Morphpublishing
J. AI-Driven Autom. Predict. Maint. Smart Technol.
open access academic journal
Peer reviewed

Optimizing End-to-End Throughput in Network-Sliced 5G Systems for Real-Time Collision Avoidance

Lindiwe Nyambi[a]

Keywords: 5G networks, latency, machine learning, network slicing, predictive resource allocation, real-time collision avoidance, ultrareliable low-latency communication (URLLC)

Abstract: Network slicing in fifth-generation (5G) wireless systems enables flexible resource allocation and service customization across diverse use cases. Real-time collision avoidance demands high end-to-end throughput and deterministic latency guarantees, since vehicles, drones, or autonomous machinery rely on rapid data exchange and decision-making logic. Traditional static configurations struggle to account for variable traffic patterns, dynamic channel conditions, and stringent reliability requirements. Adaptive slicing frameworks can integrate cross-layer optimization to ensure that capacity and latency constraints are jointly satisfied, while also maximizing network utilization under heterogeneous workloads. Resource blocks, slicing policies, and interference mitigation strategies must be dynamically orchestrated to balance throughput against latency, jitter, and reliability. End-to-end throughput, from the local edge domain to the core network, influences collision detection algorithms, alert forwarding, and actuator control for timely evasive actions. Emerging approaches leverage advanced scheduling, machine learning-based prediction, and real-time analytics to handle bursts in traffic and topological changes. Ongoing standardization and industrial collaborations focus on refining these solutions, emphasizing robust, scalable, and secure network orchestration. This work explores a comprehensive framework for throughput optimization and real-time collision avoidance in 5G-sliced systems, discussing the principal design challenges and highlighting how multi-dimensional resource management schemes can support mission-critical, latency-sensitive operations.

1. Introduction

High-speed wireless connectivity drives numerous safety-critical applications that rely on instantaneous decision-making and data exchange. Evolving 5G architectures, equipped with network slicing capabilities, hold the promise of delivering guaranteed performance for diverse verticals and use cases. Real-time collision avoidance stands out

[a]**Mpumalanga Institute of Technology, Department of Computer Science, Middelburg, South Africa.**

among these due to its dependency on ultra-reliable and low-latency data transmission, where even minor lags in communication may lead to severe outcomes. Network slicing, a mechanism that separates physical infrastructure into multiple virtual networks, allows operators to allocate dedicated resources to different service categories. This concept underpins mission-critical tasks that require guaranteed throughput, bounded latency, and consistent quality of service over time [1, 2].

End-to-end throughput in a network-sliced 5G environment spans multiple components: radio access, transport, and core layers. Data rates across these segments must remain sufficiently high to accommodate the rapid exchange of sensor data, environmental updates, and control signals. Latency and throughput optimization strategies extend beyond single-hop considerations, since modern 5G networks often involve edge computing nodes for local data processing and cloud components for more intensive analytics. Slicing segmentation can become complex under realistic traffic conditions, where multiple slices share the same physical substrate and compete for radio resources, backhaul capacity, and core functions. An effective design for collision avoidance must account for these constraints in parallel with a targeted focus on speed and responsiveness.

Network slicing solutions comprise control-plane and user-plane mechanisms that orchestrate the distribution of computational, storage, and bandwidth resources. Control-plane functions integrate with radio resource management modules to ensure that the correct slice-specific scheduling and slice admission policies are enforced. User-plane data transmissions proceed according to the allocated resources, with real-time flows prioritized based on service-level agreements. Effective slicing is not restricted to the radio domain; transport networks must also support variable traffic demands and guarantee consistent data rates. Traffic fluctuations become more pronounced in collision avoidance scenarios, as sensor readings often exhibit bursty patterns influenced by sudden obstacles or unexpected user maneuvers.

System-wide orchestration tools must coordinate heterogeneous radio environments, time-varying traffic loads, and cross-slice interference issues. End-to-end throughput can degrade when multiple coexisting slices use the same spectrum, especially if dynamic variations in traffic profiles or channel quality are not properly managed. Robust throughput requires intelligent handover procedures between base stations, resource reallocation in response to real-time measurements, and synergy between radio access technologies such as millimeter-wave and sub-6 GHz bands. While millimeter-wave frequencies can support higher throughput, they are vulnerable to path loss and blockage, necessitating backup mechanisms in lower frequency bands.

Collision avoidance adds another dimension to the problem: stringent delay constraints must be met while maintaining a high packet delivery ratio. Control loop stability in vehicular and drone navigation depends on packets traversing the network within milliseconds, with minimal jitter [3]. Techniques such as multi-access edge computing (MEC) can offload computation from distant clouds to local edge nodes, reducing round-trip delay. Nonetheless, the coordination of these nodes and the allocation of sufficient resources for analytics and communication require an overarching slicing strategy that respects both throughput and latency objectives.

