ISSN: 2519-9439

Online Reinforcement Policies for Weather-Triggered Vehicular Link Migration Decisions

George Yim

Department of Computing, Faculty of Computer and Mathematical Sciences, Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China

Abstract

The rapid evolution of connected and autonomous vehicles relies heavily on robust, high-capacity communication links to support safety-critical applications and high-bandwidth infotainment systems. While millimeter-wave and terahertz frequency bands offer unprecedented data rates, they suffer from severe susceptibility to dynamic meteorological conditions, particularly precipitation, fog, and atmospheric absorption. Traditional rule-based vertical handover mechanisms often fail to adapt to the highly dynamic and unpredictable nature of sudden weather variations, leading to suboptimal link performance, frequent disconnections, and unacceptable latency spikes. This paper proposes a novel framework utilizing online reinforcement learning to optimize weather-triggered vehicular link migration decisions in real-time. By formulating the communication environment as a Markov Decision Process, the proposed system enables a learning agent to continuously observe local meteorological data, current link quality indicators, and vehicular mobility states to make proactive link migration decisions between high-frequency and highly reliable sub-6 gigahertz backup bands. Unlike offline learning models that struggle with unseen weather anomalies, the online reinforcement learning policy continuously updates its action-value estimations, balancing the trade-off between maximizing instantaneous throughput and minimizing the signaling overhead associated with frequent handovers. Extensive experimental evaluations using synthesized meteorological traces and mobility datasets demonstrate that the proposed online policy significantly reduces packet drop rates and maintains higher average throughput during severe weather events compared to baseline static threshold methods and offline learning approaches.

Keywords: Vehicular Networks, Reinforcement Learning, Link Migration, Weather Attenuation, Millimeter Wave

1.Introduction

The paradigm of intelligent transportation systems and autonomous driving has fundamentally transformed the requirements for vehicular communication networks. Modern connected vehicles are no longer merely endpoints for basic telematics; they operate as mobile computing hubs that generate, process, and transmit vast amounts of data. Applications such as cooperative adaptive cruise control, raw sensor data sharing, high-definition map downloading, and extended reality infotainment demand communication links with ultra-low latency and multi-gigabit per second throughput. To satisfy these stringent requirements, the telecommunications industry has increasingly integrated high-frequency bands, specifically millimeter-wave and low terahertz frequencies, into vehicular communication standards. These high-frequency bands provide the massive bandwidth

ISSN: 2519-9439

necessary to support the next generation of vehicular applications. However, this transition introduces significant physical layer challenges, primarily due to the severe propagation characteristics of high-frequency electromagnetic waves. Unlike traditional sub-6 gigahertz bands, millimeter-wave signals experience high path loss, poor penetration through solid objects, and extreme sensitivity to atmospheric conditions, which severely limits their reliability in dynamic outdoor environments. Related work has addressed full-Duplex NOMA-Enabled integrated sensing and communication: joint transmit and receive beamforming optimization [1]. Among the various environmental factors impacting high-frequency vehicular communications, meteorological conditions represent one of the most volatile and disruptive elements. Atmospheric precipitation, including rain, snow, and hail, causes severe signal attenuation due to absorption and scattering effects. Rain fade is particularly detrimental, as the physical size of raindrops is comparable to the wavelength of millimeter-wave signals, leading to intense Mie scattering. Furthermore, suspended water droplets in heavy fog can severely degrade signal-to-noise ratios, rapidly reducing the effective communication range of roadside units and vehicular transceivers. In a typical urban or highway vehicular environment, weather conditions can change rapidly and unpredictably. A vehicle traversing a highway might suddenly encounter localized micro-weather phenomena, such as an intense rain squall or a dense patch of fog, causing its primary millimeter-wave communication link to degrade precipitously within seconds. Such rapid degradation can lead to catastrophic communication failures, severely compromising the safety and efficacy of cooperative autonomous driving systems.

To mitigate the unreliability of high-frequency bands, modern vehicular networks typically employ heterogeneous communication architectures, integrating multiple frequency bands and radio access technologies. In such architectures, an active communication session can be migrated from a high-frequency, high-capacity band to a lower-frequency, high-reliability band when environmental conditions deteriorate. This process, often referred to as vertical handover or link migration, is critical for maintaining session continuity and meeting quality of service guarantees. However, deciding exactly when and how to perform this migration is a complex optimization problem. If a vehicle migrates to a lower frequency band too early, it unnecessarily sacrifices the massive bandwidth available in the millimeter-wave band, leading to network congestion on the lower frequency spectrum and suboptimal application performance. Conversely, if the vehicle migrates too late, the millimeter-wave link may fail completely, resulting in dropped packets, disrupted safety services, and high latency as the system scrambles to re-establish a connection.

Traditional approaches to link migration in vehicular networks have predominantly relied on static rule-based algorithms or simple threshold-based triggers. For instance, a system might be programmed to initiate a handover to a sub-6 gigahertz band whenever the received signal strength indicator of the millimeter-wave link falls below a pre-defined threshold. While simple to implement, these static policies are fundamentally ill-equipped to handle the immense variability of vehicular environments. A fixed threshold cannot account for the trajectory of the vehicle, the varying intensity of a storm, or the differing quality of service requirements of active applications. Moreover, rule-based systems often suffer from the ping-pong effect, where fluctuating signal strengths cause the system to rapidly and repeatedly switch between bands, generating excessive signaling overhead and degrading overall throughput. As the complexity of heterogeneous networks increases, it has become evident that dynamic, intelligent decision-making frameworks are required to manage weather-triggered link migrations effectively.

