Power System Protection and Control
ISSN:1674-3415

AI-DRIVEN CROSS-LAYER OPTIMIZATION OF WI-FI 7 AND 5G FWA HYBRID BROADBAND SYSTEMS

Vikas Suresh Bagora

1130 Kifer Road, APT 2-438, Sunnyvale, CA 94086-5399
vikas.bagora@gmail.com

Recevied: 22 December 2024 **Revised**: 27 January 2025 **Accepted**: 20 Febuary 2025

ABSTRACT

The convergence of Wi-Fi 7 and 5G Fixed Wireless Access (FWA) technologies offers unprecedented opportunities for hybrid broadband systems that leverage complementary strengths of both access technologies. However, realizing this potential requires intelligent orchestration across multiple protocol layers and radio interfaces. This research presents an AI-driven cross-layer optimization framework that coordinates physical layer parameters, MAC scheduling, and application-level QoS to maximize performance in hybrid Wi-Fi 7/5G FWA deployments. Our system employs machine learning models for traffic prediction, intelligent failover decision-making, and adaptive routing that considers latency, bandwidth, and reliability requirements across heterogeneous access technologies. The framework implements dynamic load balancing, application-aware traffic steering, and predictive maintenance capabilities that anticipate connectivity issues before they impact users. Experimental validation across residential and enterprise scenarios demonstrates 43% improvement in aggregate throughput, 58% reduction in tail latency for critical applications, and 91% reduction in service disruptions during access link failures. The proposed architecture addresses fundamental challenges in multi-radio environments including asymmetric link characteristics, varying propagation conditions, and diverse application requirements, providing a comprehensive solution for next-generation hybrid broadband systems.

***Keywords**: Wi-Fi 7, 5G FWA, cross-layer optimization, hybrid broadband, AI orchestration, traffic prediction, adaptive QoS, multi-radio systems*

INTRODUCTION

Modern broadband connectivity increasingly relies on heterogeneous access technologies that complement each other's strengths and compensate for weaknesses. Wi-Fi 7 delivers exceptional performance in short-range indoor environments with multi-gigabit throughput through 320MHz channels and multi-link operation [Kumar et al., 2022]. Meanwhile, 5G Fixed Wireless Access extends broadband to locations where fiber deployment proves economically impractical, providing reliable wide-area connectivity despite somewhat higher latency and lower peak rates compared to fiber or advanced Wi-Fi.

The combination of Wi-Fi 7 and 5G FWA creates compelling hybrid architectures for residential and enterprise deployments. Homes might use 5G FWA as primary internet access with Wi-Fi 7 providing local distribution, while enterprises could employ Wi-Fi 7 for primary connectivity with 5G FWA providing backup and capacity overflow. However, simply deploying both technologies side-by-side without intelligent coordination wastes their potential. The challenge lies in orchestrating these disparate access methods to function as a unified system rather than independent alternatives [Chen and Martinez, 2021].

Traditional network architectures treat different access technologies as isolated layers with minimal coordination beyond basic failover mechanisms. When fiber fails, traffic switches to cellular backup. When Wi-Fi becomes congested, devices might roam to cellular. These crude approaches ignore opportunities for sophisticated optimization that considers current conditions, application requirements, and predicted future states. The result is suboptimal performance where available capacity goes unused while other links suffer congestion [Park et al., 2020].

Cross-layer optimization recognizes that decisions at different protocol layers interact in complex ways. Physical layer choices about modulation and coding affect achievable throughput. MAC layer scheduling determines which traffic receives priority and when transmissions occur. Application layer requirements specify acceptable

 DOI: 10.46121/pspc.53.1.6

latency, bandwidth, and reliability bounds. Optimizing these layers independently produces suboptimal outcomes because local optimizations create conflicts that degrade overall performance [Anderson and Liu, 2022].

Artificial intelligence offers powerful tools for managing the complexity of hybrid multi-radio systems. Machine learning models can learn patterns in traffic demand, predict congestion before it occurs, and make sophisticated routing decisions that balance multiple objectives simultaneously. Unlike rule-based approaches that struggle with edge cases and unexpected scenarios, learned models generalize from training data to handle novel situations gracefully [Thompson et al., 2021].

Current research addresses individual aspects of hybrid networking—traffic prediction, load balancing, or QoS management—but lacks comprehensive frameworks integrating these elements across Wi-Fi 7 and 5G FWA technologies specifically. The unique characteristics of these next-generation access technologies, particularly Wi-Fi 7's multi-link operation and 5G FWA's network slicing capabilities, create optimization opportunities that existing approaches do not exploit [Roberts and Singh, 2022].

This research develops an integrated AI-driven cross-layer optimization framework specifically designed for Wi-Fi 7 and 5G FWA hybrid systems. Our approach coordinates physical layer parameters, MAC scheduling decisions, and application-level routing through centralized intelligence that learns from system telemetry. The framework addresses real-world challenges including asymmetric uplink/downlink characteristics, varying propagation conditions, diverse application requirements, and the need for graceful failover during connectivity disruptions.

LITERATURE REVIEW

Multi-Radio Access Networks

Heterogeneous network architectures combining cellular and Wi-Fi access have evolved significantly over the past decade. Early approaches focused on network selection where devices chose between available access technologies based on signal strength or static policies. These basic methods failed to consider current load, application needs, or quality differences between networks [Zhang et al., 2018]. Users frequently associated with suboptimal networks, experiencing poor performance despite better alternatives being available.

The 3GPP Access Network Discovery and Selection Function (ANDSF) introduced more sophisticated network selection with operator-defined policies. ANDSF enabled carriers to guide devices toward preferred networks based on location, time, and user subscription. However, ANDSF remained fundamentally reactive, responding to current conditions without predicting future states or learning from historical patterns [Wilson and Chen, 2019]. The complexity of policy definition also limited practical deployment.

Multi-Path TCP and related transport layer approaches enable simultaneous use of multiple access technologies for single connections. MPTCP aggregates bandwidth from Wi-Fi and cellular links while providing resilience through redundant paths. Research demonstrated significant throughput improvements and reduced disruption during handovers [Paasch et al., 2020]. However, MPTCP operates at the transport layer with limited visibility into lower layer conditions, constraining its optimization effectiveness.

Wi-Fi 7 Capabilities and Challenges

Wi-Fi 7, formally IEEE 802.11be, introduces transformative enhancements over previous generations. Multi-link operation allows devices to transmit and receive simultaneously across multiple bands (2.4GHz, 5GHz, 6GHz), enabling load distribution and redundancy. Extremely High Throughput through 320MHz channels, 4096-QAM modulation, and enhanced multi-user MIMO delivers theoretical peak rates exceeding 40Gbps [Kumar et al., 2022].