Subsequent sections provide a deep analysis of end-to-end throughput considerations in 5G-sliced systems, focusing on how partitioned slices can be configured for real-time collision avoidance. Emphasis will be placed on the interplay between resource allocation strategies, scheduling algorithms, and system design principles that assure consistent high throughput [4, 5]. By dissecting the architectural components and exploring optimization frameworks, this discussion sheds light on how 5G networks can be shaped into reliable platforms for next-generation vehicular safety applications, autonomous navigation, and industrial automation requiring collision detection and evasion [6].

2. System Model and End-to-End Throughput Considerations

System architectures for 5G network slicing are designed to address the diverse performance requirements of heterogeneous applications while maintaining operational efficiency and scalability. Central to this approach are software-defined networking (SDN) and network function virtualization (NFV), two paradigms that enable the dynamic instantiation, scaling, and management of virtual network functions (VNFs) over shared infrastructure. These frameworks abstract the underlying physical network resources, facilitating the creation of logical network slices tailored to specific service demands. This section explores the key components, operational principles, and challenges associated with the architecture of 5G network slicing.

2.1. Frameworks and Principles

The integration of SDN and NFV into 5G slicing architectures provides a flexible and programmable control plane and a virtualized data plane. The control-plane components are responsible for traffic demand analysis, orchestration of resource allocation, and ensuring that each slice adheres to its service-level agreement (SLA). The SDN controller enables centralized management of routing, bandwidth allocation, and traffic prioritization, while the NFV framework facilitates the deployment of VNFs, such as firewalls, load balancers, and protocol converters, in a virtualized environment.

The system model follows a layered approach, where each layer enforces slice isolation and performs dynamic resource allocation. The three primary layers include:

- **Radio Access Network (RAN):** This layer focuses on resource allocation across frequency, time, and spatial dimensions to ensure reliable connectivity and performance for all slices.
- **Transport Network:** The transport network interconnects the RAN with the core network, providing high-capacity and low-latency data transport for diverse slice requirements.
- **Core Network:** The core network layer supports advanced functionalities such as authentication, mobility management, and session control, tailored to the needs of each slice.

2.2. Dynamic Resource Allocation in the RAN

The RAN is a critical component of 5G slicing, where radio resource blocks (RRBs) serve as the fundamental units for resource allocation. These blocks are dynamically aggregated to meet the capacity and latency requirements of each slice. The allocation process leverages three dimensions:

1. **Frequency:** Spectrum slices are dynamically assigned to ensure coexistence and efficiency.
2. **Time:** Scheduling techniques allocate time slots to slices based on their priority and demand.
3. **Spatial:** Beamforming and multiple-input multiple-output (MIMO) techniques optimize spatial resource allocation, improving capacity and reducing interference.

For instance, during high-demand scenarios, such as traffic congestion at intersections or dense drone cluster operations, slice reconfiguration dynamically assigns additional resources to mission-critical flows. This ensures throughput assurance and minimizes packet delays for latency-sensitive applications.

Table 1. Key Characteristics of RAN Resource Allocation for 5G Network Slicing

Dimension	Allocation Methodology	Objective
Frequency	Spectrum slicing based on traffic demand	Efficient spectrum utilization
Time	Dynamic time-slot scheduling	Latency reduction and fairness
Spatial	Beamforming and MIMO techniques	Enhanced capacity and interference mitigation

2.3. Slice Isolation and Traffic Management

To ensure performance guarantees for individual slices, 5G architectures enforce strict slice isolation. This prevents resource contention and interference among slices, particularly when mission-critical services coexist with best-effort traffic. Isolation mechanisms include:

- **Bandwidth Reservation:** Dedicated bandwidth is allocated to high-priority slices, ensuring their performance is not affected by congestion.
- **Priority Scheduling:** Traffic is scheduled based on slice priorities, allowing low-latency flows to access resources promptly.
- **Interference Coordination:** Advanced coordination techniques mitigate inter-slice interference, improving overall network efficiency.

Traffic management in 5G slicing involves balancing the needs of high-priority slices with shared-resource slices. For example, mission-critical applications such as autonomous vehicle communication or remote surgery require stringent latency and reliability guarantees. These are achieved through dedicated resource provisioning and priority-based traffic scheduling.

2.4. End-to-End Throughput Optimization

End-to-end throughput in 5G network slicing depends on coordinated decisions across the RAN, transport network, and core network. The orchestration process includes:

1. **RAN Optimization:** Dynamic allocation of RRBs and interference coordination.
2. **Transport Network Optimization:** Path selection and bandwidth allocation based on slice requirements.
3. **Core Network Optimization:** Efficient deployment of VNFs and load balancing.

The interplay of these components ensures that critical slices, such as ultra-reliable low-latency communication (URLLC), maintain high throughput and low latency, while slices for enhanced mobile broadband (eMBB) and massive machine-type communications (mMTC) achieve their respective performance targets.