ISSN: 2519-9439

In recent years, artificial intelligence and machine learning have emerged as powerful tools for optimizing complex network management tasks. Reinforcement learning, in particular, offers a promising paradigm for dynamic link migration, as it enables an artificial agent to learn optimal decision-making policies through continuous interaction with the environment. Asynchronous deep reinforcement learning provides a training reference for policies updated from multiple interacting environments [2]. By observing the current state of the network, taking actions, and receiving numerical rewards based on the outcomes, a reinforcement learning agent can discover sophisticated migration strategies that balance multiple conflicting objectives. However, the majority of existing research applying reinforcement learning to vehicular communications has focused on offline training methodologies. In an offline approach, the agent is trained extensively on historical datasets or within simulated environments, and the resulting static policy is deployed to the vehicles. While offline policies can achieve reasonable performance under conditions that closely match their training data, they exhibit profound fragility when exposed to novel situations, such as unprecedented weather anomalies, unexpected traffic congestion patterns, or newly deployed network infrastructure. Yang and colleagues examine implicit constraint approach for offline Multi-Agent reinforcement learning [3].

To address the limitations of static rules and offline learning, this paper proposes a comprehensive online reinforcement learning policy specifically designed for weather-triggered vehicular link migration decisions. Online reinforcement learning distinguishes itself by continuously updating its internal model and policy while operating in the live environment. This continuous adaptation allows the agent to dynamically respond to shifting weather patterns, changing network topologies, and varying application demands in real-time. The core contribution of this work lies in the formulation of a highly customized Markov Decision Process that explicitly incorporates real-time meteorological sensor data into the state space, alongside traditional networking metrics and vehicular mobility parameters. By providing the learning agent with direct visibility into the environmental conditions causing signal degradation, the proposed system can proactively initiate link migrations before catastrophic communication failures occur.

Furthermore, this paper introduces a novel reward function that carefully balances the maximization of long-term application throughput against the minimization of migration penalties. Link migration is not a cost-free operation; it incurs significant signaling overhead, temporary data buffering, and brief interruptions in transmission. The proposed reward mechanism intelligently penalizes unnecessary band switching, effectively mitigating the ping-pong effect without relying on arbitrary time-to-trigger timers. The online nature of the algorithm ensures that this balance is continuously refined based on the immediate feedback from the network environment. Through the deployment of a deep Q-network architecture tailored for online updates, the proposed framework achieves high computational efficiency, making it suitable for implementation on standard vehicular onboard units and edge computing nodes. Related work has addressed nonvolatile optical switch with a programmable pn heterojunction [4].

The remainder of this comprehensive academic paper is structured as follows. Section 2 provides an exhaustive review of the background and related literature, detailing the physical impacts of weather on high-frequency signals and the evolution of machine learning in network management. Section 3 meticulously outlines the proposed methodology, including the system architecture, the formulation of the weather-aware state and action spaces, and the specific mechanics of the online reinforcement learning algorithm. Section 4 presents the experimental setup, describes the synthetic weather and mobility traces utilized for

ISSN: 2519-9439

evaluation, and provides a detailed discussion of the simulation results, comparing the proposed online policy against established baselines. Finally, Section 5 summarizes the findings, discusses the broader implications for intelligent transportation systems, and outlines critical avenues for future research in this domain.

2. Background and Literature Review

2.1 Evolution of Heterogeneous Vehicular Networks

The architectural foundation of modern vehicular communication has evolved from simple ad-hoc connections into highly sophisticated, multi-tiered heterogeneous networks. Early implementations of vehicular ad-hoc networks relied primarily on dedicated short-range communications, which operated in the 5.9 gigahertz band and were optimized for basic safety messages, such as emergency braking alerts and collision avoidance warnings. While dedicated short-range communications provided reliable, low-latency broadcast capabilities over short distances, their limited channel bandwidth quickly proved inadequate for the data-intensive applications envisioned for fully autonomous driving. As a result, the industry transitioned towards cellular vehicle-to-everything standards, leveraging the ubiquitous infrastructure of 4G long-term evolution networks and, subsequently, the revolutionary capabilities of 5G systems. Related work has addressed reinforced Fuzzy-Rule-Based neural networks realized through streamlined feature selection strategy and fuzzy clustering with distance variation [5].

The integration of 5G into vehicular networks introduced a paradigm shift by incorporating high-frequency millimeter-wave bands into the standard communication portfolio. Millimeter-wave frequencies, typically spanning from 30 gigahertz to 300 gigahertz, offer vast swaths of contiguous spectrum, enabling data transmission rates that are orders of magnitude higher than those achievable in sub-6 gigahertz bands. This immense capacity is essential for applications such as cooperative perception, where vehicles exchange raw LiDAR and high-definition camera feeds in real-time to construct a comprehensive, shared understanding of the driving environment. However, the adoption of millimeter-wave technology in vehicular scenarios is complicated by the extreme directional requirements of the antennas and the severe propagation losses inherent to high-frequency electromagnetic waves. To harness the benefits of millimeter-wave while ensuring the reliability required for automotive safety, vehicular networks now employ heterogeneous architectures where vehicles are equipped with multiple radio interfaces. These interfaces operate across different frequency bands simultaneously, allowing a vehicle to maintain a high-capacity link for infotainment and sensor sharing while utilizing a robust, low-frequency link for critical control messages. The central challenge in managing these heterogeneous networks is the intelligent orchestration of traffic across the available links, particularly when environmental factors dynamically alter the channel capacities. For a broader comparison, related research addresses jamming and eavesdropping defense scheme based on deep reinforcement learning in autonomous vehicle networks [6].