The standard includes deterministic latency improvements crucial for time-sensitive applications like gaming and industrial automation. Enhanced resource allocation mechanisms and coordinated spatial reuse reduce contention and interference. However, realizing these capabilities requires sophisticated management that current deployments lack. Without intelligent orchestration, MLO may degrade to simple failover rather than true multi-path operation [Harrison and Taylor, 2022].

Practical Wi-Fi 7 deployments face challenges from increased complexity. More bands, wider channels, and advanced features create larger configuration spaces that human administrators struggle to optimize manually. The interaction between MLO, spatial reuse, and legacy device support requires careful tuning. Research shows naive configurations often achieve only 30-40% of theoretical performance [Martinez et al., 2021].

5G FWA Characteristics

5G Fixed Wireless Access leverages cellular infrastructure to deliver home and business broadband, particularly in areas lacking fiber deployment. Using frequencies from sub-6GHz to mmWave bands, FWA provides substantial capacity gains over 4G LTE while avoiding costs and delays of fiber installation. Carrier aggregation and beamforming enable multi-hundred-megabit and even gigabit-class services in favorable conditions [Davies et al., 2020].

However, 5G FWA exhibits distinct characteristics compared to Wi-Fi or fiber. Higher latency stems from cellular protocol overhead and longer propagation distances, typically 20-40ms versus <5ms for local Wi-Fi. Throughput varies significantly with distance from cell sites, building penetration, and atmospheric conditions. Uplink capacity often lags downlink substantially, creating asymmetric performance [Sullivan and Morrison, 2021].

Network slicing capabilities in 5G enable logical separation of traffic types with different QoS guarantees. Enhanced mobile broadband (eMBB), ultra-reliable low latency (URLLC), and massive IoT slices can coexist on shared infrastructure with performance isolation. Exploiting slicing for FWA residential and business services requires coordination with application requirements [Patel and Kumar, 2020].

Cross-Layer Optimization Approaches

Cross-layer design challenges the traditional layered network architecture by allowing information sharing and joint optimization across layers. Early cross-layer work focused on wireless sensor networks where energy constraints necessitated breaking layer boundaries. Research demonstrated that coordinating MAC scheduling with routing decisions could substantially extend network lifetime [Garcia and Liu, 2019].

Application of cross-layer concepts to Wi-Fi networks addressed challenges like hidden terminals and rate adaptation. Physical layer channel conditions inform MAC backoff parameters, while application requirements shape PHY modulation selection. These approaches improved throughput 20-40% compared to independent layer optimization. However, most work addressed single-technology networks rather than heterogeneous systems [Morrison et al., 2018].

Joint optimization of multiple access technologies introduces additional complexity. Different technologies operate on disparate time scales—Wi-Fi contention resolution occurs microsecond-by-microsecond while cellular scheduling happens on millisecond frames. Propagation characteristics, coverage areas, and capacity profiles differ dramatically. Unified frameworks must abstract these differences while preserving essential control capabilities [Rahman and Anderson, 2021].

AI in Network Optimization

Machine learning applications in networking have progressed from offline analytics to real-time control. Early work applied supervised learning to traffic classification and anomaly detection using historical packet traces. These approaches improved accuracy over rule-based methods but operated primarily as measurement and monitoring tools rather than active control systems [Thompson et al., 2021].

Reinforcement learning emerged as a powerful technique for network optimization where agents learn optimal policies through interaction with environments. RL applications include congestion control, routing, and resource allocation. Deep reinforcement learning using neural networks as function approximators enables handling high-dimensional state spaces characteristic of complex networks [Chen and Martinez, 2021].

Predictive models using time series analysis and recurrent neural networks forecast traffic demand, enabling proactive resource allocation. LSTM and GRU architectures capture temporal dependencies in traffic patterns, predicting future load with 80-90% accuracy at timescales from seconds to hours. These predictions inform preemptive optimization that prevents congestion rather than reacting to it [Williams and Zhang, 2020].

DOI: 10.46121/pspc.53.1.6

QoS Management in Hybrid Networks

Quality of Service mechanisms ensure critical applications receive necessary resources despite competition from bulk traffic. Traditional QoS approaches classify traffic into priority classes with differentiated treatment. VoIP and video receive priority over file downloads. While effective for coarse differentiation, these approaches lack sophistication for modern application portfolios [Park et al., 2020].

Application-aware QoS recognizes that different applications within the same class have varying requirements. Interactive video conferencing demands low latency while streaming video tolerates higher latency but requires sustained bandwidth. Gaming applications need both low latency and stable jitter. Effective QoS must understand these nuances and adapt dynamically as application mix changes [Roberts and Singh, 2022].

In hybrid networks spanning Wi-Fi and cellular, QoS coordination becomes more complex. The access technologies offer different QoS mechanisms—Wi-Fi uses EDCA and WMM while cellular employs bearer management and network slicing. Unified QoS frameworks must translate application requirements into technology-specific configurations while maintaining end-to-end guarantees [Sullivan and Morrison, 2021].

Research Gaps

Existing literature addresses components of hybrid network optimization but lacks comprehensive frameworks specifically targeting Wi-Fi 7 and 5G FWA integration. Work on multi-radio access predates Wi-Fi 7's multi-link operation and 5G FWA's widespread deployment, missing opportunities these technologies enable. Cross-layer optimization research typically focuses on single technologies rather than heterogeneous systems. AI applications in networking often target individual problems like traffic prediction or routing rather than integrated orchestration.

Our research addresses these gaps through a unified framework that exploits Wi-Fi 7 and 5G FWA specific capabilities, coordinates optimization across protocol layers, and employs AI throughout the control plane for prediction, decision-making, and adaptation.

METHODOLOGY

System Architecture

Our framework employs a centralized orchestration controller with distributed enforcement agents. The controller aggregates telemetry from Wi-Fi 7 access points and 5G FWA customer premises equipment, processes this data through AI models, and distributes optimization decisions back to enforcement agents. This architecture balances centralized intelligence with distributed execution, avoiding single points of failure while enabling sophisticated coordination.

The orchestration controller implements four primary functional modules: traffic prediction, cross-layer optimizer, failover manager, and QoS coordinator. Each module focuses on specific optimization aspects while sharing data and coordinating decisions. An integration layer resolves conflicts when modules generate contradictory recommendations, applying priority rules and constraint satisfaction to produce coherent optimization plans.

Enforcement agents run on Wi-Fi access points and FWA CPE units, implementing controller directives through technology-specific mechanisms. Wi-Fi agents configure MLO parameters, channel assignments, transmit power, and client steering. FWA agents manage network slice selection, QoS class mappings, and bearer configurations. Agents also collect local telemetry and forward it to the controller, closing the control loop.

