predictor that uses a soft-gated recurrent unit (GRU). The model was trained on live network data to estimate resource provisioning based on traffic per slice. The predicted traffic was then fed into various DL models to perform end-to-end resource allocation and optimization to meet SLAs. However, the convergence rate of the proposed technique was reported to be low. The authors of [85] developed a two-step algorithm for tenants to identify the minimum number of resources required for each VNF in a service chain to meet a specific end-to-end latency and a fast search algorithm for adjusting the network slice to guarantee QoS during changes in traffic demand.

A low-cost auto-resizing approach for adjusting slice size with traffic fluctuation was also demonstrated. Gharehgoli et al. [86] investigated resource optimization in end-to-end NS and defined the utility function of the InP using non-convex mixed-integer non-linear programming to reflect the difference between revenue and cost. The intricacies of this formulation were addressed through the use of deep RL techniques to maximize the utility of the MNO by considering numerous actions and states.

According to Anuar et al. [87], the efficient management of network function deployment in a hybrid cloud infrastructure, consisting of central and edge computing resources, is crucial to easily create ancillary slices by deploying distributed mission-critical services (VNFs) in the edge cloud to ensure that critical slices continue to be served even in case of a sudden increase in traffic flow. The authors then studied VNF deployment in 5G NS with a focus on mission-critical communication and used a stochastic Markov decision process that comprised states, actions, state transitions, and a reward function to model the problem. A reinforcement learning-based technique that could automatically find a near-optimal deployment solution while minimizing logical network delays and costs was developed. Table 4 provides a summary of the open issues/limitations of the existing research.

Table 4 Open Issues/Limitations of Existing Research

Ref.	Problem	Solution Approach	Open Issues/Limitations
[49]	Future networks are vulnerable to a spike in user connection requests or a burst in users per unit area; thus, the wireless system must dynamically adjust, particularly for essential services.	Demonstrated the architectural design and proof-of-concept implementation of an NS framework for future networks based on virtualized multi-region infrastructures to deliver customized slices.	On the testbed, the authors only demonstrated the time it takes to create and scale a network slice. There was no analysis of placement and resource allocation algorithms provided.
[52]	Several techniques including precise, heuristics, and meta-heuristics, can be used to optimize resources in NS.	Tested the proposed algorithm via Global Resource Capacity, Monte Carlo Tree Search, and RL under profit for InP and virtual resource mapping computational complexity.	Finding the balance between complexity and KPI improvement, enabled by dynamic resource allocation and sharing,

SURVEY ARTICLE

[54]	The VNE problem often leads to MILP, which is NP-hard, necessitating various heuristic methods to solve it.	Presented a novel formulation for E2E resource provisioning that avoids the intractable complexity of solving a MILP	Pushing the optimization algorithm to the slice manager reduces global solution convergence and may result in a suboptimal solution.
[55]	VNF bugs may invalidate some slices, and dramatic changes in traffic demands may necessitate slice reconfiguration or slice recovery processes. Existing work does not adequately address these challenges.	Investigated a unified architecture for slice recovery and reconfiguration in a robust network slicing strategy.	The authors utilized a variable neighborhood search heuristic approach with a low convergence rate and a significant signal overhead. They assumed that VNFs are only assigned to one network slice.
[56]	Iterative trials of VNF embedding are performed until the correct number of VNFs is found.	Investigated the determination and embedding of NSs without prior knowledge of the topology of the VN and resource allocation.	Dynamic SLA and mobility were not considered. The 5G service-based architecture and KPIs were not used for the evaluation.
[57]	Existing resource allocation algorithms do not consider demand uncertainties during each time interval.	Investigated a resource provisioning technique to increase network slicing reliability for a partially unknown number of users whose resource requirements are unpredictable.	InP could not update some slices' already-provisioned resources during the slice lifespan.
[60]	Current NS methods do not guarantee the fair distribution of resources across users in a slice.	Investigated an NSRM technique for assigning the necessary resources to each slice in an LTE network. NSRM isolated users on the same slice and fairly shared distributed bandwidth.	The mechanism only ensures that resources are distributed uniformly; thus, it does not consider differentiated resource requirements.
[61]	Existing research employs myopic algorithms that allocate available resources to maximize only the current utility of network resources without regard to future changes.	Proposed a policy learning algorithm for determining which action (resource assignment) to do for the gNodeB across a time horizon in a given state (available resources and requests).	Even if resources are allocated but not used, they cannot be adjusted before the service lifetime. The reward function focuses on only user satisfaction.
[29]	Existing mobile traffic ML predictors usually function at the BS level, but NFV operations mostly take place in datacenters to manage tens to thousands of BSs.	Proposed a cost-effective, proactive, data-driven and automated solution for network resource utilization that anticipates future capacity needs and allocates resources dynamically.	Predicted results may not reflect the present state of the network since it is done at the datacenter where there is a significant delay from the base station.
[69]	The challenge of updating the status of slices grows as the network size or user demand increases, however, current rigid NS solutions prevent joint optimization of resources in 5GC and 5G RAN	Developed a data-driven vector network that allows for quick updates on the status of slices and an inbuilt balanced method for reconfiguring them.	The authors considered only links as physical resources to evaluate the proposed algorithm.
[83]	DL provides accurate	Use both approaches for	Time was split into prediction

SURVEY ARTICLE

	predictions in RAN slicing but only with large datasets, whereas RL can respond quickly to a dynamic environment but with an extensive learning process.	collaborative RAN slicing. For dealing with small-scale network dynamics, DL tackles large-scale resource allocation and RL addresses online resource scheduling.	windows (PWs), each with DL and RL, resulting in periodic traffic prediction and signaling overhead.
[88]	Current DRP systems are either limited to specific scenarios or unable to deliver a low-complexity, combined sizing of all entities while maintaining system stability simultaneously.	Developed a resource dimensioning heuristic approach to identify the necessary computational and virtual resources to allocate to NSOS for a specific workload, with the aim of attaining the maximum response time for NSOS.	Excessive SORs are blocked if service demand increases. Similarly, as service demand falls, resource utilization falls.