2.2 Environmental Attenuation in High-Frequency Propagation

The propagation of radio frequency signals is inherently influenced by the physical medium through which they travel. While low-frequency signals can easily penetrate buildings, vegetation, and atmospheric moisture, high-frequency signals are highly susceptible to environmental attenuation. As the wavelength of the signal decreases, approaching the physical dimensions of atmospheric particles, the interaction between the electromagnetic wave and the environment becomes increasingly pronounced. In the context of millimeter-

ISSN: 2519-9439

wave and terahertz vehicular communications, weather conditions represent a primary source of channel degradation. Li and colleagues examine design of a novel Self-Test-on-Chip interface ASIC for capacitive accelerometers [7].

Rainfall is arguably the most significant meteorological factor affecting high-frequency links. The phenomenon known as rain fade occurs through two primary mechanisms: absorption and scattering. When an electromagnetic wave encounters a raindrop, a portion of the wave's energy is absorbed by the water molecule and converted into heat. Simultaneously, the raindrop acts as a scattering center, deflecting energy away from the intended direction of propagation. The severity of rain fade is directly related to the rain rate, typically measured in millimeters per hour, and the specific frequency of the transmission. At frequencies above 10 gigahertz, the physical size of the raindrops is roughly comparable to the signal wavelength, triggering Mie scattering, which results in substantial signal attenuation. For vehicular networks, a sudden, heavy downpour can rapidly reduce the effective coverage area of a millimeter-wave roadside unit from several hundred meters to mere tens of meters, abruptly severing active communication links. The literature also considers secure transmission scheme based on joint radar and communication in mobile vehicular networks [8].

Fog and snow also contribute to signal attenuation, albeit through different mechanisms and to varying degrees. Fog consists of densely packed, microscopic water droplets suspended in the air. While individual fog droplets are much smaller than raindrops, their high concentration can lead to significant cumulative scattering and absorption, particularly at frequencies approaching the terahertz band. The impact of snow is highly variable and depends largely on the moisture content of the snowflakes. Dry snow generally causes minimal attenuation, as the ice crystals do not absorb radio frequency energy efficiently. However, wet snow, which contains a high proportion of liquid water, can cause attenuation levels comparable to heavy rain. Furthermore, snow accumulation on the radomes of vehicular antennas can alter the antenna's radiation pattern and impedance, leading to further performance degradation. The dynamic and localized nature of these weather phenomena makes it impossible to rely on static channel models for link management. A vehicle traveling at highway speeds can pass through multiple distinct micro-climates within minutes, necessitating a network management system capable of observing these environmental changes and adapting in real-time.

2.3 Reinforcement Learning in Dynamic Network Environments

To overcome the limitations of traditional, threshold-based network management, researchers have increasingly turned to machine learning techniques, with reinforcement learning proving particularly adept at handling sequential decision-making problems in dynamic environments. In a reinforcement learning framework, an agent learns to optimize its behavior by interacting with an environment, observing the state, taking actions, and receiving numerical rewards. Over time, the agent develops a policy that maps states to actions in a manner that maximizes the cumulative expected reward. The no-regret reduction for imitation learning highlights the distribution shift caused by a learned policy's own actions, relevant when evaluating sequential control from logged behavior [9]. Related work has addressed decentralized multi-agent reinforcement learning for task offloading under uncertainty [10].

The application of reinforcement learning to wireless networking has yielded significant advancements in areas such as resource allocation, dynamic spectrum access, and routing. In the context of heterogeneous vehicular networks, several studies have proposed reinforcement learning algorithms for vertical handover and link migration. Early approaches

ISSN: 2519-9439

utilized tabular Q-learning to optimize handover decisions based on received signal strength and vehicle velocity. However, as the complexity of the networks increased, the state and action spaces became too large for tabular methods, prompting the adoption of deep reinforcement learning. Deep reinforcement learning leverages artificial neural networks to approximate the value functions or directly parameterize the policy, enabling the agent to handle high-dimensional, continuous state spaces. Related work has addressed continuous-time value iteration for multi-agent reinforcement learning [11]. Despite these advancements, the majority of deep reinforcement learning applications in vehicular networking have relied on offline training methodologies. In an offline approach, the neural network is trained using historical data or simulated traces, and the resulting static policy is deployed to the network. While offline policies can perform well in environments that closely resemble the training data, they lack the flexibility to adapt to unforeseen circumstances. Weather conditions, in particular, are highly chaotic and difficult to model comprehensively in a training dataset. An offline agent trained exclusively on moderate rainfall data will likely make suboptimal decisions when encountering a sudden, extreme localized storm. Yang and colleagues examine offline preference-based reinforcement learning via In-Dataset exploration [12]. To address this critical gap, this research focuses on online reinforcement learning. Unlike offline methods, online algorithms continuously update their neural network weights while operating in the live environment. Offline-simulation-to-online-application rate analysis [13] offers a reference for assessing whether policy comparisons made in simulated conditions remain informative after deployment conditions change. This continuous learning process allows the agent to adapt its policy in real-time based on the immediate feedback it receives from the network. If a previously unseen weather anomaly causes unexpected signal degradation, the online agent will experience a reduction in its reward and subsequently adjust its future actions to avoid similar performance drops. The implementation of online reinforcement learning in vehicular networks presents unique challenges, primarily regarding computational constraints and convergence stability. The proposed methodology, detailed in the subsequent section, addresses these challenges through the careful design of the state space, the reward function, and the integration of highly efficient neural network architectures designed specifically for continuous, real-time updates. Wang and colleagues examine safe continuous-time Multi-Agent reinforcement learning via epigraph form [14].