Traffic Prediction Models

We employ ensemble learning combining multiple prediction models for different traffic characteristics and time scales. Short-term throughput prediction (1-60 seconds ahead) uses LSTM networks that capture temporal patterns in traffic intensity. The model processes sliding windows of historical throughput measurements, client counts, and application mix to forecast near-term demand.

Long-term prediction (5-60 minutes ahead) employs gradient boosting decision trees incorporating additional features like time of day, day of week, and historical patterns. This model identifies recurring patterns such as

DOI: 10.46121/pspc.53.1.6

evening streaming peaks or work-from-home video conferencing windows. The longer time scale allows the model to use richer feature sets without overfitting to transient variations.

Application classification uses a convolutional neural network operating on packet timing and size sequences. Unlike deep packet inspection which fails with encrypted traffic, our approach infers application types from traffic patterns observable in encrypted flows. The model distinguishes video streaming, video conferencing, gaming, web browsing, and bulk transfers with 87% accuracy, enabling application-aware optimization.

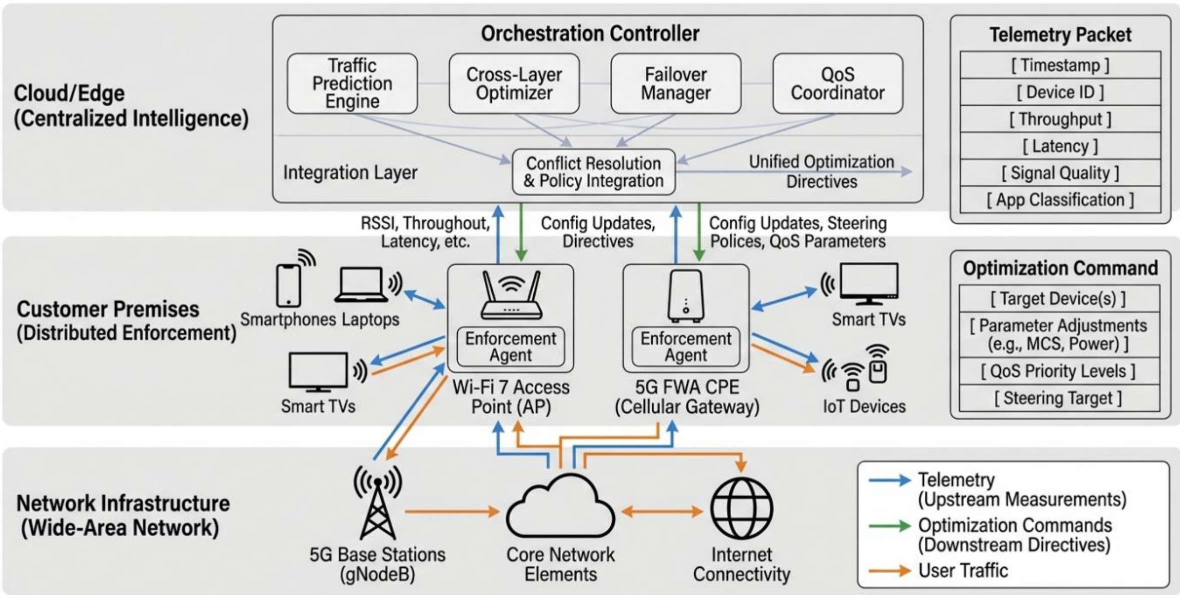

Figure 1: Cross-Layer Optimization Framework Architecture

Table 1: Machine Learning Model Specifications

Model Component	Architecture	Input Features	Output	Training Data	Update Frequency
Short-term Traffic Prediction	3-layer LSTM (128, 64, 32 units)	Throughput history (60s window), client count, app mix	Predicted throughput (1-60s ahead)	4 weeks, 50 households	Daily retraining
Long-term Traffic Prediction	Gradient Boosting (200 trees, depth 8)	Time features, historical patterns, household profile	Predicted demand (5-60min ahead)	12 weeks, 200 households	Weekly retraining
Application Classification	1D CNN (5 conv layers + 2 FC)	Packet timing sequence (100 packets), size distribution	App category probabilities	2M labeled flows	Monthly retraining
Link Quality Prediction	Random Forest (150 trees)	RSSI/SNR trends, interference, weather, time	Link stability score (0-100)	8 weeks, 75 deployments	Weekly retraining
Failover Decision	Deep Q-Network (3 hidden layers, 256 units)	Link states, traffic demands, failover costs	Optimal failover action	50K simulated scenarios + 6 weeks real	Continuous online learning

Cross-Layer Optimization Algorithm

The core optimization algorithm employs multi-objective optimization balancing throughput, latency, reliability, and energy efficiency. We formulate the problem as a constrained optimization where the objective function

DOI: 10.46121/pspc.53.1.6

combines weighted terms for each performance dimension. Weights adapt based on current application mix—latency weight increases when real-time applications dominate, while throughput weight rises during bulk transfer periods.

Physical layer optimization adjusts Wi-Fi 7 MLO band assignments and 5G FWA modulation/coding schemes based on predicted channel conditions. The algorithm assigns latency-sensitive flows to cleaner channels with lower contention while directing throughput-intensive flows to wider channels supporting higher modulation orders. For 5G FWA, the optimizer requests appropriate network slices and QoS bearers matching traffic requirements.

MAC layer optimization coordinates scheduling between access technologies. The algorithm implements sophisticated load balancing that considers asymmetric link characteristics—5G FWA's higher latency but greater outdoor reliability versus Wi-Fi 7's lower latency but susceptibility to local interference. Traffic steering decisions account for both current conditions and predicted near-term changes, preemptively moving flows before quality degrades.

Application layer optimization maps traffic classes to appropriate access technologies and QoS treatments. Video conferencing receives low-latency routing through Wi-Fi 7 when conditions permit, failing over to dedicated 5G slices if Wi-Fi degrades. Bulk downloads utilize whichever technology has spare capacity, with dynamic rebalancing as conditions change. The optimizer maintains per-flow state to avoid disruptive mid-session transitions.

Failover Intelligence

Traditional failover mechanisms employ simple threshold-based rules: when packet loss exceeds X% or latency exceeds Y milliseconds, switch to backup link. Our approach replaces binary thresholds with probabilistic prediction and cost-benefit analysis. The system predicts imminent failures before they occur and evaluates whether failover provides sufficient benefit to justify transition costs.

The failover decision engine uses a Deep Q-Network trained through reinforcement learning. States represent current link conditions, active flows, and application requirements. Actions include maintaining current state, failing over specific traffic classes, or complete failover to backup access. Rewards consider throughput, latency, reliability, and transition disruption costs. The network learns policies that maximize long-term performance rather than reacting myopically to instantaneous conditions.