4. DISCUSSION, OPEN ISSUES, AND FUTURE RESEARCH DIRECTIONS

This paper aims to uncover some major issues in existing research dealing with resource allocation and optimization in 5G network slicing. Based on a thorough examination of the state-of-the-art literature, the following key unresolved issues, which represent some of the important themes for future study in the overall domain of resource management in 5G NS are identified and elaborated.

4.1. Network Slicing Framework

4.1.1. Slicing Strategy

The 5G network slice design adopts a common framework referred to as Network Slice Instance (NSI), which divides the network into multiple slices. An NSI includes network functions from the CN and RAN, along with all the necessary functionalities and resources, to support a specific set of communication services. NSI is managed by a Network Slice Subnet Instance (NSSI) while providing communication services [37]. This framework has mainly been adopted to provide greater flexibility for dynamic slices that handle heterogeneous network demands. With this framework, network slicing (NS) can be applied to various domains within the network, producing different types of slices [89]. NS in the CN only involves the allocation of cloud resources, which can easily be reduced to the problem of assigning resources to VM [90]. This makes slicing in the core more efficient, but it decreases service customization and resource isolation since all services and tenants share a significant portion of the network. Personalized slicing that provides customized functionalities requires dedicated radio access resources via cloud RAN (C-RAN) paradigms, allowing for the customization of scheduling algorithms [91]. However, this approach increases system complexity and restricts the possibilities for radio resource sharing or dynamic reconfiguration of virtual network functions (VNFs). As a result, operators must choose between customization of services, efficiency in resource management, and the complexity of the system, or find a compromise among these options. However, the trade-offs and ramifications of these choices are unaddressed in the existing literature.

4.1.2. Reconfigurable Slicing

The frequency of reconfiguring the VNFs in NS is a crucial aspect to consider. It is limited in reality by the technological limitations of the slicing method used. For instance, slicing at the antenna level can take several minutes to switch the radio-frequency front end and reset the transport network. In radio resource management, reconfigurations may be restricted by signaling overhead and fixed reconfigurations can lead to under or overutilization of resources. To overcome these challenges, dynamic spectrum slicing can be utilized to adjust to changes in network traffic. In [92], a policy-based dynamic spectrum slicing system that takes network traffic variations into account was introduced. The inter-arrival time of user traffic was modeled using a Markov process, which accounts for correlations between consecutive user traffic points, enabling more accurate predictions. In the case of VM orchestration, instantiation, and migration delays limit the timescale, and therefore, further research is needed to determine the optimal time steps or frequency for resource reallocation (reconfiguration interval) for resource provisioning algorithms to help network operators.

4.1.3. Security in Network Slicing

Dynamic network slicing and network resource sharing can present security challenges. This is because the security policies for network slices that provide different services may vary, and security becomes more complex with network slicing in a multi-domain architecture and multi-tenancy schemes. Additionally, the various technologies, including SDN and NFV, used in the different layers of the NS framework and SLAs between NS entities pose significant and complex security challenges. These challenges emanate from aspects like security across multiple domains, security of the slicing orchestrator, security between slices, and security within a slice. Moreover, different 5G applications have different latency requirements, and URLLC services have stricter latency requirements compared to eMBB. As a result, URLLC slice authentication must be simple and lightweight. A feasible solution is to design efficient adaptive security mechanisms and deploy them in an intelligent slicing framework.

4.1.4. Intelligent Slicing/Service Function Chaining

This is a method of implementing various network functions in a VM running on one or more nodes in one or more domains [53]. Intelligent SFC aims to optimize customization, efficiency, and complexity, or enhance one over the others while addressing security issues. However, network slicing is not just a technology, it is also a commercial strategy involving various actors delivering services to users. The optimal way of mapping the business demands of tenants and customers to service providers' infrastructure has not been adequately characterized, affecting the assignment and management of slice resources and optimization. Adequate characterization of customers' business requirements into Service Level Agreements (SLAs) between network slicing parties is necessary for adaptive business-model-driven network services and easy SLA fulfillment. The amount of system automation required to minimize manual effort and human involvement is also an open area for investigation. Additionally, the placement of the SDN controller has implications for network scalability, network performance, and network reliability. Placing the controller in the core network can lead to increased network performance and scalability, but it can also lead to single points of failure and increased latency [47]. On the other hand, distributing the controller throughout the network domain can enhance network reliability but can also result in increased complexity and decreased network performance [48]. Further investigation is necessary to determine the best placement of the SDN controller in the network slicing framework to strike a balance between these trades-offs and deliver an efficient network slicing solution.

4.2. Resource Provisioning and Optimization

4.2.1. Efficient Provisioning

Effective slice resource provisioning methods must be resilient to changes in resource demands. Network function chains are often deployed in a network slice on a best-effort basis, but this strategy doesn't ensure the availability of sufficient physical resources to meet changing demands. As a result, some researchers have adopted a probabilistic approach that aims to guarantee the satisfaction of resource requirements while being robust against uncertainties [57].

in utilizing Federated learning for resource allocation and optimization in network slicing.