The transport and core network segments in 5G slicing architectures are critical enablers of end-to-end performance. These segments interconnect the radio access network (RAN) with core data centers, where advanced control-plane and user-plane functionalities operate [7, 8]. This section delves into the mechanisms of slicing in transport networks and the role of the core network in ensuring high throughput, low latency, and reliability. It also highlights the challenges and solutions for meeting the stringent requirements of 5G applications [9].

Table 2. Comparison of Throughput Optimization Strategies Across 5G Network Layers

Network Layer	Optimization Techniques	Performance Metrics
RAN	Dynamic RRB allocation, interference coordination	Throughput, latency, reliability
Transport Network	Path selection, bandwidth allocation	Latency, jitter, capacity
Core Network	VNF deployment, load balancing	Processing efficiency, scalability

2.5. Transport Network Architecture

Transport networks connect the RAN to the core network through a hierarchy of links, typically divided into fronthaul, midhaul, and backhaul segments. Each segment has specific characteristics and constraints:

- **Fronthaul:** Connects remote radio units (RRUs) to centralized baseband units (BBUs). High data rates and low latency are paramount in this segment.
- **Midhaul:** Links the BBUs to aggregation points, facilitating resource pooling and load balancing.
- **Backhaul:** Transports aggregated traffic from the midhaul to the core network. It demands high bandwidth and reliable connectivity.

The effectiveness of slicing in transport networks depends on partitioning these segments to isolate traffic belonging to different slices. Optical fiber capacity, microwave backhaul channels, or millimeter-wave links are partitioned into logical tunnels that maintain slice isolation. Each tunnel is provisioned with specific bandwidth and latency guarantees, preventing performance degradation in one slice from impacting others [10, 11].

2.5.1. Bandwidth Provisioning and Isolation Mechanisms

Bandwidth provisioning in transport networks is achieved through techniques such as wavelength-division multiplexing (WDM) in optical networks and dynamic bandwidth allocation in microwave backhaul. Logical partitions, or tunnels, are implemented using technologies such as Multiprotocol Label Switching (MPLS) or Segment Routing (SR). These mechanisms ensure that the following requirements are met:

- **Isolation:** Traffic belonging to different slices remains independent, avoiding cross-slice interference.
- **Performance Guarantees:** Each tunnel is configured to meet the specific throughput and latency requirements of its slice.
- **Scalability:** The architecture supports the dynamic creation and scaling of tunnels based on traffic demands.

Sophisticated scheduling tools at switches and routers play a vital role in reducing jitter and maintaining consistent throughput. For example, queue management algorithms such as Weighted Fair Queuing (WFQ) prioritize high-priority slices and ensure fair resource allocation across all slices.

2.6. Core Network Role in 5G Slicing

The core network in a 5G architecture handles advanced control-plane and user-plane functions, ensuring the seamless operation of diverse network slices. Virtualized frameworks, powered by NFV, enable the deployment of

Table 3. Key Characteristics of Transport Network Slicing in 5G

Technique	Purpose	Key Features
Wavelength-Division Multiplexing (WDM)	Partitioning optical fiber capacity	High bandwidth, low latency
Multiprotocol Label Switching (MPLS)	Logical tunnel creation	Traffic isolation, scalability
Weighted Fair Queuing (WFQ)	Queue management	Jitter reduction, prioritization

critical components such as session management functions, user-plane gateways, and edge computing platforms. These components are hosted in centralized or distributed data centers, depending on the specific slice requirements.

2.6.1. Virtualized Frameworks and Network Functions

Core network elements leverage NFV to deploy and scale virtualized network functions (VNFs) dynamically. Key VNFs include:

- **Session Management Functions (SMFs):** Manage user sessions, ensuring reliable connectivity and mobility support.
- **User-Plane Gateways (UPFs):** Handle data forwarding and traffic aggregation for user devices.
- **Edge Computing Platforms:** Provide localized processing power for latency-sensitive applications.

These functions interact with the transport network and the RAN to deliver the end-to-end performance required by 5G slices. The availability of compute resources for tasks such as machine learning, sensor fusion, or advanced analytics directly impacts throughput and latency. For example, collision avoidance applications for autonomous vehicles rely on the swift execution of sensor fusion algorithms at the edge.

2.6.2. Scalability and Traffic Bursts

The core network must adapt to sudden increases in traffic volume, which can result from events such as sensor data surges or video streaming spikes. Horizontal scaling of compute nodes within edge data centers is a common solution, allowing the network to handle increased workloads without sacrificing performance. Additionally, resource orchestration platforms dynamically allocate processing power to VNFs based on real-time demands.

Table 4. Core Network Functions and Their Impact on Throughput

Core Network Function	Role	Performance Impact
Session Management Functions (SMFs)	Manage user sessions	Ensure connectivity and reliability
User-Plane Gateways (UPFs)	Data forwarding and aggregation	Improve throughput and scalability
Edge Computing Platforms	Localized processing of tasks	Reduce latency and enhance responsiveness

2.7. End-to-End Throughput Considerations

End-to-end throughput in 5G slicing is influenced by several factors across the RAN, transport network, and core network:

- **Raw Link Capacity:** The bandwidth of radio links, optical fibers, or microwave channels directly affects throughput.
- **Traffic Isolation:** Effective slicing mechanisms ensure that high-usage slices do not degrade the performance of others.
- **Network Function Responsiveness:** The execution speed of VNFs, especially in edge data centers, determines the network's ability to handle latency-sensitive tasks.