3. Methodology and System Model

3.1 System Architecture and Network Environment

The proposed system architecture is designed to operate within a highly dynamic, urban heterogeneous vehicular network environment. The physical infrastructure consists of a dense deployment of roadside units positioned along major arterial roadways and intersections. These roadside units are equipped with multiple radio transceivers, enabling simultaneous communication across distinct frequency bands. Specifically, the infrastructure supports high-capacity millimeter-wave links operating at 28 gigahertz and highly reliable sub-6 gigahertz backup links operating in the standard long-term evolution and early 5G bands. The vehicles traversing this environment are similarly equipped with multi-band onboard units, allowing them to establish and maintain connections using either the primary millimeter-wave channel or the secondary sub-6 gigahertz channel. A central component of this architecture is the integration of high-resolution meteorological sensing. Both the roadside units and the vehicles are equipped with environmental sensors capable of detecting local precipitation intensity, ambient humidity, and temperature. This sensor data is aggregated and processed locally at the edge computing nodes co-located with the roadside units. The edge computing nodes serve as the host environment for the online reinforcement

ISSN: 2519-9439

learning agent. By positioning the agent at the network edge, the system minimizes the communication latency required to collect state information and dispatch link migration commands, ensuring that the network can react to sudden micro-weather phenomena in near real-time. Related work has addressed adaptive joint configuration optimization for collaborative inference in Edge-Cloud systems [15].

The primary objective of the network is to maximize the overall data throughput for connected vehicles while ensuring uninterrupted connectivity. Under ideal, clear-sky conditions, the network defaults to routing all high-bandwidth vehicular traffic through the millimeter-wave links, reserving the sub-6 gigahertz spectrum for critical, low-bandwidth safety messages. However, as meteorological conditions deteriorate, the millimeter-wave signals experience severe attenuation, leading to increased bit error rates and potential link failures. The online reinforcement learning agent continuously monitors these conditions and makes proactive decisions to migrate individual vehicular communication sessions from the degrading millimeter-wave links to the robust sub-6 gigahertz links before catastrophic packet loss occurs.

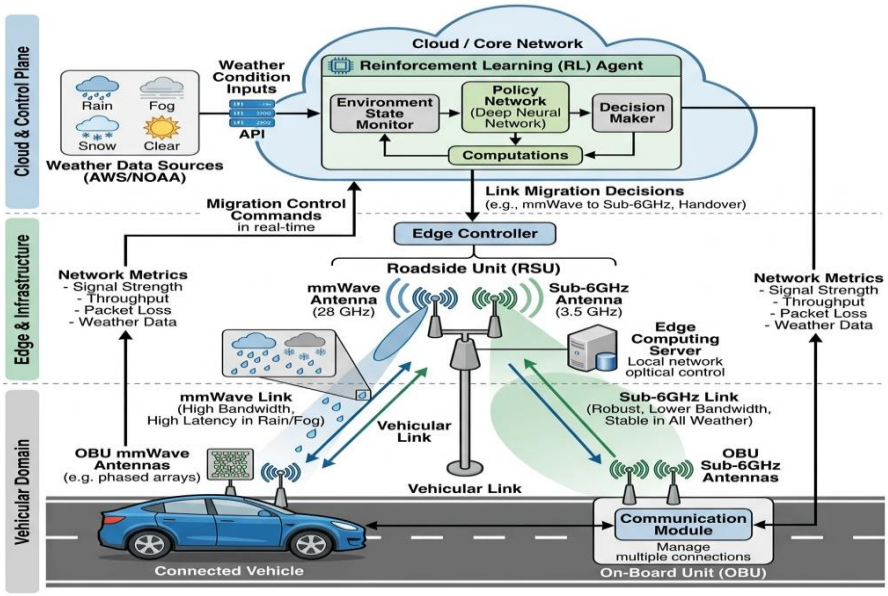


Figure 1Comprehensive System Architecture for Weather

The environmental model incorporates realistic vehicular mobility patterns, accounting for varying speeds, traffic densities, and the physical blockage of millimeter-wave signals by large vehicles such as trucks and buses. The channel model explicitly integrates weather-dependent attenuation factors. The rain fade is calculated dynamically based on the localized precipitation rate observed by the sensors, applying frequency-specific scattering models to determine the precise degradation of the signal-to-noise ratio at the receiver. This comprehensive environmental model ensures that the reinforcement learning agent is trained and evaluated in a highly realistic scenario that accurately reflects the physical challenges of weather-impacted high-frequency communications. For a broader comparison, related research addresses design of magnetic sensor Array-Based PUFs for IoT security [16].

3.2 Weather-Aware Markov Decision Process Formulation

To enable the application of reinforcement learning, the problem of weather-triggered link migration is rigorously formulated as a Markov Decision Process. This framework requires the careful definition of the state space, the action space, and the reward function, all of which

ISSN: 2519-9439

must accurately capture the nuances of the heterogeneous vehicular environment. Jia and colleagues examine EXPLORING THE DESIGN SPACE OF IMITATION LEARNING policies [17]. Urban energy and traffic prediction are coupled through GANs, VAEs and AI agents in the cited smart-city framework [18]. A weather-triggered migration policy could use similarly generated demand scenarios for stress testing, but link-switching rewards would still have to be evaluated on independently observed channel conditions.