Others, however, have adopted a myopic approach that considers past demands when provisioning current requests using heuristic and meta-heuristic algorithms [29]. The literature suggests that ML approaches produce superior outcomes compared to these methods because ML methods consider both past and future demands before allocating the appropriate number of resources to meet the demand while maximizing the profits of the InP [83]. ML techniques are also better suited for solving large optimization problems than traditional optimization approaches. Traditional approaches, such as starting with an initial solution and progressing to the optimal solution through each iteration [65], are slow but guarantee global convergence for small problems. However, for large problems, convergence depends on the choice of initial approximation. On the other hand, ML approaches, such as RL, minimize the complexity of problems that require stringent constraints, by stating such constraints as rewards or penalties. In network slicing, ML methods can easily categorize slice requests, lease, and release resources, optimize network performance, and reduce the complexity of implementing traditional optimization algorithms and handling massive data analytics and continuous functions. The challenge lies in implementing these techniques with the network InP at the lowest possible cost in processing, latency, and signaling. ML methods often require large amounts of data and the size of the data affects the computational complexity and processing time. For example, if the ML algorithm needs to analyze a large and continuous state space, the time to train the model or the memory requirements will increase. These challenges can be addressed with the use of Federated learning models.

4.2.2. Federated Learning for Resource Provisioning

Federated learning is seen as a breakthrough in natural language processing (NLP) for prediction and resource management [93]. While these models can be developed using traditional ML techniques [76]–[78], the dynamic nature of 5G networks makes the training period required for accurate decisions expensive. Moreover, high-quality data sets are necessary to produce reliable models with valuable insights and accurate judgments. However, these models can be challenging to implement due to the shortage of available data sets and data privacy concerns from existing operators. Traditional ML methods also have privacy and latency issues that can be addressed by using federated learning. This approach eliminates the need to transfer data from end devices to centralized servers, protecting user privacy, reducing decision-making time, and reducing network congestion [93]. The use of Federated learning has been demonstrated in [94] to train double-deep Q-learning agents at the network edge for caching and computational offloading while maintaining user privacy. Despite these developments, there remain challenges

4.2.3. Optimization Under Mobility

SURVEY ARTICLE

The mobility of users in 5G network slicing presents significant challenges for resource allocation and optimization, as outlined in [42], [43]. These challenges stem from the handovers between different access networks with various radio access technologies, the high density of 5G networks, and stringent handover requirements in specific 5G applications like autonomous vehicles. These challenges also make it difficult to maintain network slice isolation. Therefore, it's crucial to develop unique on-demand slicing schemes and mobility-aware slicing algorithms to tackle mobility issues. One proposed solution is the Lagrangian dual decomposition-based approach presented in [95]. The authors developed a handover strategy for resource distribution in multiple 5G scenarios in the context of network slicing. Along with the Lagrangian dual decomposition-based solution, which relaxes binary variables into continuous ones, game theoretic approaches can also be explored to tackle the challenges of mobility-aware slicing. Table 5 provides a summary of the open issues and potential areas for future research that were discussed in this section.

Table 5 Open Issues and Future Research Directions

Open Issue	Future Research Direction
Optimal Slicing Strategy	Investigating the trade-offs between customization of services, efficiency in resource management, and complexity of the system.
Reconfigurable slicing	Investigating the optimal frequency of reconfiguration of the VNFs under various network conditions.
Reconfigurable slicing	Designing efficient adaptive security mechanisms to address various security issues in 5G network slicing.
Intelligent Slicing	Characterizing the business requirements to make them suitable for machine learning algorithms.
SDN Controller Placement	Investigating the optimal location of the SDN controller to ensure efficient operation of the network.
Efficient Provisioning and Optimization	Designing computationally efficient machine learning algorithms to ensure optimal allocation and utilization of resources.

5. CONCLUSION

This paper highlights the significance of network slicing in 5G and the need for efficient resource allocation and optimization methods for cost-effectiveness for both network operators and users. The paper presents the current state-of-the-art solutions for 5G network slicing using SDN and NFV which are widely accepted as solutions to the resource management and orchestration challenges in 5G NS. It begins with an introduction to 5G network design, followed by a discussion on the enabling technologies of 5G network slicing, SDN, and NFV, and their role in the 5G network slicing. A comprehensive literature review and comparison of different 5G resource management strategies, such as resource allocation and optimization, is also presented, in relation to the 5G network architecture. Finally, the paper highlights persistent research challenges, alternative solutions, and future research directions in 5G network slicing resource management.

ACKNOWLEDGEMENT

This research was funded by the KNUST Engineering Education Project (KEEP), Ghana with project number P164546.

REFERENCES

[1] M. A. Siddiqi, H. Yu, and J. Joung, “5G ultra-reliable low-latency communication implementation challenges and operational issues with IoT devices,” Electronics (Switzerland), vol. 8, no. 9. pp. 1–18, 2019. doi: 10.3390/electronics8090981.

[2] G. Kalem, O. Vayvay, B. Sennaroglu, and H. Tozan, “Technology Forecasting in the Mobile Telecommunication Industry: A Case Study Towards the 5G Era,” Engineering Management Journal, vol. 33, no. 1, pp. 15–29, Jan. 2021, doi: 10.1080/10429247.2020.1764833.

[3] R. K. Gupta Akhil Jha, “A Survey of 5G Network : Architecture and Emerging Technologies,” IEEE Access, vol. 3, p. 27, 2015.