Packet forwarding policies at switches and routers are dynamically adapted to maintain throughput thresholds. For example, during collision avoidance maneuvers for autonomous vehicles, packet forwarding is prioritized to ensure minimal delay and reliable communication.

Inter-slice interference management arises when multiple slices operate concurrently in close geographic proximity. Cellular reuse patterns can expose collision avoidance slices to potential interference from other services. Overlapping frequency channels, high user density, or suboptimal scheduling can all degrade throughput. Coordinated multi-point (CoMP) transmission and beamforming can mitigate interference by aligning transmissions across base stations. Algorithms for slice-aware power control prevent oversaturated links, ensuring that mission-critical slices achieve adequate signal quality. Intelligent resource reuse planning and predictive analytics help operators forecast traffic levels and adapt the slicing parameters before congestion worsens.

End-to-end throughput is also influenced by mobility patterns of devices requesting the collision avoidance service. Handovers between base stations must be seamless to prevent data interruptions. Robust throughput is maintained through algorithms that select the target cell or beam with the best capacity while minimizing reconfiguration delays. Multi-connectivity, where a device connects to multiple base stations or frequency layers, can boost throughput by aggregating resources. Redundant transmissions over multiple slices may further enhance reliability for critical packets, though at the expense of resource overhead. Dynamic orchestrators in the control plane can coordinate these multi-connectivity features, deciding when to maintain additional connections based on forecasted collision risk or network congestion.

Data encapsulation and protocol overhead also matter when aiming for high throughput. Various tunnel setups, such as encapsulation for slicing or encryption for security, introduce header overhead that reduces the net payload capacity. Optimizing protocols within the 5G core can minimize such overhead, yet slicing requires encapsulation layers to maintain isolation and ensure privacy. The trade-off between robust isolation and peak throughput must be carefully navigated, especially in safety-critical applications. Achieving better throughput may necessitate refinements to encapsulation strategies or near-real-time adaptation of protocol parameters.

Edge computing platforms, placed at or near the base stations, aim to shorten the data path for collision avoidance applications. These platforms handle tasks such as sensor data aggregation, object detection, and motion planning. Throughput in these segments depends on tight coupling between radio resources and computing resources. Data must be swiftly offloaded, processed, and returned to the user or actuator. The synergy between slicing and edge computing is a cornerstone of next-generation collision avoidance, as it offloads significant computational tasks away from centralized clouds. Lower round-trip times enhance system responsiveness, and throughput is stabilized by distributing data flows closer to the end-users.

Achieving robust end-to-end throughput for real-time collision avoidance also hinges on implementing admission control policies that anticipate high-demand scenarios. Traffic forecasting is feasible via machine learning that analyzes historical data and current sensor input patterns. When traffic spikes are predicted, the slicing orchestrator can reserve additional bandwidth, scale computing resources, or reassign frequency blocks. Dynamic adaptation at all layers ensures that throughput remains above a critical threshold, giving collision avoidance algorithms adequate margins to operate effectively. Once demand subsides, resources can be reallocated to other slices to maximize overall network utilization.

3. Network Slicing Strategies and Resource Allocation

The concept of network slicing is integral to 5G systems, enabling operators to create multiple logical networks, each tailored to specific application requirements, over a shared physical infrastructure. Specialized orchestrators govern this process, dividing infrastructure resources among various slices with varying levels of granularity. This section explores the principles of slicing, the strategies for managing network resources dynamically, and the mechanisms for ensuring end-to-end throughput across multiple slices.

3.1. Slicing Orchestration and Granularity Levels

Network slicing orchestrators are the central entities responsible for the management and allocation of physical and virtual resources among multiple slices. These orchestrators interact with software-defined networking (SDN) controllers and network function virtualization (NFV) platforms to ensure seamless resource provisioning. The granularity of slicing can be categorized into three primary levels:

1. **Coarse-Grained Slicing:** Allocates large portions of network capacity, such as frequency bands or entire transport links, to individual slices. This approach is typically used for services with predictable, high-demand traffic, such as enhanced mobile broadband (eMBB).
2. **Medium-Grained Slicing:** Involves partitioning the network into smaller units, such as specific time slots or segments of optical fiber capacity. This level of granularity is suited for applications like ultra-reliable low-latency communication (URLLC), where fine control of resources is necessary.
3. **Fine-Grained Slicing:** Focuses on allocating individual resource blocks (e.g., radio resource blocks or processing units) to slices based on real-time demand. This is critical for highly dynamic applications, such as collision avoidance or mission-critical industrial automation.