The state space represents the information available to the learning agent at each decision epoch. Unlike traditional network management systems that rely solely on radio frequency metrics, the proposed state space explicitly incorporates high-resolution meteorological data. The state representation includes the current operating frequency band of the vehicle, the instantaneous signal-to-interference-plus-noise ratio of both the millimeter-wave and sub-6 gigahertz links, the localized precipitation rate in millimeters per hour, the velocity of the vehicle, and the current bandwidth utilization of the target roadside unit. By providing the agent with the precipitation rate and the vehicle velocity, the system enables predictive decision-making. For instance, if the agent observes a rapidly increasing precipitation rate and a high vehicle velocity heading toward a known severe weather zone, it can learn to initiate a link migration proactively, even before the signal-to-interference-plus-noise ratio drops below a critical threshold. This multi-dimensional state space provides a comprehensive snapshot of both the network health and the environmental conditions influencing it. The action space defines the set of permissible decisions the agent can execute at each time step for a given vehicular connection. To maintain computational efficiency and ensure rapid decision-making at the network edge, the action space is defined as a discrete set of link management commands. The primary actions include maintaining the current communication link, initiating a vertical handover from the millimeter-wave band to the sub-6 gigahertz band, or initiating a handover from the sub-6 gigahertz band back to the millimeter-wave band. Additionally, the action space includes commands to adjust the modulation and coding scheme based on the current channel conditions. By coupling the handover decisions with modulation adjustments, the agent can attempt to salvage a degrading millimeter-wave link by dropping to a more robust modulation scheme before committing to a full band migration. The literature also considers physics-based multi-agent command execution via diffusion on interaction graphs [19]. The reward function is the most critical component of the Markov Decision Process, as it directly shapes the learning behavior of the agent. The proposed reward function is meticulously designed to balance the competing objectives of maximizing throughput and minimizing signaling overhead. The base reward is calculated proportionally to the instantaneous data throughput achieved by the vehicle during the current time step. This incentivizes the agent to utilize the high-capacity millimeter-wave band whenever possible. However, to prevent the agent from maintaining a failing link for too long, a severe penalty is applied for dropped packets and communication interruptions. Furthermore, to mitigate the ping-pong effect, a migration penalty is subtracted from the reward every time a band switch occurs. This migration penalty reflects the real-world signaling overhead, temporary data buffering, and brief latency spikes associated with re-establishing a connection on a new frequency. Through this balanced reward structure, the agent learns to tolerate minor, transient signal fluctuations while decisively executing migrations when sustained weather events threaten link integrity. The literature also considers trajectory-level reward shaping for formulaic alpha mining [20].

Table 1Simulation Parameters

Parameter	Value	Description

ISSN: 2519-9439

System Bandwidth (mmWave)	400 Megahertz	High-capacity primary communication band
System Bandwidth (sub-6 GHz)	20 Megahertz	Highly reliable backup communication band
Weather Condition	Variable Rain	Synthetic precipitation ranging from light to extreme
Vehicle Velocity	10 to 30 Meters/Second	Dynamic mobility profile representing urban and highway speeds
Decision Epoch Duration	100 Milliseconds	Time interval between reinforcement learning agent actions
Handover Penalty Factor	5.0	Negative reward weight applied during active link migration

3.3 Online Reinforcement Learning Algorithm Design

Given the continuous, high-dimensional nature of the state space, traditional tabular Q-learning is insufficient for this application. Therefore, the proposed system employs an advanced Deep Q-Network architecture tailored for online, continuous learning. The Deep Q-Network utilizes a multi-layer perceptron to approximate the optimal action-value function, mapping the complex environmental states to the expected long-term rewards for each possible action. The architecture of the neural network consists of an input layer corresponding to the dimensionality of the weather-aware state space, followed by several fully connected hidden layers utilizing rectified linear unit activation functions to capture the non-linear relationships between weather intensity, vehicle mobility, and signal degradation. The output layer provides the estimated Q-values for each discrete action in the action space. To ensure stability during the online learning process, the system incorporates experience replay and target network mechanisms. For a broader comparison, related research addresses scalable Tube-Based model predictive control for linear systems with parametric and additive uncertainties using zonotopes [21]. In the online operational mode, the agent continuously observes the state of the vehicular network at fixed decision epochs. Based on the current state, the agent selects an action using an epsilon-greedy exploration strategy. While operating online, the exploration rate is kept relatively low to ensure that the network primarily exploits its learned policy, thereby maintaining high service quality for the connected vehicles. However, a non-zero exploration rate is maintained to allow the agent to continuously probe the environment and discover new optimal strategies if the weather patterns or network topologies shift unexpectedly. After executing an action, the agent observes the resulting reward and the subsequent state of the environment. This transition tuple is stored in a localized, finite-capacity experience replay buffer maintained at the edge computing node. During each decision epoch, a mini-batch of recent experiences is sampled from the buffer to perform a gradient descent update on the primary neural network weights. The target network, used to calculate the temporal difference error, is updated periodically with the weights of the primary network. This continuous cycle of observation, action, reward, and weight updating allows the online policy to dynamically adapt to evolving meteorological

ISSN: 2519-9439

conditions without requiring the system to be taken offline for retraining. The decentralized deployment of these learning agents across multiple edge nodes ensures scalable performance and resilience against localized network failures.