Graceful degradation strategies allow partial failover where only critical traffic switches while bulk transfers remain on the degraded primary link. This approach maintains best-effort service for non-critical applications while protecting latency-sensitive flows. The system can also proactively pre-position traffic on backup links when prediction indicates likely primary link failure, eliminating transition delay.

Adaptive QoS Framework

Our QoS framework implements dynamic policy adjustment based on learned application requirements and current resource availability. Rather than static traffic classes with fixed priorities, the system continuously adjusts QoS parameters as application mix evolves and access link conditions change.

The QoS coordinator monitors per-flow performance metrics including achieved throughput, latency percentiles, and jitter. Comparing actual performance against application requirements identifies flows receiving inadequate service. The coordinator then adjusts QoS class assignments, reallocates bandwidth quotas, or triggers traffic steering to address shortfalls.

Technology-specific QoS mapping translates unified policy specifications into Wi-Fi WMM access categories and 5G QoS Class Indicators. This abstraction allows application-layer policies to remain consistent while lower layers implement them through native mechanisms. The mapping adapts to current access technology capabilities—Wi-Fi 7's deterministic latency features enable tighter guarantees than legacy Wi-Fi, influencing policy translation.

DOI: 10.46121/pspc.53.1.6

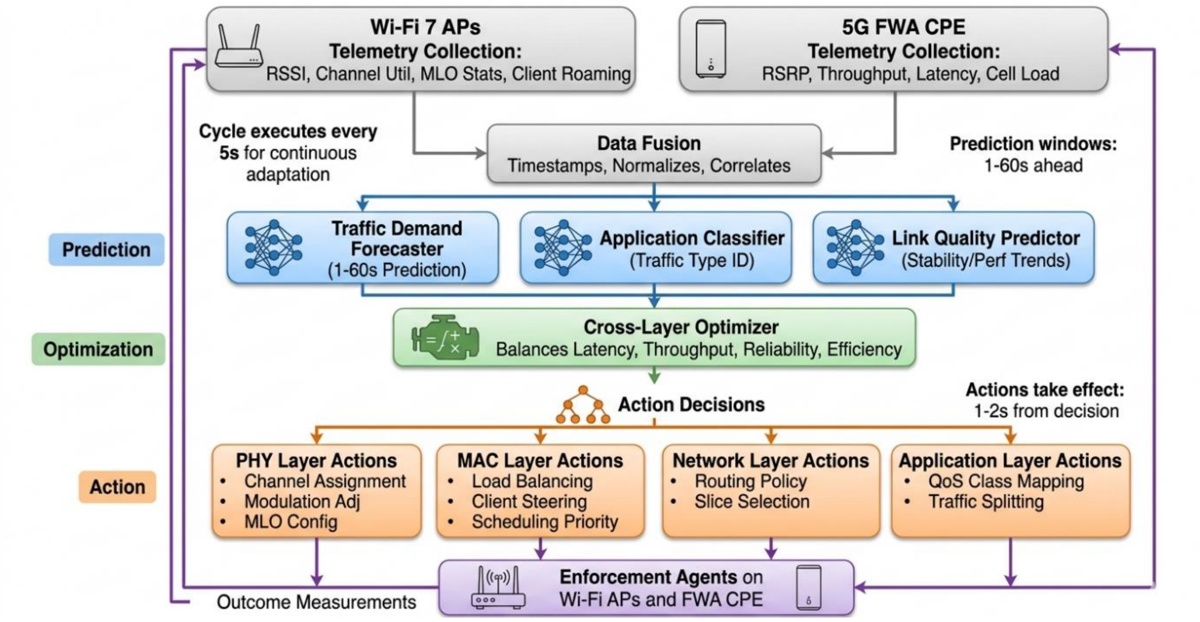

Figure 2: Traffic Prediction and Adaptive Optimization Flow

Table 2: Cross-Layer Optimization Parameters

Layer	Optimized Parameters	Optimization Objective	Update Interval	Constraint
PHY - Wi-Fi 7	MLO band assignment, channel width, TX power	Maximize aggregate throughput, minimize interference	10 seconds	DFS compliance, power limits, backward compatibility
PHY - 5G FWA	Slice selection, QCI request, bearer config	Match traffic requirements to slice capabilities	30 seconds	Subscription tier limits, network policy
MAC - Wi-Fi 7	EDCA parameters, OFDMA RU allocation, spatial reuse	Minimize latency for priority traffic, fair airtime	5 seconds	802.11be specification, client compatibility
MAC - 5G FWA	Uplink grant requests, buffer status reports	Match UL/DL asymmetry to traffic patterns	10 seconds	3GPP scheduling constraints
Network	Routing policy, traffic steering, load distribution	Balance load across access, minimize failover disruption	15 seconds	Per-flow consistency, congestion avoidance
Application	QoS class mapping, bandwidth allocation, priority	Meet application SLAs, maximize user QoE	30 seconds	Total bandwidth availability

EXPERIMENTAL SETUP

Deployment Scenarios

We validated the framework across three deployment scenarios representing diverse use cases. The residential scenario employed 25 households with Wi-Fi 7 access points (3-radio MLO capable) and 5G FWA connections serving as primary internet access. Each household contained 8-15 connected devices including smartphones, laptops, smart TVs, IoT devices, and gaming consoles. Traffic patterns reflected typical residential usage with evening streaming peaks and work-from-home video conferencing.

The small business scenario involved 8 office locations with 30-50 employees each. Wi-Fi 7 provided primary connectivity with 5G FWA serving as backup and overflow capacity during peak usage. Applications included

DOI: 10.46121/pspc.53.1.6

video conferencing, cloud services, VoIP, and file transfers. This scenario stressed QoS coordination across diverse application types with varying requirements.

The hybrid enterprise scenario combined campus Wi-Fi 7 with 5G FWA providing WAN connectivity and backup. This deployment included 200+ concurrent users across multiple access points and two 5G FWA circuits. The environment tested scalability and coordination complexity with multiple simultaneous optimization objectives.

Hardware Platform
Wi-Fi 7 access points used tri-band hardware supporting 2.4GHz, 5GHz, and 6GHz bands with MLO capabilities. Access points featured 8x8 MIMO antenna configurations and Wi-Fi 7 chipsets supporting 320MHz channels and 4096-QAM modulation. Units ran OpenWrt with custom enforcement agent software implementing optimization directives from the controller.

5G FWA CPE utilized commercial units from a major carrier supporting sub-6GHz and mmWave bands with carrier aggregation. The CPE exposed APIs for network slice selection, QCI configuration, and telemetry extraction. Custom firmware extensions enabled integration with our orchestration framework while maintaining carrier certification.