The choice of slicing granularity depends on the specific application requirements, network conditions, and the trade-offs between performance and complexity. Coarse-grained slicing is simpler to implement but may result in resource underutilization, whereas fine-grained slicing offers higher efficiency at the cost of increased computational overhead.

3.2. Static vs. Dynamic Slicing Strategies

The management of resources across slices can be broadly categorized into static and dynamic slicing strategies. Each approach has distinct advantages and trade-offs, depending on the network scenario and application requirements.

Table 5. Comparison of Slicing Granularity Levels in 5G Networks

Granularity Level	Characteristics	Applications
Coarse-Grained Slicing	Large capacity partitions, low flexibility	eMBB, video streaming
Medium-Grained Slicing	Moderate partitioning, balance between flexibility and simplicity	URLLC, industrial IoT
Fine-Grained Slicing	Individual resource block allocation, high flexibility	Collision avoidance, drone swarms

3.2.1. Static Slicing

Static slicing involves fixed allocation of network resources, ensuring that specific slices receive a predetermined share of bandwidth, processing power, or other resources. This strategy is often employed for mission-critical applications where performance guarantees are non-negotiable. For instance, collision avoidance slices may reserve dedicated frequency bands or optical links to ensure uninterrupted operation during peak traffic periods.

While static slicing provides reliable performance for high-priority slices, it can lead to inefficient resource utilization during off-peak periods. For example, if collision avoidance traffic subsides, the reserved resources may remain underutilized, potentially affecting the overall network efficiency.

3.2.2. Dynamic Slicing

Dynamic slicing relies on real-time monitoring, predictive analytics, and reconfiguration algorithms to allocate resources flexibly. This approach addresses the limitations of static slicing by adapting to variations in network load, user density, and device mobility. Dynamic slicing involves:

- **Real-Time Monitoring:** Network elements continuously collect metrics such as throughput, latency, and congestion levels to identify changing traffic patterns.
- **Predictive Analytics:** Machine learning models predict future demand based on historical data and current trends, enabling proactive resource allocation.
- **Automated Reconfiguration:** Orchestration platforms dynamically adjust resource allocations, ensuring that each slice meets its performance requirements.

For example, during peak congestion at a busy intersection, dynamic slicing algorithms can allocate additional radio resource blocks (RRBs) to collision avoidance slices. Once the congestion subsides, these resources can be reallocated to other slices, such as best-effort streaming or file transfers [12, 13].

3.3. Adaptive Resource Allocation for Collision Avoidance

Collision avoidance systems are among the most demanding applications in 5G networks, requiring ultra-low latency, high reliability, and dynamic adaptability. The performance of these systems depends on how effectively resources are allocated in response to traffic events, such as vehicle congestion or drone swarms [14].

3.3.1. Peak Demand Scenarios

In scenarios where peak demand arises intermittently, such as during traffic jams or unexpected obstacles, collision avoidance slices must be prioritized. Dynamic slicing mechanisms can temporarily assign additional bandwidth, processing power, or routing capacity to these slices, ensuring timely communication between devices. Key techniques include:

- **Priority-Based Scheduling:** High-priority packets are prioritized in scheduling queues, reducing delays and ensuring timely delivery.
- **Bandwidth Augmentation:** Additional spectrum is allocated to collision avoidance slices during peak periods.
- **Latency Minimization:** Low-latency paths are dynamically selected to reduce communication delays.

3.3.2. Off-Peak Optimization

During off-peak periods, dynamic slicing frameworks release unused resources from collision avoidance slices and redistribute them to other slices. This improves overall network efficiency while maintaining the readiness to respond to sudden demand spikes [4, 15, 16].

Table 6. Resource Allocation Strategies for Collision Avoidance in 5G Networks

Scenario	**Resource Allocation Method**	**Outcome**
Peak Demand	Priority scheduling, bandwidth augmentation	Reduced latency, high reliability
Off-Peak Periods	Resource redistribution to other slices	Improved network efficiency
Real-Time Reconfiguration	Automated adjustments based on traffic patterns	Optimal resource utilization

3.4. End-to-End Throughput Optimization

Achieving high end-to-end throughput in 5G slicing architectures requires coordinated efforts across the RAN, transport network, and core network. The following considerations are critical:

3.4.1. Interplay Between Static and Dynamic Policies

While dynamic slicing is essential for handling fluctuating demand, certain slices may benefit from a hybrid approach that combines static and dynamic policies. For example, a collision avoidance slice might reserve a baseline level of resources statically while dynamically acquiring additional capacity during peak demand.

3.4.2. Network Function Responsiveness

End-to-end throughput is influenced by the execution speed of network functions, particularly in edge data centers. Functions such as sensor fusion, machine learning inference, and analytics must be executed promptly to avoid bottlenecks. Horizontal scaling of compute nodes and efficient orchestration of virtualized network functions (VNFs) are vital for maintaining throughput under heavy traffic conditions.