[4] NGMN Alliance, “Description of Network Slicing Concept by NGMN Alliance,” Ngmn 5G P1, vol. 1, no. September, p. 19, 2016, [Online]. Available: https://www.ngmn.org/uploads/media/160113_Network_Slicing_v1_0.pdf

[5] 5GPPP, “View on 5G Architecture,” 5G Architecture White Paper, no. February, p. 182, 2020, doi: 10.5281/zenodo.3265031.

[6] T. Cisco and A. Internet, “Cisco: 2020 CISO Benchmark Report,” Computer Fraud & Security, vol. 2020, no. 3, pp. 4–4, Jan. 2020, doi: 10.1016/S1361-3723(20)30026-9.

SURVEY ARTICLE

[7] K. Cengiz and M. Aydemir, “Next-Generation infrastructure and technology issues in 5G systems,” Journal of Communications Software and Systems, vol. 14, no. 1, pp. 33–39, 2018, doi: 10.24138/jcomss.v14i1.422.

[8] J. Long and O. Büyüköztürk, “Collaborative duty cycling strategies in energy harvesting sensor networks,” Computer-Aided Civil and Infrastructure Engineering, vol. 35, no. 6, pp. 534–548, Jun. 2020, doi: 10.1111/mice.12522.

[9] N. Salhab, R. Langar, and R. Rahim, “5G network slices resource orchestration using Machine Learning techniques,” Computer Networks, vol. 188, no. August 2020, p. 107829, 2021, doi: 10.1016/j.comnet.2021.107829.

[10] A. Gausseran, “Optimization algorithms for Network Slicing for 5G,” 2021.

[11] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, “5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges,” Computer Networks, vol. 167, no. 2020, 2020, doi: 10.1016/j.comnet.2019.106984.

[12] L. Bonati, M. Polese, S. D’Oro, S. Basagni, and T. Melodia, “Open, Programmable, and Virtualized 5G Networks: State-of-the-Art and the Road Ahead,” Computer Networks, vol. 182, no. July. Elsevier B.V., p. 107516, 2020. doi: 10.1016/j.comnet.2020.107516.

[13] M. O. Ojijo and O. E. Falowo, “a Survey on Slice admission Control Strategies and Optimization Schemes in 5G Network,” IEEE Access, vol. 8, pp. 14977–14990, 2020, doi: 10.1109/aCCESS.2020.2967626.

[14] W. Ejaz, S. K. Sharma, S. Saadat, M. Naeem, A. Anpalagan, and N. A. Chughtai, “A comprehensive survey on resource allocation for CRAN in 5G and beyond networks,” Journal of Network and Computer Applications, vol. 160, no. March. Elsevier Ltd, p. 102638, 2020. doi: 10.1016/j.jnca.2020.102638.

[15] A. Mughees, M. Tahir, M. A. Sheikh, and A. Ahad, “Towards energy efficient 5G networks using machine learning: Taxonomy, research challenges, and future research directions,” IEEE Access, vol. 8, pp. 187498–187522, 2020, doi: 10.1109/ACCESS.2020.3029903.

[16] N. Slamnik-Kriještorac, H. Kremo, M. Ruffini, and J. M. Marquez-Barja, “Sharing Distributed and Heterogeneous Resources toward End-to-End 5G Networks: A Comprehensive Survey and a Taxonomy,” IEEE Communications Surveys and Tutorials, vol. 22, no. 3, pp. 1592–1628, 2020, doi: 10.1109/COMST.2020.3003818.

[17] O. Idowu-bismark, O. Kennedy, R. Husbands, and M. Adedokun, “5G Wireless Communication Network Architecture and Its Key Enabling Technologies,” vol. 12, no. April, pp. 70–82, 2019.

[18] P. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, “Design considerations for a 5G network architecture,” IEEE Communications Magazine, vol. 52, no. 11, pp. 65–75, Nov. 2014, doi: 10.1109/MCOM.2014.6957145.

[19] K. M. S. Huq, S. A. Busari, J. Rodriguez, V. Frascolla, W. Bazzi, and D. C. Sicker, “Terahertz-Enabled Wireless System for Beyond-5G Ultra-Fast Networks: A Brief Survey,” IEEE Netw, vol. 33, no. 4, pp. 89–95, 2019, doi: 10.1109/MNET.2019.1800430.

[20] O. O. Erunkulu, “5G Mobile Communication Applications : A Survey and Comparison of Use Cases,” IEEE Access, vol. 9, pp. 97251–97295, 2021, doi: 10.1109/ACCESS.2021.3093213.

[21] M. Vaezi and Y. Zhang, “Virtualization and Cloud Computing,” Wireless Networks (United Kingdom), pp. 11–31, 2017, doi: 10.1007/978-3-319-54496-0_2.

[22] G. Brown, “Service-Oriented 5G Core Networks,” Huawei, no. February, p. 12, 2017, [Online]. Available: www-file.huawei.com/-/media/CORPORATE/PDF/white paper/Heavy Reading Whitepaper-Service-Oriented 5G Core Networks.pdf

[23] K. V. Cardoso, C. B. Both, L. R. Prade, C. J. A. Macedo, and V. H. L. Lopes, “A softwarized perspective of the 5G networks,” Jun. 2020, Accessed: Jul. 16, 2022. [Online]. Available: http://arxiv.org/abs/2006.10409

[24] V. Thirupathi, C. Sandeep, S. Naresh Kumar, and P. Pramod Kumar, “A comprehensive review on sdn architecture, applications and major benifits of SDN,” International Journal of Advanced Science and Technology, vol. 28, no. 20, pp. 607–614, 2019.