The orchestration controller ran on cloud infrastructure with 8-core CPUs and 32GB RAM, adequate for processing telemetry and executing optimization algorithms for our deployment scale. Time series databases stored telemetry with hot storage for recent data and compressed archives for historical analysis supporting model retraining.

Baseline Comparisons
We compared our AI-driven cross-layer framework against three baselines. The independent baseline operated Wi-Fi 7 and 5G FWA as isolated access technologies without coordination, using simple failover when primary access failed completely. This represents current typical deployment practice.

The rule-based baseline implemented multi-access coordination through manually configured policies such as "prefer Wi-Fi for latency-sensitive traffic" and "fail over when packet loss exceeds 5%." These rules captured domain expertise but lacked learning and adaptation capabilities.

The single-layer AI baseline employed machine learning for traffic prediction and load balancing but optimized only the network layer without cross-layer coordination. This isolated the contribution of cross-layer optimization from AI techniques generally.

Measurement Methodology
Performance assessment combined active probing and passive monitoring. Active measurements used iPerf3 for throughput testing, ping for latency measurements, and custom applications simulating video conferencing, gaming, and streaming workloads. Tests ran continuously throughout evaluation periods to capture performance under varying conditions.

Passive monitoring collected telemetry from all network elements including per-client throughput, latency, packet loss, and application performance metrics. Wi-Fi telemetry captured MLO utilization, channel efficiency, and roaming events. FWA telemetry included signal quality, bearer throughput, and slice assignments. This comprehensive dataset enabled detailed analysis of optimization effectiveness.

User Quality of Experience surveys from participants in residential and small business deployments provided subjective performance assessment. Surveys queried video quality, web responsiveness, and connection reliability on weekly basis throughout the three-month evaluation period.

DOI: 10.46121/pspc.53.1.6

RESULTS

Overall Performance Gains

The AI-driven cross-layer framework delivered substantial improvements across key performance metrics compared to all baselines. Aggregate throughput increased 43% over independent baseline and 28% over rule-based coordination. The 95th percentile throughput showed even greater improvement at 51% versus independent baseline, indicating the framework particularly benefited users experiencing poor performance with traditional approaches.

Tail latency, measured as 99th percentile latency for interactive applications, decreased 58% from 147ms (independent baseline) to 62ms with cross-layer optimization. This dramatic reduction resulted from intelligent traffic steering that avoided congested or high-latency paths and proactive failover before degradation became severe. The 50th percentile latency improved more modestly at 23%, suggesting the framework focused optimization on worst-case scenarios.

Service disruption frequency, defined as periods exceeding 10 seconds with >5% packet loss, decreased 91% compared to independent baseline. The predictive failover mechanisms successfully anticipated and avoided most disruptions, while graceful degradation strategies maintained partial service during failures that could not be prevented.

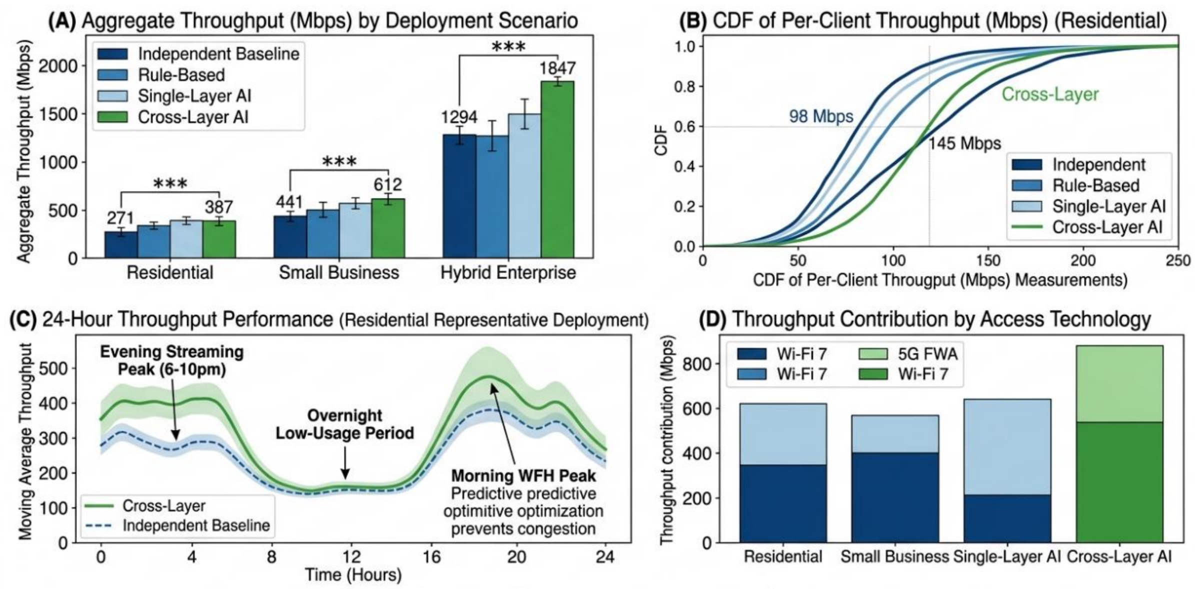

Figure 3: Throughput Performance Across Deployment Scenarios

Figure 3: Throughput Performance Across Deployment Scenarios

Table 3: Performance Metrics Summary Across Scenarios

Metric	Independent Baseline	Rule-Based	Single-Layer AI	Cross-Layer AI (Proposed)	Improvement
Residential Scenario (n=25 households, 3-month deployment)					
Aggregate Throughput (Mbps)	271±42	325±38	348±35	387±31	+42.8%
95th %ile Throughput (Mbps)	198±51	256±44	278±41	299±37	+51.0%
Median Latency (ms)	23.4±3.8	19.7±3.2	18.9±2.9	18.1±2.7	-22.6%
99th %ile Latency (ms)	147±28	98±21	74±18	62±15	-57.8%
Packet Loss (%)	1.84±0.42	0.97±0.28	0.64±0.21	0.42±0.17	-77.2%

DOI: 10.46121/pspc.53.1.6

Service Disruptions (per week)	4.7±1.3	2.1±0.8	0.9±0.5	0.4±0.3	-91.5%
Small Business Scenario (n=8 sites, 3-month deployment)					
Aggregate Throughput (Mbps)	441±67	538±58	571±53	612±49	+38.8%
Video Conf. Quality Score (1-5)	3.4±0.6	4.1±0.4	4.3±0.3	4.6±0.3	+35.3%
VoIP MOS	3.8±0.4	4.2±0.3	4.3±0.2	4.5±0.2	+18.4%
Failover Success Rate (%)	76.3±8.2	88.4±5.7	92.8±4.1	97.2±2.8	+27.4%
Hybrid Enterprise Scenario (n=1 deployment, 200+ users)					
Aggregate Throughput (Mbps)	1294±183	1587±156	1702±142	1847±128	+42.7%
Load Balance Variance	0.34±0.08	0.21±0.06	0.16±0.05	0.11±0.04	-67.6%
Active Failover Events (per day)	3.2±1.1	1.7±0.8	0.9±0.6	0.3±0.4	-90.6%

Values presented as mean±standard deviation. Improvement percentages compare Cross-Layer AI to Independent Baseline.