Resource orchestration extends to scheduling algorithms at the base station level. Each slice receives a share of transmission opportunities, determined by sophisticated schedulers that account for queue lengths, channel quality, and latency requirements. Weighted round-robin or proportional fair scheduling can be adapted to incorporate slice priorities, with collision avoidance flows gaining precedence. These mechanisms must be integrated with slice-level admission control, ensuring that the total load does not exceed what can be handled while maintaining performance guarantees. By correlating slice traffic forecasts with radio conditions, scheduling can deliver consistent throughput over time.

Transport network slicing is achieved through segment routing or MPLS-based tunnels, with traffic engineered paths set up for each slice. Bandwidth on these paths is reserved according to policies, preventing one slice from saturating the links and affecting others. For collision avoidance, high-priority packets may be routed over low-latency links, leveraging a robust path diversity approach if feasible. Multi-path transport protocols further boost throughput by dispersing traffic over multiple links. These protocols can handle link failures and congestion by rerouting flows quickly, mitigating throughput degradation. This arrangement often requires close interaction between the control-plane orchestrator and routing controllers in the transport network.

Core slicing primarily involves deploying VNF instances that handle user-plane data in a slice-aware fashion. A resource allocation mechanism within the core ensures that sufficient CPU, memory, and storage are assigned to the collision avoidance slice. Horizontal scaling of these virtual functions can accommodate higher throughput during surge periods. Load balancers direct flows to underutilized virtual machines or containers, distributing the processing tasks. This model lowers the risk of bottlenecks in the user-plane, maintaining the high data rates necessary for real-time collision avoidance. However, the overhead in instantiating or migrating VNF instances must be minimized to avoid slow adjustments.

Machine learning (ML) algorithms can play a transformative role in dynamic slicing. Predictions of mobility patterns, user demands, or channel states can feed into the orchestration process, enabling preemptive modifications of resource allocations. Through such forecasting, an operator may enlarge the collision avoidance slice before a rush hour begins or during periods of expected inclement weather when accident risk is elevated. Reinforcement learning frameworks that reward stable throughput and minimal latency can learn optimal slicing decisions over time, adjusting parameters such as bandwidth allocation, scheduling weights, and VNF scaling policies. Coupled with real-time analytics, these methods unlock greater autonomy and efficiency in slice management.

End-to-end isolation, a key objective of slicing, sometimes conflicts with resource-sharing efficiency. Collision avoidance slices call for assured throughput and low latency, suggesting a high degree of isolation. On the other hand, total isolation can cause underutilized resources, raising operational costs. Solutions blend the two by providing partial isolation with flexible boundaries, ensuring that mission-critical slices maintain priority access while allowing opportunistic use of leftover capacity by other services. This approach requires robust slice-level service-level agreements (SLAs) that delineate guaranteed minimum throughput and maximum allowable latency.

Orchestrators also interface with policy engines to interpret business objectives and regulatory constraints. Certain safety regulations might mandate a dedicated collision avoidance slice with strict performance thresholds. Enforcement of these policies influences how capacity is segmented among slices. Fine-grained real-time analytics can detect policy violations, such as subpar throughput or excessive delay, and trigger immediate remedial actions. These actions might involve admitting fewer best-effort users or reallocating spectrum from less critical services. A multi-layered monitoring system that supervises key performance indicators (KPIs) at radio, transport, and core layers supplies the orchestrator with data to maintain SLA compliance.

Another element in resource allocation is the provisioning of adequate capacity to the edge computing nodes.

Collision avoidance computations sometimes involve complex sensor fusion, object detection, and trajectory planning. The ingestion and distribution of high-resolution images, lidar outputs, or radar readings require stable throughput between the user equipment and the edge node. To that end, slicing frameworks ensure that the edge nodes have robust uplink and downlink connectivity. Schedulers can prioritize collision avoidance data traffic toward edge servers, while edge computing policies determine the CPU and GPU resource shares reserved for real-time analytics. If throughput or latency thresholds risk being exceeded, the system can seamlessly hand off computation tasks to a neighboring edge node with available capacity.

4. Scheduling Mechanisms and Real-Time Collision Avoidance

Scheduling in 5G-sliced networks addresses multiple priorities: average throughput, latency, fairness, and user experience. Real-time collision avoidance adds an additional requirement for ultra-reliable low-latency communication (URLLC). Schedulers are tasked with meeting strict delay budgets, often on the order of milliseconds, while simultaneously guaranteeing throughput targets. These algorithms work at the intersection of slice-level resource allocation and physical-layer constraints such as channel quality indicators and interference patterns.

In a typical scheduling scenario, the base station gathers real-time indicators of channel state, queue occupancy, and slice requirements. The scheduler assigns resource blocks in each transmission interval according to a priority function that weighs instantaneous throughput demands, guaranteed bit rates for high-priority slices, and fairness among users. For collision avoidance data flows, the priority function emphasizes low packet delay, awarding immediate transmission opportunities to these flows if they are queued. Although this approach satisfies urgent traffic, it can penalize best-effort slices. Balancing these interests becomes more complicated as multiple high-priority services might coexist, such as telemedicine or industrial automation with parallel low-latency demands.