[25] S. H. Haji et al., “Comparison of Software Defined Networking with Traditional Networking,” Asian Journal of Research in Computer Science, no. May, pp. 1–18, 2021, doi: 10.9734/ajrcos/2021/v9i230216.

[26] T. Bakhshi, “State of the art and recent research advances in software defined networking,” Wirel Commun Mob Comput, vol. 2017, p. 36, 2017, doi: 10.1155/2017/7191647.

[27] S. Yadav and S. Singh, “SDN and NFV in 5G : Advancements and Challenges,” International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology ISSN, vol. 7, no. 5, pp. 101–110, 2018, [Online]. Available: www.ijcsmc.com%0AInternational

[28] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, “SoftCell : Scalable and Flexible Cellular Core Network Architecture Categories and Subject Descriptors,” no. 163, pp. 163–174.

[29] D. Bega, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, “DeepCog: Optimizing Resource Provisioning in Network Slicing with AI-Based Capacity Forecasting,” IEEE Journal on Selected Areas in Communications, vol. 38, no. 2, pp. 361–376, 2020, doi: 10.1109/JSAC.2019.2959245.

[30] S. van Rossem et al., “Deploying elastic routing capability in an SDN/NFV-enabled environment,” 2015 IEEE Conference on Network Function Virtualization and Software Defined Network, NFV-SDN 2015, pp. 22–24, 2016, doi: 10.1109/NFV-SDN.2015.7387398.

[31] A. O. Nyanteh, M. Li, M. F. Abbod, and H. Al-Raweshidy, “CloudSimHypervisor: Modeling and Simulating Network Slicing in Software-Defined Cloud Networks,” IEEE Access, vol. 9, pp. 72484–72498, 2021, doi: 10.1109/ACCESS.2021.3079501.

[32] ETSI, “Network Functions Virtualisation (NFV) Release 3 ; Evolution and Ecosystem ; Report on Network Slicing Support with ETSI NFV Architecture Framework,” 2017. [Online]. Available: https://ipr.etsi.org/

[33] M. Chahbar, G. Diaz, A. Dandoush, C. Cerin, and K. Ghoumid, “A Comprehensive Survey on the E2E 5G Network Slicing Model,” IEEE Transactions on Network and Service Management, vol. 18, no. 1, pp. 49–62, 2021, doi: 10.1109/TNSM.2020.3044626.

[34] B. NGMN Alliance, R. el Hattachi, and J. Erfanian, “NGMN 5G White Paper,” 2015.

[35] D. Ficzere, “Complex network theory to model 5G Network Slicing,” pp. 19–22, 2022.

[36] A. Papageorgiou, A. Fernández-Fernández, S. Siddiqui, and G. Carrozzo, “On 5G network slice modelling: Service-, resource-, or deployment-driven?,” Comput Commun, vol. 149, no. June 2019, pp. 232–240, 2020, doi: 10.1016/j.comcom.2019.10.024.

[37] T. Specification and G. Services, “3gpp ts 28.530,” 2021.

[38] X. Li et al., “Network Slicing for 5G: Challenges and Opportunities,” IEEE Internet Comput, vol. 21, no. 5, 2018, doi: 10.1109/MIC.2018.326150452.

[39] J. Mei, “An intelligent self-sustained RAN slicing framework for diverse service provisioning in 5G-beyond and 6G network s A pattern recognition framework for detecting changes Triboelectric nanogenerators enabled internet of things : A survey Network Vehicles :,” vol. 1, no. 3, pp. 281–294, 2020, doi: 10.23919/ICN.2020.0019.

[40] O. Mauricio, C. Rendon, and S. Member, “Scalability and Performance Analysis in 5G Core Network slicing,” vol. 8, 2020, doi: 10.1109/ACCESS.2020.3013597.

[41] Z. Kotulski, T. W. Nowak, M. Sepczuk, and M. A. Tunia, “5G networks: Types of isolation and their parameters in RAN and CN slices,” Computer Networks, vol. 171, 2020, doi: 10.1016/j.comnet.2020.107135.

[42] R. Singh et al., “Analysis of Network Slicing for Management of 5G Networks Using Machine Learning Techniques,” vol. 2022, 2022.

SURVEY ARTICLE

[43] R. A. Addad, G. S. Member, D. Leonel, and C. Dutra, "Toward Using Reinforcement Learning for Trigger Selection in Network Slice Mobility," vol. 39, no. 7, pp. 2241–2253, 2021.

[44] S. Sridharan, "A Literature Review of Network Function Virtualization (NFV) in 5G Networks," vol. 68, no. 10, pp. 49–55, 2020, doi: 10.14445/22312803/IJCTT-V68I10P109.

[45] Z. Kotulski, T. W. Nowak, M. Sepczuk, and M. A. Tunia, "5G networks : Types of isolation and their parameters in RAN and CN slices," vol. 171, 2020, doi: 10.1016/j.comnet.2020.107135.

[46] and J.-L. G. Mat´ıas Richart, Javier Baliosian, Joan Serrat, "End-to-end network slicing enabled through network function virtualization," in 2017 IEEE Conference on Standards for Communications and Networking (CSCN), Sep. 2017, pp. 30–35. doi: 10.1109/CSCN.2017.8088594.