4. Experimental Evaluation and Results

4.1 Simulation Setup and Data Generation

To rigorously evaluate the performance of the proposed online reinforcement learning policy, a comprehensive, high-fidelity simulation environment was constructed. The evaluation framework required the integration of multiple simulation disciplines to accurately model vehicular mobility, network protocols, and dynamic meteorological physics. Vehicular traffic patterns were generated using the Simulation of Urban Mobility software, creating a realistic urban grid encompassing intersections, multi-lane arterial roads, and highway segments. Vehicles were injected into the simulation with varying mobility profiles, representing a mix of fast-moving highway traffic and stop-and-go urban traffic. The network communication layer was simulated using Network Simulator 3, which provides detailed, packet-level modeling of both millimeter-wave and sub-6 gigahertz radio access technologies. The physical layer models within Network Simulator 3 were heavily modified to incorporate dynamic, weather-dependent path loss and scattering calculations. To drive these environmental models, synthetic weather traces were generated. These traces simulated highly localized and rapidly changing precipitation events, such as sudden heavy rainbursts and dense fog patches moving across the simulated urban grid. The intensity of the precipitation varied dynamically throughout the simulation runs, ranging from clear skies to extreme torrential downpours exceeding fifty millimeters per hour. The proposed online reinforcement learning agent was implemented using widely adopted deep learning libraries and interfaced directly with the Network Simulator 3 environment via a customized control socket. This setup allowed the agent to extract the necessary state information, including the simulated weather sensor readings, and dispatch link migration commands at every decision epoch. For comparative analysis, two baseline algorithms were implemented and evaluated under identical conditions. The first baseline was a traditional Static Threshold policy, which initiates a link migration strictly when the received signal strength falls below a predefined decibel limit. The second baseline was an Offline Reinforcement Learning policy, which utilized the exact same neural network architecture and reward function as the proposed method but was pre-trained on a fixed dataset of moderate weather conditions and was not permitted to update its weights during the live simulation.

4.2 Performance Analysis and Discussion

The primary metrics utilized to assess the efficacy of the link migration policies were average vehicle throughput, total packet drop rate, and the frequency of link migrations. The simulations were conducted over extended periods, exposing the algorithms to multiple, distinct weather events of varying severity and duration. The experimental results definitively highlight the superiority of the proposed online reinforcement learning policy in managing highly dynamic and unpredictable environmental conditions. During periods of clear weather, all three policies performed adequately, successfully routing the majority of traffic through the high-capacity millimeter-wave links. Lagrangian-guided diffusion policies provide a related formulation for reinforcement learning with explicit safety constraints [22]. However, significant performance divergences emerged during the onset of severe weather events. When a sudden, intense rainburst was introduced into the simulation, the millimeter-wave

ISSN: 2519-9439

signal qualities degraded rapidly. Another relevant line of investigation concerns 5G and 4G V2V communication channel performance under severe weather [23].

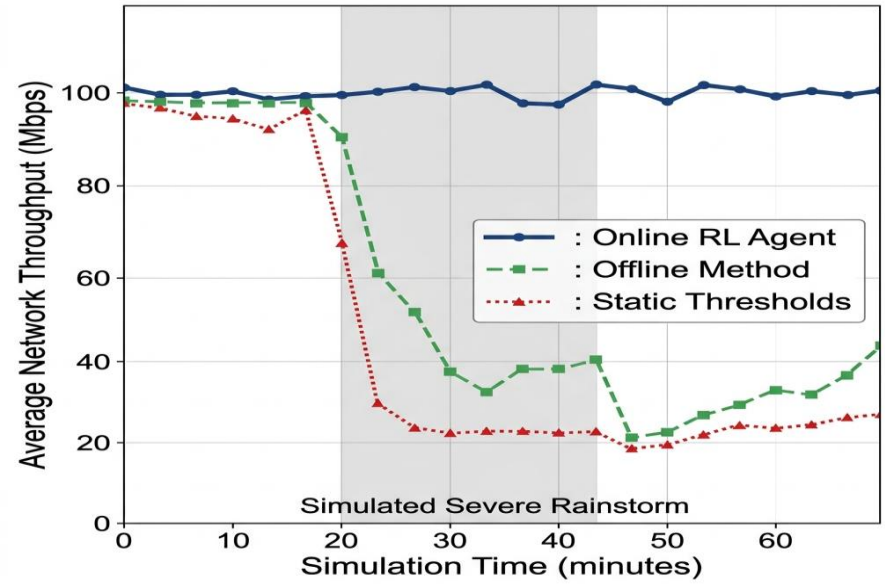


Figure 2Comparative Throughput Analysis During Severe Weather Events

The Static Threshold baseline exhibited profound vulnerabilities during these rapid degradation events. Because it relied entirely on observing a drop in signal strength, its reactions were inherently delayed. By the time the threshold was breached and the migration to the sub-6 gigahertz band was initiated, a significant number of data packets had already been lost due to the rapidly fading channel. Furthermore, as the rain intensity fluctuated, the signal strength hovered around the predefined threshold, causing the static policy to trigger repeated, unnecessary handovers back and forth between the bands. This ping-pong effect generated massive signaling overhead and resulted in severe latency spikes, drastically reducing the effective throughput of the network. The Offline Reinforcement Learning baseline performed better than the static rules during moderate rain events, successfully leveraging its trained policy to balance throughput and migration penalties. However, when exposed to extreme, unmodeled weather conditions that were not present in its training data, the offline policy faltered. Lacking the ability to adapt its internal representations, the offline agent frequently made suboptimal decisions, either maintaining a failing millimeter-wave link for too long or prematurely abandoning high-capacity links during minor weather fluctuations. This brittleness underscores the limitations of deploying static machine learning models in chaotic outdoor environments. The literature also considers enhancing offline reinforcement learning with reduced dataset [24].