Resource allocation can leverage multi-carrier features of 5G, distributing collision avoidance traffic across multiple carrier frequencies or even different radio technologies. Aggregating capacity across frequency bands increases the effective throughput and provides backup paths if one carrier experiences congestion or poor channel conditions. Dual connectivity in 5G allows a user device to connect simultaneously to a master and a secondary base station, effectively doubling the resources available. Collision avoidance flows benefit from this setup by having redundancy in their data paths, mitigating the impact of localized interference or fading.

Frame structure configuration also plays a part in real-time scheduling. Shorter transmission time intervals (TTIs) can reduce latency, though at the expense of additional overhead. In pursuit of near-instantaneous scheduling, 5G defines mini-slots that occupy a fraction of a standard TTI, facilitating rapid dispatch of time-sensitive packets. Throughput considerations remain crucial, since an overly aggressive reduction in slot length can degrade spectral efficiency. Maintaining a balanced approach helps collision avoidance flows achieve minimal delay without unduly compromising overall throughput.

Adaptive modulation and coding schemes optimize link performance by matching the modulation order and coding rate to current channel conditions. This adaptation ensures that high-throughput modes are used when signal quality is good, while robust modes are selected during channel fading or interference events. Collision avoidance slices benefit from robust modes that guarantee packet delivery, though throughput may temporarily suffer. A well-designed scheduler can exploit favorable channel conditions to enhance throughput for collision avoidance slices while reserving more robust transmission modes for less critical slices when needed.

Real-time collision avoidance necessitates consistent communication with minimal jitter, prompting the use of guaranteed scheduling intervals for periodic control messages. A slice might reserve a fraction of each subframe for collision avoidance traffic, ensuring that crucial telemetry or control data can be transmitted even in congested conditions. In such setups, competing slices must adapt to the leftover resources, and overall throughput distribution must still align with each slice's SLA. Dynamic bandwidth partitioning can be triggered when the collision avoidance application detects an imminent hazard, granting additional resources for a short period to expedite data exchange.

Interference management is a critical scheduling concern, especially in dense deployments that serve multiple slices in overlapping frequencies. Coordinated scheduling across neighboring cells can mitigate co-channel interference by synchronizing resource allocations or beam directions. Collision avoidance slices stand to gain from advanced interference cancellation techniques, such as interference shaping and space-frequency block coding, that enhance signal-to-interference-plus-noise ratios. Precise alignment of beams and power control can protect mission-critical transmissions, increasing the effective throughput and reliability of the collision avoidance slice.

Prioritization of collision avoidance traffic can extend beyond the physical layer to incorporate queue management mechanisms. Packets carrying collision alert or sensor data might be placed in high-priority queues, ensuring they are transmitted first. These queues typically have smaller buffers to reduce the risk of excessive queuing delays. If the queue size is too small, however, bursty traffic may experience packet drops. Intelligent dynamic buffer sizing, informed by real-time traffic analytics, can mitigate this challenge by adjusting buffer thresholds according to load and channel conditions. This approach preserves throughput while offering minimal latency for safety-critical traffic.

Routing and scheduling synergy becomes vital in multi-hop or multi-layer networks. Collisions can occur not only at the radio interface but also within transport or edge segments if scheduling is misaligned or if routing fails to account for dynamic slice demands. Traffic engineering solutions can coordinate link-level scheduling decisions with high-level routing to prevent bottlenecks. For instance, if a certain backhaul segment is nearing capacity, the network orchestrator can reroute some traffic around the congested link or adjust scheduling weights to prioritize collision avoidance flows. This holistic approach integrates radio-level scheduling with broader end-to-end management, ensuring that throughput objectives are consistently met.

Real-time collision avoidance often employs advanced algorithms such as simultaneous localization and mapping (SLAM), cooperative perception, or machine learning-based hazard prediction. These algorithms require continuous, high-throughput data from sensors and neighboring devices to detect potential collisions. When a risk is identified, the network must dispatch alerts or commands to relevant vehicles or drones within milliseconds. A stable throughput that can handle bursts of data during hazard detection phases is crucial. Large sensor data spikes could occur when an autonomous vehicle encounters complex objects requiring high-resolution analysis. Scheduling and resource allocation frameworks that adapt to these spikes in real-time serve as the backbone of collision avoidance.

5. Conclusion

Efficient end-to-end throughput management in 5G-sliced networks emerges as a cornerstone for achieving real-time collision avoidance in safety-critical environments. System design must integrate flexible slicing policies, robust scheduling algorithms, and dynamic resource allocation methods to handle fluctuating traffic conditions without sacrificing reliability or latency performance. Edge computing, in conjunction with advanced network orchestration, empowers mission-critical applications to capitalize on localized data processing and reduced round-trip times. Combining these architectural elements ensures that collision detection and avoidance mechanisms function predictably, minimizing the risk of catastrophic events.