[47] A. Abdulghaffar and A. Mahmoud, "Modeling and Evaluation of Software Defined Networking Based 5G Core Network Architecture," pp. 10179–10198, 2021, doi: 10.1109/ACCESS.2021.3049945.

[48] L. Ma, X. Wen, L. Wang, Z. Lu, and R. Knopp, "An SDN / NFV Based Framework for Management and Deployment of Service Based 5G Core Network," no. October, pp. 86–98, 2018.

[49] T. Lin, S. Marinova, and A. Leon-Garcia, "Towards an end-to-end network slicing framework in multi-region infrastructures," in Proceedings of the 2020 IEEE Conference on Network Softwarization: Bridging the Gap Between AI and Network Softwarization, NetSoft 2020, 2020, pp. 413–421. doi: 10.1109/NetSoft48620.2020.9165408.

[50] R. A. Addad, M. Bagaa, T. Taleb, D. Leonel, and C. Dutra, "Optimization Model for Cross-Domain Network Slices in 5G Networks," vol. 19, no. 5, pp. 1156–1169, 2020.

[51] G. Dandachi, A. de Domenico, and D. T. Hoang, "An Artificial Intelligence Framework for Slice Deployment and Orchestration in 5G Networks," vol. 6, no. 2, pp. 858–871, 2020.

[52] D. Irawan, N. R. Syambas, A. A. N. Ananda Kusuma, and E. Mulyana, "Network Slicing Algorithms Case Study:Virtual Network Embedding," in 2020 14th International Conference on Telecommunication Systems, Services, and Applications (TSSA, Nov. 2020, pp. 1–5. doi: 10.1109/TSSA51342.2020.9310856.

[53] R. A. Addad, M. Bagaa, T. Taleb, D. Leonel, and C. Dutra, "Optimization Model for Cross-Domain Network Slices in 5G Networks," vol. 19, no. 5, pp. 1156–1169, 2020.

[54] J. Khamse-Ashari, G. Senarath, I. Bor-Yaliniz, and H. Yanikomeroglu, "An agile and distributed mechanism for inter-domain network slicing in next-generation mobile networks," IEEE Trans Mob Comput, 2021, doi: 10.1109/TMC.2021.3061613.

[55] R. Wen et al., "On robustness of network slicing for next-generation mobile networks," IEEE Transactions on Communications, vol. 67, no. 1, pp. 430–444, 2019, doi: 10.1109/TCOMM.2018.2868652.

[56] Q. T. Luu, S. Kerboeuf, A. Mouradian, and M. Kieffer, "A Coverage-Aware Resource Provisioning Method for Network Slicing," IEEE/ACM Transactions on Networking, vol. 28, no. 6, pp. 2393–2406, 2020, doi: 10.1109/TNET.2020.3019098.

[57] Q. Luu, S. Kerboeuf, M. Kieffer, and S. Member, "Uncertainty-Aware Resource Provisioning for Network Slicing," vol. 18, no. 1, pp. 79–93, 2021.

[58] Q.-T. Luu, S. Kerboeuf, and M. Kieffer, "Foresighted Resource Provisioning for Network Slicing," in 2021 IEEE 22nd International Conference on High Performance Switching and Routing (HPSR), Jun. 2021, vol. 2021-June, no. 1, pp. 1–8. doi: 10.1109/HPSR52026.2021.9481832.

[59] N. Kazemifard and V. Shah-Mansouri, "Minimum delay function placement and resource allocation for Open RAN (O-RAN) 5G networks," Computer Networks, vol. 188, no. May 2020, p. 107809, 2021, doi: 10.1016/j.comnet.2021.107809.

[60] A. S. D. Alfoudi, S. H. S. Newaz, A. Otebolaku, G. M. Lee, and R. Pereira, "An Efficient Resource Management Mechanism for Network Slicing in a LTE Network," IEEE Access, vol. 7, pp. 89441–89457, 2019, doi: 10.1109/ACCESS.2019.2926446.

[61] Y. Shi, Y. E. Sagduyu, and T. Erpek, "Reinforcement Learning for Dynamic Resource Optimization in 5G Radio Access Network Slicing," 2020.

[62] S. M. A. Kazmi and R. Hussain, "Effects of Differentiated 5G Services on Computational and Radio Resource Allocation Performance," vol. 18, no. 2, pp. 2226–2241, 2021.

[63] D. Bega, M. Gramaglia, A. Garcia-Saavedra, M. Fiore, A. Banchs, and X. Costa-Perez, "Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management," IEEE Communications Magazine, vol. 58, no. 6, pp. 32–38, 2020, doi: 10.1109/MCOM.001.1900653.

[64] Z. Wang, Y. Wei, F. Richard Yu, and Z. Han, "Utility Optimization for Resource Allocation in Multi-Access Edge Network Slicing: A Twin-Actor Deep Deterministic Policy Gradient Approach," IEEE Trans Wirel Commun, pp. 1–14, 2022, doi: 10.1109/TWC.2022.3143949.

[65] P. Borylo, M. Tornatore, P. Jaglarz, N. Shahriar, P. Chołda, and R. Boutaba, "Latency and energy-aware provisioning of network slices in cloud networks," Comput Commun, vol. 157, no. October 2019, pp. 1–19, 2020, doi: 10.1016/j.comcom.2020.03.050.

[66] F. Fossati, S. Moretti, P. Perny, S. Secci, and S. Member, "Multi-Resource Allocation for Network Slicing," vol. 28, no. 3, pp. 1311–1324, 2020.