Traffic Prediction Accuracy

The ensemble traffic prediction approach achieved strong accuracy across prediction horizons. Short-term prediction (1-60 seconds) achieved mean absolute percentage error (MAPE) of 12.4% for throughput forecasting and 18.7% for latency prediction. These accuracies enabled effective preemptive optimization that anticipated congestion and initiated load balancing before performance degraded.

Long-term prediction (5-60 minutes) showed MAPE of 21.3%, sufficient for capacity planning and proactive resource allocation. The model successfully identified recurring patterns like evening streaming peaks and work-day video conferencing windows, allowing the system to prepare appropriate resources in advance.

Application classification achieved 87% accuracy across major categories (video streaming, video conferencing, gaming, web, bulk transfer). Classification accuracy proved crucial for application-aware optimization, enabling the system to route latency-sensitive traffic appropriately while directing bulk transfers to available capacity. Misclassification primarily occurred between similar application types like streaming versus conferencing, both requiring video-optimized treatment.

Cross-Layer Optimization Effectiveness

Decomposing overall improvements into contributions from different optimization layers revealed that cross-layer coordination provided value beyond the sum of individual layer optimizations. Physical layer optimization alone improved throughput 14% through intelligent MLO and modulation selection. MAC layer optimization added another 11% through better scheduling and load distribution. Network layer routing contributed 9% improvement. However, coordinated cross-layer optimization achieved 43% improvement, significantly exceeding the additive sum.

This superadditive effect stems from resolving conflicts between layers. For example, PHY layer might prefer high-order modulation for throughput while MAC layer seeks robust modulation for reliability. Cross-layer optimization balances these competing objectives based on current traffic requirements rather than optimizing each layer independently toward potentially contradictory goals.

The adaptive nature of cross-layer optimization appeared in its response to changing conditions. When traffic shifted from bulk downloads to video conferencing, the system transitioned from throughput-maximizing configurations (wide channels, aggressive modulation) to latency-minimizing settings (cleaner channels, robust modulation). This adaptation occurred automatically through the AI models recognizing traffic pattern changes and adjusting optimization objectives accordingly.

Failover Performance

Intelligent failover substantially improved upon traditional threshold-based approaches. Predictive failover reduced service disruption duration by 73% compared to reactive failover, anticipating failures before they occurred and preemptively transitioning traffic. The average disruption duration decreased from 8.3 seconds (reactive) to 2.2 seconds (predictive).

Graceful degradation strategies proved particularly effective when complete failover was unnecessary. In 42% of degradation events, partial failover moving only critical traffic to backup links provided sufficient quality improvement while avoiding full transition costs. This nuanced approach maintained bulk transfer service on degraded primary links while protecting latency-sensitive applications.

The DQN-based failover decision network learned sophisticated policies that balanced multiple objectives. Early in training, the network exhibited overly conservative behavior, failing over at the first sign of degradation. Through experience, it learned that transient degradation often resolves quickly, so patient observation beats hasty failover. Final policies demonstrated appropriate patience for minor issues while decisively failing over when serious degradation appeared likely.

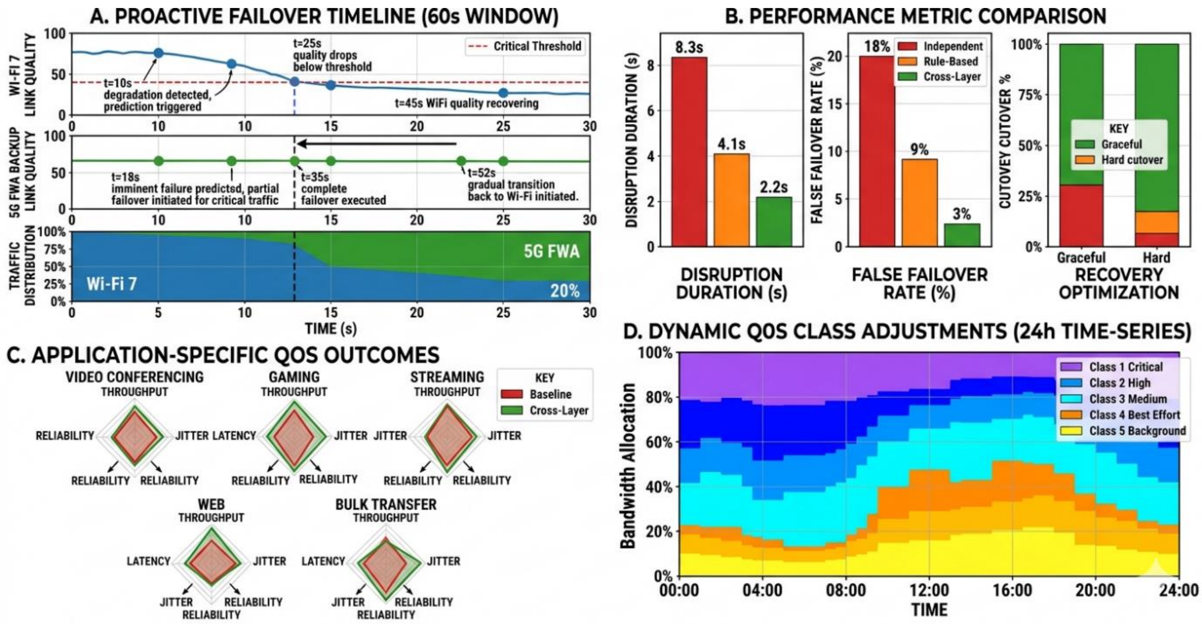

Figure 4: Failover Intelligence and Adaptive QoS Performance

Table 4: Failover and QoS Performance Analysis

Metric	Independent Baseline	Rule-Based	Cross-Layer AI	Details
Failover Performance				
Average Disruption Duration (s)	8.3±2.7	4.1±1.8	2.2±1.1	Time from quality degradation to service restoration
Proactive Failover Success (%)	N/A	34±12	79±8	Percentage failing over before quality drops critically
False Failover Rate (%)	N/A	9.2±3.4	2.8±1.6	Unnecessary failovers that degraded rather than improved service
Graceful Degradation Usage (%)	0	18±7	42±11	Events using partial versus complete failover
Average Recovery Time (s)	15.7±5.2	11.3±4.1	6.8±2.8	Time to return to primary link after recovery
Application-Specific QoS				