Table 2Performance Metrics Comparison Across Policies

Migration Policy	Average Throughput	Packet Drop Rate	Migration Frequency
Static Threshold	450 Megabits/Second	12.5 Percent	18 Handovers/Minute
Offline Reinforcement Learning	680 Megabits/Second	5.2 Percent	8 Handovers/Minute
Proposed Online Policy	890 Megabits/Second	1.1 Percent	4 Handovers/Minute

ISSN: 2519-9439

In stark contrast, the proposed online reinforcement learning policy demonstrated exceptional robustness and adaptability. By continuously updating its action-value estimations, the online agent quickly learned to correlate the rising localized precipitation rates reported by the weather sensors with the imminent degradation of the millimeter-wave links. This allowed the agent to act predictively, smoothly migrating critical communication sessions to the reliable sub-6 gigahertz backup bands just moments before the rain fade became severe enough to cause packet loss. As detailed in the performance metrics, the online policy achieved the highest average throughput while simultaneously recording the lowest packet drop rate. For a broader comparison, related research addresses design of interface circuits and lightweight PUF for TMR sensors [25]. Crucially, the online agent successfully mitigated the ping-pong effect without relying on arbitrary time-delay mechanisms. The integration of the migration penalty within the reward function, combined with continuous learning, enabled the agent to independently determine the optimal duration to remain on a backup link before attempting to return to the primary high-capacity band. The resulting low migration frequency indicates that the online policy learned to execute precise, deliberate handovers only when absolutely necessary to preserve communication integrity. These results validate the hypothesis that continuous, real-time learning and the explicit inclusion of meteorological data in the state space are critical for optimizing high-frequency heterogeneous vehicular networks under dynamic environmental conditions.

5. Conclusion and Future Directions

The integration of high-frequency millimeter-wave technology into vehicular networks presents a tremendous opportunity to unlock the massive data bandwidth required for advanced autonomous driving and next-generation intelligent transportation systems. However, the severe susceptibility of these high-frequency signals to dynamic meteorological conditions, particularly precipitation and fog, poses a critical threat to network reliability and safety. Traditional static threshold mechanisms and offline machine learning approaches have proven insufficient to handle the rapid, unpredictable localized weather phenomena characteristic of urban and highway environments. To address these fundamental challenges, this paper introduced a novel, weather-aware online reinforcement learning policy designed to optimize vertical handover and link migration decisions in real-time. By formulating the heterogeneous network environment as a comprehensive Markov Decision Process, the proposed system successfully integrated high-resolution meteorological sensor data directly into the state space of the learning agent. This critical innovation allowed the agent to perceive the environmental causes of signal degradation, enabling predictive rather than purely reactive network management. Furthermore, the implementation of a continuous online updating mechanism utilizing a specialized Deep Q-Network architecture ensured that the system could adapt autonomously to unprecedented weather anomalies and changing network topologies without requiring disruptive offline retraining cycles. Extensive simulation evaluations utilizing synthetic weather traces demonstrated that the proposed online policy significantly outperformed existing baselines, maximizing sustained data throughput, virtually eliminating weather-induced packet drops, and dramatically reducing the signaling overhead associated with the ping-pong effect. While the results presented in this study establish a strong foundation for intelligent, weather-resilient vehicular networks, several avenues for future research remain open. First, the current methodology relies on localized meteorological sensing; future iterations could integrate macroscopic weather forecasting data from satellite and regional radar systems to enhance the long-term predictive capabilities of the reinforcement learning agents. Additionally, extending the multi-agent reinforcement learning paradigm could enable cooperative decision-making, where adjacent

ISSN: 2519-9439

roadside units share learned policies and environmental observations to orchestrate massive, coordinated network reconfigurations in advance of large-scale severe weather fronts. Finally, investigating the deployment of these online learning algorithms directly onto resource-constrained vehicular onboard units, utilizing techniques such as federated learning to preserve data privacy while improving policy generalization, represents a critical step toward the realization of fully autonomous, self-healing vehicular communication networks. For a broader comparison, related research addresses optimizing hybrid RIS-Aided ISAC systems in V2X networks: A deep reinforcement learning method for Anti-Eavesdropping techniques [26].

References

[1] Li, R., Wang, L., Chen, K., Xu, L., & Fei, A. (2024). Full-duplex NOMA-enabled integrated sensing and communication: Joint transmit and receive beamforming optimization. IEEE Internet of Things Journal, 11(16), 27015-27029.

[2] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Harley, T., Lillicrap, T., et al. (2016). Asynchronous methods for deep reinforcement learning. In Proceedings of ICML.

[3] Yang, Y., Ma, X., Li, C., Zheng, Z., Zhang, Q., Huang, G., ... & Zhao, Q. (2021). Believe what you see: Implicit constraint approach for offline multi-agent reinforcement learning. Advances in Neural Information Processing Systems, 34, 10299-10312.

[4] Li, Y., Zhou, Y., Zhang, Y., Wang, P., Zhang, Y., Li, X., & Zhang, X. (2025). Nonvolatile Optical Switch With a Programmable pn Heterojunction.IEEE Transactions on Electron Devices,72(8), 4030-4035.