[67] Y. Li et al., "Understanding the ecosystem and addressing the fundamental concerns of commercial MVNO," IEEE/ACM Transactions on Networking, vol. 28, no. 3, pp. 1364–1377, 2020, doi: 10.1109/TNET.2020.2981514.

[68] M. Gharbaoui, B. Martini, and P. Castoldi, "Programmable and Automated Deployment of Tenant-Managed SDN Network Slices," Proceedings of IEEE/IFIP Network Operations and Management Symposium 2020: Management in the Age of Softwarization and Artificial Intelligence, NOMS 2020, 2020, doi: 10.1109/NOMS47738.2020.9110302.

[69] H. Yang et al., "Data-Driven Network Slicing from Core to RAN for 5G Broadcasting Services," IEEE Transactions on Broadcasting, vol. 67, no. 1, pp. 23–32, 2021, doi: 10.1109/TBC.2020.3031742.

[70] B. I. N. Han, J. I. Lianghai, and S. Member, "Slice as an Evolutionary Service : Genetic Optimization for Inter-Slice Resource Management in 5G Networks," IEEE Access, vol. 6, pp. 33137–33147, 2020, doi: 10.1109/ACCESS.2018.2846543.

[71] A. Mpatziakas, S. Papadopoulos, A. Drosou, and D. Tzovaras, "Multi-objective Optimisation for Slice-aware Resource Orchestration in 5G Networks," no. Icin, pp. 79–86, 2020.

[72] B. B. Haile and E. Mutafungwa, "A Data-Driven Multiobjective Optimization Framework for Hyperdense 5G Network Planning," pp. 169423–169443, 2020, doi: 10.1109/ACCESS.2020.3023452.

[73] R. A. Addad, M. Bagaa, T. Taleb, D. L. C. Dutra, and H. Flinck, "Optimization model for cross-domain network slices in 5g networks," IEEE Trans Mob Comput, vol. 19, no. 5, pp. 1156–1169, 2020, doi: 10.1109/TMC.2019.2905599.

[74] A. A. Abdellatif, A. Mohamed, A. Erbad, and M. Guizani, "Dynamic Network Slicing and Resource Allocation for 5G-and-Beyond Networks," pp. 262–267, 2022.

[75] H. Fourati, R. Maaloul, and L. Chaari, A survey of 5G network systems : challenges and machine learning approaches, no. 0123456789. Springer Berlin Heidelberg, 2020. doi: 10.1007/s13042-020-01178-4.

[76] L. A. Garrido, A. Dalgkitsis, K. Ramantas, and C. Verikoukis, "Machine Learning for Network Slicing in Future Mobile Networks: Design and Implementation," 2021 IEEE International Mediterranean Conference on Communications and Networking, MeditCom 2021, pp. 23–28, 2021, doi: 10.1109/MeditCom49071.2021.9647571.

[77] H. Chergui and C. Verikoukis, "Big Data for 5G Intelligent Network Slicing Management," no. August, pp. 56–61, 2020.

[78] F. Debbabi, R. Jmal, L. C. Fourati, and A. Ksentini, "Algorithmics and Modeling Aspects of Network Slicing in 5G and Beyonds Network:

Survey," IEEE Access, vol. 8, pp. 162748–162762, 2020, doi: 10.1109/ACCESS.2020.3022162.

[79] D. Yan, X. Yang, and L. Cuthbert, "Regression-based K nearest neighbours for resource allocation in network slicing," Wireless Telecommunications Symposium, vol. 2022-April, 2022, doi: 10.1109/WTS53620.2022.9768174.

[80] Y. Liu, J. Ding, Z. Zhang, and X. Liu, "CLARA : A Constrained Reinforcement Learning Based Resource Allocation Framework for Network Slicing," pp. 1427–1437, 2021.

[81] B. Han, D. Feng, and H. D. Schotten, "A Markov Model of Slice Admission Control," IEEE Networking Letters, vol. 1, no. 1, pp. 2–5, 2018, doi: 10.1109/lnet.2018.2873978.

[82] C. Ssengonzi, O. P. Kogeda, and T. O. Olwal, "A survey of deep reinforcement learning application in 5G and beyond network slicing and virtualization," Array, vol. 14, no. April, p. 100142, 2022, doi: 10.1016/j.array.2022.100142.

[83] M. Yan, G. Feng, J. Zhou, Y. Sun, and Y. Liang, "Intelligent Resource Scheduling for 5G Radio," IEEE Trans Veh Technol, vol. 68, no. 8, pp. 7691–7703, 2019, doi: 10.1109/TVT.2019.2922668.

[84] H. Chergui, C. Verikoukis, and S. Member, "Offline SLA-Constrained Deep Learning for 5G Networks Reliable and Dynamic End-to-End Slicing," vol. 38, no. 2, pp. 350–360, 2020.

[85] Q. Xu, J. Wang, and K. Wu, "Learning-Based Dynamic Resource Provisioning for Network Slicing with Ensured End-to-End Performance Bound," vol. 7, no. 1, pp. 28–41, 2020.

[86] A. Gharehgoli, A. Nouruzi, S. Member, N. Mokari, P. Azmi, and M. R. Javan, "AI-based Robust Resource Allocation in End-to-End Network Slicing under Demand and CSI Uncertainties," no. Ml, pp. 1–18.

[87] A. Othman and N. A. Nayan, "Automated Deployment of Virtual Network Function in 5G Network Slicing Using Deep Reinforcement Learning," vol. 10, pp. 61065–61079, 2022.