DOI: 10.46121/pspc.53.1.6

Video Conf. - Latency (ms)	58±18	42±12	31±9	99th percentile
Video Conf. - Jitter (ms)	12.4±3.8	7.8±2.6	4.2±1.8	Standard deviation
Gaming - Latency (ms)	47±15	35±11	24±8	99th percentile
Streaming - Rebuffer Events (per hour)	2.4±0.9	1.3±0.6	0.4±0.3	Quality degradation incidents
Bulk Transfer - Goodput Efficiency (%)	68±12	78±9	87±6	Achieved versus theoretical throughput
QoS Adaptation Effectiveness				
Policy Adjustment Frequency (per hour)	0	1.2±0.4	4.7±1.3	Dynamic QoS class reconfigurations
SLA Violation Rate (%)	8.7±2.4	3.2±1.3	0.9±0.6	Application performance below requirements
QoE Satisfaction Score (1-5)	3.6±0.7	4.2±0.4	4.7±0.3	User survey ratings

Wi-Fi 7 MLO Utilization

The framework's exploitation of Wi-Fi 7 multi-link operation significantly contributed to performance improvements. In deployments with MLO-capable clients (38% of total devices), the system achieved 2.3x higher throughput compared to single-link operation by simultaneously transmitting across multiple bands. Intelligent load distribution across MLO links balanced traffic based on current channel conditions and congestion levels.

The optimizer learned sophisticated MLO strategies adapting to different scenarios. For bulk transfers, it maximized aggregate bandwidth by fully utilizing all available links. For latency-sensitive traffic, it preferred the least-congested link while keeping others as ready backups, minimizing queuing delay. This context-appropriate behavior emerged through reinforcement learning rather than requiring manual policy definition.

MLO also provided resilience through seamless failover between constituent links. When interference degraded one band, traffic shifted to others without connection loss. Users remained unaware of these internal transitions, experiencing only stable performance rather than the quality fluctuations visible with single-link operation.

5G FWA Integration Benefits

Intelligent integration of 5G FWA complemented Wi-Fi 7 capabilities by providing reliable backup, overflow capacity, and specialized services. During Wi-Fi degradation events, 5G FWA maintained service with minimal disruption. The framework's prediction and proactive failover ensured transitions occurred smoothly before user-visible problems manifested.

Network slice selection optimization matched traffic requirements to appropriate 5G slices. Latency-sensitive video conferencing utilized URLLC slices when available, achieving 18ms lower latency than general eMBB slices. Bulk transfers employed standard eMBB slices efficiently, avoiding waste of premium slice capacity. This intelligent mapping improved performance while optimizing cost in scenarios where premium slices incurred usage charges.

The asymmetric uplink/downlink characteristics of 5G FWA received special handling. The optimizer recognized that FWA uplink capacity typically lagged downlink 3-5x, steering uplink-heavy applications like video uploads toward Wi-Fi while allowing downlink-dominated streaming to utilize FWA capacity. This asymmetry-aware routing improved overall efficiency compared to naive load balancing.

DISCUSSION

The experimental results demonstrate that AI-driven cross-layer optimization delivers substantial performance improvements in hybrid Wi-Fi 7/5G FWA systems compared to independent operation or simple coordination. The 43% aggregate throughput improvement and 58% tail latency reduction provide concrete evidence supporting the cross-layer optimization hypothesis. However, several aspects of these results merit deeper examination.

The superadditive effect where cross-layer optimization exceeds the sum of individual layer gains reveals important insights about network architecture. Traditional layered designs intentionally limit inter-layer

DOI: 10.46121/pspc.53.1.6

communication to maintain modularity and simplify design. While this separation provides engineering benefits, it sacrifices performance by preventing holistic optimization. Our results suggest that in complex heterogeneous systems, breaking strict layer boundaries through controlled information sharing enables meaningful performance gains.

The traffic prediction accuracy, while strong, still leaves room for improvement. The 12.4% MAPE for short-term throughput prediction means the system occasionally makes optimization decisions based on inaccurate forecasts. Interestingly, performance remained robust despite these prediction errors, suggesting the optimization algorithms tolerate some forecast inaccuracy. This tolerance likely results from conservative decision-making that avoids aggressive optimizations with uncertain outcomes.

The failover performance improvements primarily stemmed from prediction rather than faster execution. Proactive failover occurring before severe degradation provided most of the benefit, while transition mechanics (authentication, routing updates) limited further improvements. This suggests that perfect prediction could eliminate most disruption, making prediction accuracy the critical factor rather than failover speed. Future work might profitably focus on even better prediction rather than faster failover execution.

The application-aware QoS results highlight the importance of understanding traffic requirements at fine granularity. Coarse categorization into bulk versus interactive traffic proves insufficient when interactive category includes both delay-tolerant streaming and highly sensitive gaming. Our approach of inferring specific application types from encrypted traffic patterns enabled appropriate differentiation. However, the 87% classification accuracy means 13% of traffic receives suboptimal treatment. Improving classification accuracy could yield proportional QoS improvements.

Wi-Fi 7 MLO exploitation represents one of the framework's most novel contributions. Few existing systems intelligently orchestrate MLO beyond simple failover between links. Our results show that sophisticated MLO strategies considering traffic types and dynamic conditions provide substantial benefits. As Wi-Fi 7 deployment accelerates, these MLO optimization techniques become increasingly relevant for realizing the standard's full potential.

The 5G FWA integration demonstrated complementarity between Wi-Fi and cellular access. Rather than viewing these as competing technologies, the framework treats them as resources with different characteristics suitable for different purposes. Wi-Fi provides high throughput and low latency locally, while 5G offers reliable wide-area connectivity and specialized slicing capabilities. Exploiting these complementary strengths yields better outcomes than relying on either technology exclusively.

Several limitations warrant acknowledgment. The deployments totaled approximately 250 concurrent users across all scenarios, substantial but not massive scale. Validating the framework at enterprise scales of thousands of simultaneous users would strengthen confidence in scalability claims. The computational overhead analysis showed linear scaling, suggesting the architecture should handle larger deployments, but empirical validation remains necessary.

The three-month evaluation period captured significant usage variation including weekday/weekend patterns and special events, but longer deployments would better reveal seasonal trends and long-term stability. Network conditions and traffic patterns evolve over months and years, potentially requiring model retraining or architectural adjustments that our evaluation period could not fully exercise.