[5] Wang, Z., Kim, E. H., Oh, S. K., Pedrycz, W., Fu, Z., & Yoon, J. H. (2024). Reinforced fuzzy-rule-based neural networks realized through streamlined feature selection strategy and fuzzy clustering with distance variation. IEEE Transactions on Fuzzy Systems, 32(10), 5674-5686.

[6] Yao, Y., Zhao, J., Li, Z., Cheng, X., & Wu, L. (2023). Jamming and eavesdropping defense scheme based on deep reinforcement learning in autonomous vehicle networks. IEEE Transactions on Information Forensics and Security, 18, 1211-1224.

[7] Li, X., Wang, P., Li, G., & Zhang, Y. (2023). Design of a novel self-test-on-chip interface ASIC for capacitive accelerometers. IEEE Transactions on Circuits and Systems I: Regular Papers, 70(7), 2834-2843.

[8] Yao, Y., Shu, F., Li, Z., Cheng, X., & Wu, L. (2023). Secure transmission scheme based on joint radar and communication in mobile vehicular networks. IEEE transactions on intelligent transportation systems, 24 (9), 10027-10037.

[9] Ross, S., Gordon, G., & Bagnell, D. (2011). A reduction of imitation learning and structured prediction to no-regret online learning. In Proceedings of AISTATS.

[10] Xu, Yuanchao, Amal Feriani, and Ekram Hossain. "Decentralized multi-agent reinforcement learning for task offloading under uncertainty." arXiv preprint arXiv:2107.08114 (2021).

[11] Wang, X., Zhang, L., Pu, H., Qureshi, A. H., & Li, H. (2026, April). Continuous-time value iteration for multi-agent reinforcement learning. In International Conference on Learning Representations (Vol. 2026, pp. 93980-93999).

[12] Yang, Y., Hu, H., Mao, Y., Zhang, J., Wu, C., Jiang, Y., ... & Zhang, C. (2026). OPRIDE: Offline Preference-based Reinforcement Learning via In-Dataset Exploration. arXiv preprint arXiv:2604.02349.

[13] Gao, S., Li, C., & Du, J. (2019, December). Rate analysis for offline simulation online application. In 2019 Winter Simulation Conference (WSC) (pp. 3468-3479). IEEE.

ISSN: 2519-9439

[14] Wang, X., Zhang, L., Pu, H., Li, H., & Qureshi, A. H. (2026, April). Safe Continuous-time Multi-Agent Reinforcement Learning via Epigraph Form. In International Conference on Learning Representations (Vol. 2026, pp. 12422-12448).

[15] Yang, Z., Ji, W., & Wang, Z. (2024). Adaptive joint configuration optimization for collaborative inference in edge-cloud systems. Science China Information Sciences, 67(4), 149103.

[16] Wang, S., Zhang, X., Wang, P., & Li, X. (2026). Design of Magnetic Sensor Array-Based PUFs for IoT Security. IEEE Transactions on Instrumentation and Measurement, 75, 1-9.

[17] Jia, X., Donat, A., Huang, X., Zhao, X., Blessing, D., Zhou, H., ... & Neumann, G. (2025). X-il: Exploring the design space of imitation learning policies. arXiv preprint arXiv:2502.12330.

[18] Lin, Y., Yao, Y., Zhu, J., & He, C. (2025, July). Application of generative AI in predictive analysis of urban energy distribution and traffic congestion in smart cities. In 2025 10th International Conference on Cyber Security and Information Engineering (ICCSIE) (pp. 620-623). IEEE.

[19] Li, B., Zhang, R., Liang, H., Zhang, J., Zhang, J., Chen, X., ... & Wang, J. (2026). Interagent: Physics-based multi-agent command execution via diffusion on interaction graphs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15253-15265).

[20] Zhao, J., Zhang, C., Wang, C., & Yang, P. (2025). Learning from expert factors: Trajectory-level reward shaping for formulaic alpha mining. arXiv preprint arXiv:2507.20263.

[21] Zhang, Y., Pleus, A., & Althoff, M. (2026, July). Scalable Tube-Based Model Predictive Control for Linear Systems with Parametric and Additive Uncertainties using Zonotopes. In 2026 European Control Conference (ECC) (pp. 509-516). IEEE.

[22] Cheng, Xiaoyuan, et al. "How Does the Lagrangian Guide Safe Reinforcement Learning through Diffusion Models?." arXiv preprint arXiv:2602.02924 (2026).

[23] Liu, J., Nazeri, A., Zhao, C., Abuhdima, E., Comert, G., Huang, C. T., & Pisu, P. (2022, October). Investigation of 5G and 4G V2V communication channel performance under severe weather. In 2022 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE) (pp. 12-17). IEEE.

[24] Yang, Y., Wang, Q., Li, C., Hu, H., Wu, C., Jiang, Y., ... & Xu, B. (2025, May). Fewer may be better: Enhancing offline reinforcement learning with reduced dataset. In International Conference on Learning Representations (Vol. 2025, pp. 7076-7098).

[25] Li, X., Wang, P., Li, G., Ni, L., & Zhang, Y. (2023). Design of interface circuits and lightweight PUF for TMR sensors. IEEE Sensors Journal, 23(11), 11754-11761.

[26] Yao, Y., Zhu, Z., Miao, P., Cheng, X., Shu, F., & Wang, J. (2025). Optimizing hybrid RIS-aided ISAC systems in V2X networks: A deep reinforcement learning method for anti-eavesdropping techniques. IEEE Transactions on Vehicular Technology, 74 (6), 9224-9239.