[88] I. Afolabi, J. Prados-Garzon, M. Bagaa, T. Taleb, and P. Ameigeiras, "Dynamic resource provisioning of a scalable E2E network slicing orchestration system," IEEE Trans Mob Comput, vol. 19, no. 11, pp. 2594–2608, 2020, doi: 10.1109/TMC.2019.2930059.

[89] C. Marquez et al., "Resource Sharing Efficiency in Network Slicing," IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, vol. 16, no. 3, pp. 909–923, 2019, doi: 10.1109/TNSM.2019.2923265.

[90] S. T. Arzo, R. Bassoli, F. Granelli, S. Member, F. H. P. Fitzek, and S. Member, "Study of Virtual Network Function Placement in 5G Cloud Radio Access Network," vol. 17, no. 4, pp. 2242–2259, 2020.

[91] M. Maule, J. Vardakas, and C. Verikoukis, "5G RAN Slicing: Dynamic Single Tenant Radio Resource Orchestration for eMBB Traffic within a Multi-Slice Scenario," IEEE Communications Magazine, vol. 59, no. 3, pp. 110–116, 2021, doi: 10.1109/MCOM.001.2000770.

[92] A. A. Gebremariam, M. Chowdhury, M. Usman, A. Goldsmith, and F. Granelli, "SoftSLICE: Policy-based dynamic spectrum slicing in 5G cellular networks," IEEE International Conference on Communications, vol. 2018-May, no. February, 2018, doi: 10.1109/ICC.2018.8422148.

[93] S. R. A. N. Slicing, H. Chergui, L. Blanco, C. Verikoukis, and S. Member, "Statistical Federated Learning for Beyond 5G," 2021, doi: 10.1109/TWC.2021.3109377.

[94] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning," IEEE Netw, vol. 33, no. 5, pp. 156–165, 2019, doi: 10.1109/MNET.2019.1800286.

[95] H. Zhang, N. Liu, X. Chu, K. Long, A. Aghvami, and V. C. M. Leung, "Network Slicing Based 5G and Future Mobile Networks : Mobility , Resource Management , and Challenges," no. January, 2017, doi: 10.1109/MCOM.2017.1600940.

Authors

Simon Atuah Asakipaam is an Assistant Lecturer in the Department of Electrical and Electronics Engineering, Tamale Technical University, Ghana. He received his BSc and MPhil degrees in Telecommunication Engineering from the Kwame Nkrumah University of Science and Technology, Ghana, in 2017 and 2020 respectively. He is currently pursuing his PhD. Degree in Telecommunication Engineering from the same university. His areas of interest include Software-Defined Networking, Wireless sensor networks, 5G technology, Artificial Intelligence, etc. Simon is presently investigating efficient resource provisioning and utilization in 5G network slicing.

Prof. Jerry John Kponyo is the Dean of the Quality Assurance and Planning Office of the Kwame Nkrumah University of Science and Technology (KNUST) under the Vice-Chancellor's Office. He is the former Dean of the Faculty of Electrical and Computer Engineering, KNUST. Prior to becoming Dean of the Faculty of Electrical and Computer Engineering he was Head of Electrical Engineering Department. He is currently the Deputy Project Lead of the KNUST Engineering Education Project (KEEP), a 5.5 Million Dollar Africa Center of Excellence (ACE) Impact project sponsored by the World Bank with a focus on Digital Development and Energy. He is Co-Founder of the Responsible AI Network (RAIN) Africa, which is a collaborative effort between KNUST and TUM Germany. Between 2016 and 2019 he was a visiting Professor at ESIGELEC, France on a staff mobility programme where he taught postgraduate courses in Business Intelligence and conducted research with staff of ESIGELEC. He has done extensive research in IoT, intelligent systems and AI and currently leads the Emerging Networks and Technologies Research Laboratory at the Faculty of Electrical and Computer Engineering, KNUST which focuses on digital development technologies research. Prof Kponyo's Ph.D. research focused on applying AI to solving a traffic problem in Vehicular Ad hoc Networks (VANETs). He has published over 50 articles in refereed Journals and Conference proceedings. He is a member of the Ghana Institution of Engineers. Prof. Jerry John Kponyo is currently the coordinator of the West Africa Sustainable Engineering Network for Development (WASEND). Prof Kponyo is the PI and Scientific Director of the Responsible Artificial Intelligence Lab which is a 1 Million Canadian Dollar grant sponsored by IDRC and GIZ. He is also PI for the Partner-Afrika project which is sponsored by BMZ.

Kwame Oteng Gyasi received his BSc Telecommunication Engineering degree from the Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, in 2011. He obtained his MSc and PhD Information and Communication Engineering degrees from the University of Electronics Science and Technology of China, in 2014 and 2018, respectively. He joined KNUST as a Lecturer in 2020 and has since been teaching courses in Applied Electricity, Engineering in Society, Electromagnetic Field Theory and Wireless Communication Networks. He also briefly worked as a Lecturer at Christian Service University College and GIMPA in 2018 and 2019 respectively. His research interests include slot antennas, planar monopole antennas, planar circularly polarized antennas, printed wideband antennas and telecommunication systems.

SURVEY ARTICLE

How to cite this article:

Simon Atuah Asakipaam, Jerry John Kponyo, Kwame Oteng Gyasi, "Resource Provisioning and Utilization in 5G Network Slicing: A Survey of Recent Advances, Challenges, and Open Issues", International Journal of Computer Networks and Applications (IJCNA), 10(2), PP: 201-216, 2023, DOI: 10.22247/ijcna/2023/220736.