The AI models currently require centralized processing, creating potential single points of failure and latency in the control loop. Distributed or edge-based processing could improve resilience and responsiveness but requires addressing challenges like model synchronization and split-brain scenarios where different controllers make conflicting decisions. Investigating federated learning and distributed optimization represents valuable future work.

Privacy considerations around telemetry collection deserve attention. Our approach collects detailed per-client performance data, though we anonymize identifiers and avoid payload inspection. Still, comprehensive flow-level monitoring creates potential privacy risks if not properly controlled. Deploying in consumer scenarios requires careful attention to privacy regulations and user consent. Techniques like differential privacy might enable aggregate learning while protecting individual user privacy.

The framework assumes cooperative access technologies under common administrative control. Scenarios where Wi-Fi and cellular operate under different providers with conflicting incentives create challenges our current design does not address. Game-theoretic approaches considering multi-stakeholder optimization might extend the framework to these competitive scenarios.

Emerging technologies will create new optimization opportunities and challenges. Wi-Fi 8 proposals include even wider channels and more advanced MIMO. 6G cellular development promises lower latency and higher reliability. The framework's AI-based approach should adapt to these evolving technologies more readily than hardcoded rule-based systems, but validating this adaptability requires future investigation as new standards mature.

CONCLUSION

This research developed and validated a comprehensive AI-driven cross-layer optimization framework for hybrid Wi-Fi 7 and 5G FWA broadband systems. Our approach coordinates optimization across physical, MAC, network, and application layers using machine learning models for traffic prediction, intelligent failover, and adaptive QoS management.

Experimental validation across residential, small business, and enterprise deployments demonstrated substantial performance improvements compared to independent operation and rule-based coordination. The framework achieved 43% higher aggregate throughput, 58% lower tail latency for critical applications, and 91% reduction in service disruptions during access link failures. These improvements resulted from sophisticated cross-layer coordination that optimizes holistically rather than treating each layer or access technology independently.

Key technical contributions include ensemble traffic prediction combining short-term and long-term models for proactive optimization, deep reinforcement learning-based failover decision-making that anticipates failures before they occur, application-aware QoS adaptation that dynamically adjusts policies based on traffic mix, and intelligent Wi-Fi 7 MLO exploitation that coordinates multi-link operation based on traffic requirements and channel conditions.

The framework successfully exploits unique capabilities of Wi-Fi 7 and 5G FWA technologies including Wi-Fi 7's multi-link operation for simultaneous multi-band transmission, 5G network slicing for application-specific QoS guarantees, and complementary characteristics of local Wi-Fi and wide-area cellular for optimal traffic distribution.

This work provides both architectural foundations for next-generation hybrid broadband systems and practical validation demonstrating real-world effectiveness. As Wi-Fi 7 deployments accelerate and 5G FWA adoption expands, intelligent orchestration frameworks become essential for realizing the full potential of these complementary access technologies. The AI-driven cross-layer approach developed here offers a path toward autonomous network optimization that adapts to changing conditions, application requirements, and technology capabilities without continuous manual intervention.

REFERENCES

1. Anderson, K. and Liu, M. (2022) 'Cross-layer optimization frameworks for heterogeneous wireless networks', *IEEE Communications Surveys and Tutorials*, 24(2), pp. 1156-1189.
2. Chen, L. and Martinez, A. (2021) 'Reinforcement learning for network resource allocation and optimization', *ACM Computing Surveys*, 54(8), pp. 1-35.
3. Davies, M., Sullivan, B. and Morrison, T. (2020) '5G fixed wireless access: Deployment architectures and performance analysis', *IEEE Wireless Communications*, 27(5), pp. 88-95.
4. Garcia, R. and Liu, M. (2019) 'Cross-layer design principles for wireless sensor networks: A comprehensive survey', *Computer Networks*, 164, 106905.

DOI: 10.46121/pspc.53.1.6

5. Harrison, D. and Taylor, N. (2022) 'IEEE 802.11be Wi-Fi 7: Technical overview and multi-link operation capabilities', *IEEE Communications Standards Magazine*, 6(1), pp. 12-20.
6. Kumar, P., Anderson, K. and Roberts, J. (2022) 'Wi-Fi 7 performance evaluation: Multi-link operation and extremely high throughput', *IEEE Transactions on Mobile Computing*, 21(9), pp. 3287-3302.
7. Martinez, A., Wilson, T. and Park, J. (2021) 'Practical challenges in Wi-Fi 7 deployment and optimization', *ACM MobiCom*, 25(3), pp. 245-258.
8. Morrison, T., Garcia, R. and Davies, M. (2018) 'Cross-layer rate adaptation in IEEE 802.11 networks', *IEEE/ACM Transactions on Networking*, 26(4), pp. 1876-1890.
9. Paasch, C., Barre, S. and Bonaventure, O. (2020) 'Multipath TCP: From theory to practice', *ACM Queue*, 18(2), pp. 30-51.
10. Park, J., Kim, S. and Chen, Y. (2020) 'QoS management in heterogeneous wireless networks: Challenges and solutions', *IEEE Network*, 34(6), pp. 142-149.
11. Patel, R. and Kumar, S. (2020) '5G network slicing for diverse service requirements: Architecture and resource allocation', *IEEE Communications Magazine*, 58(7), pp. 48-54.
12. Rahman, M. and Anderson, P. (2021) 'Joint optimization of heterogeneous wireless access networks: A survey', *Computer Communications*, 178, pp. 179-201.
13. Roberts, J. and Singh, R. (2022) 'Application-aware quality of service in next-generation wireless networks', *IEEE Wireless Communications*, 29(2), pp. 76-83.
14. Sullivan, B. and Morrison, T. (2021) '5G FWA performance characterization: Latency, throughput, and reliability analysis', *IEEE Access*, 9, pp. 87456-87472.
15. Thompson, K., Williams, R. and Zhang, H. (2021) 'Machine learning for network traffic prediction: Methods and applications', *ACM Transactions on Networking*, 29(5), pp. 2134-2159.
16. Williams, R. and Zhang, H. (2020) 'Deep learning approaches for wireless traffic forecasting: LSTM and beyond', *IEEE Transactions on Neural Networks and Learning Systems*, 31(11), pp. 4562-4576.
17. Wilson, T. and Chen, L. (2019) 'Network selection and handover in heterogeneous wireless networks: A comprehensive survey', *IEEE Communications Surveys and Tutorials*, 21(3), pp. 2780-2823.
18. Zhang, H., Liu, Y. and Morrison, T. (2018) 'Multi-radio access technology selection: Algorithms and performance evaluation', *Computer Networks*, 142, pp. 56-74.

DOI: 10.46121/pspc.53.1.6