Use cases involving vehicular networks, drone swarms, and industrial automation underscore the importance of tightly coupling slicing mechanisms with domain-specific requirements. Traffic surges and bursty sensor data flows demand the capacity to adapt on the fly, reallocating spectrum, bandwidth, and computing resources where they are needed most. This flexibility hinges on the interplay between control-plane intelligence and well-tuned user-plane configurations, spanning the radio access domain, transport segments, and core network. Machine learning techniques, augmented by real-time analytics, promise to refine slicing operations further by predicting demand and optimizing resource provisioning before congestion or performance degradation occurs.

Future research directions include deeper studies of the interplay between low-latency scheduling, interference management, and machine learning-based orchestration approaches that aim to strike a balance between high throughput and reliable performance. Ongoing standardization efforts, coupled with proofs of concept in both academia and industry, are shaping the evolving 5G ecosystem toward robust collision avoidance capabilities. The complex challenge of aligning throughput, latency, and reliability in a multi-slice context is being tackled through continuous innovation at the protocol, system architecture, and algorithmic levels. This integrated strategy holds the key to future autonomous operations where human lives, infrastructure integrity, and reliable connectivity converge in real time.

References

[1] V. Loginov, A. Troegubov, A. Lyakhov, and E. Khorov, "Enhanced collision resolution methods with mini-slot support for 5g nr-u," *IEEE Access*, vol. 9, pp. 146 137–146 152, 2021.

[2] S. Barmpounakis, G. Tsiatsios, M. Papadakis, E. Mitsianis, N. Koursioumpas, and N. Alonistioti, "Collision avoidance in 5g using mec and nfv: The vulnerable road user safety use case," *Computer Networks*, vol. 172, p. 107150, 2020.

[3] F. A. Farahani, S. B. Shouraki, and Z. Dastjerdi, "Generating control command for an autonomous vehicle based on environmental information," in *International Conference on Artificial Intelligence and Smart Vehicles*. Springer, 2023, pp. 194–204.

[4] M. Vaezi, Z. Ding, and H. V. Poor, "Multiple access techniques for 5g wireless networks and beyond," 2019.

[5] G. H. Sim, A. Loch, A. Asadi, V. Mancuso, and J. Widmer, "5g millimeter-wave and d2d symbiosis: 60 ghz for proximity-based services," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 140–145, 2017.

[6] S. Bhat and A. Kavasseri, "Multi-source data integration for navigation in gps-denied autonomous driving environments," *International Journal of Electrical and Electronics Research*, vol. 12, no. 3, pp. 863–869, 2024.

[7] M. Boban, A. Kousaridas, K. Manolakis, J. Eichinger, and W. Xu, "Use cases, requirements, and design considerations for 5g v2x," *arXiv preprint arXiv:1712.01754*, 2017.

[8] S. Scott-Hayward and E. Garcia-Palacios, "Multimedia resource allocation in mmwave 5g networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 240–247, 2015.

[9] S. Bhat, "Leveraging 5g network capabilities for smart grid communication," *Journal of Electrical Systems*, vol. 20, no. 2, pp. 2272–2283, 2024.

[10] M. Gharba, H. Cao, S. Gangakhedkar, J. Eichinger, A. R. Ali, K. Ganesan, V. Jain, S. Lapoehn, T. Frankiewicz, T. Hesse *et al.*, "5g enabled cooperative collision avoidance: System design and field test," in *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2017, pp. 1–6.

[11] D. Garcia-Roger, J. F. de Valgas, J. F. Monserrat, N. Cardona, and N. Incardona, "Hardware testbed for sidelink transmission of 5g waveforms without synchronization," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2016, pp. 1–6.

[12] A. Mazin, M. Elkourdi, and R. D. Gitlin, "Comparison of slotted aloha-noma and csma/ca for m2m communications in iot networks," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2018, pp. 1–5.

[13] G. Damigos, T. Lindgren, S. Sandberg, and G. Nikolakopoulos, "Performance of sensor data process offloading on 5g-enabled uavs," *Sensors*, vol. 23, no. 2, p. 864, 2023.

[14] S. M. Bhat and A. Venkitaraman, "Hybrid v2x and drone-based system for road condition monitoring," in *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. IEEE, 2024, pp. 1047–1052.

[15] P. Kutsevol, V. Loginov, E. Khorov, and A. Lyakhov, "Analytical study of license-assisted access in 5g networks," in *2019 IFIP Networking Conference (IFIP Networking)*. IEEE, 2019, pp. 1–9.

[16] D. Hetzer, M. Muehleisen, A. Kousaridas, and J. Alonso-Zarate, "5g connected and automated driving: Use cases and technologies in cross-border environments," in *2019 European conference on networks and communications (EuCNC)*. IEEE, 2019, pp. 78–82.