Soft Computing (2024) 28:5275–5285
https://doi.org/10.1007/s00500-023-09281-8

APPLICATION OF SOFT COMPUTING

Dynamic link utilization empowered by reinforcement learning for adaptive storage allocation in MANET

R. P. Prem Anand[1] · V. Senthilkumar[2] · Gokul Kumar[3] · A. Rajendran[4] · A. Rajaram[5]

Accepted: 14 September 2023 / Published online: 20 October 2023

Abstract
In modern wireless networks, mobile nodes often deal with the challenge of maintaining a sufficient number of data packets due to limited storage capacity within each cluster. It adversely impacts network performance by compromising data quality during transmissions. The ensuing delays, caused by data packets awaiting storage allocation, result in reduced throughput and increased end-to-end latency. To effectively address these issues, we present a Dynamic Link Utilization with Reinforcement Learning (DLU-RL) method, which is designed to optimize storage allocation for communication data packets, significantly enhancing network performance. Instead of static allocation, DLU-RL employs dynamic strategies guided by reinforcement learning algorithms. This innovative method not only tackles storage constraints but also proactively adapts to varying network conditions and traffic patterns. In our approach, we first perform a comprehensive analysis of storage capacities across all nodes, establishing a baseline for dynamic resource allocation. The DLU-RL framework then swiftly assigns storage space based on real-time demand and priority, optimizing storage utilization on the fly. As a result of implementing DLU-RL, substantial enhancements in throughput and concurrent minimization of end-to-end delays are achieved. This research not only contributes to efficient storage allocation techniques but also pioneers the integration of reinforcement learning for wireless communication network performance optimization. The proposed framework signifies a paradigm shift in storage management, offering adaptability, efficiency, and real-time optimization to tackle the evolving challenges of wireless communication.

Keywords Position middle storage space allocation · Dynamic link utilization with reinforcement learning · Network quality · Data storage optimization · Data transmission

✉ R. P. Prem Anand
premanandphd@gmail.com

V. Senthilkumar
Thil4ever@gmail.com

Gokul Kumar
gokulkumar_ece@yahoo.co.in

A. Rajendran
rajendran.a@kce.ac.in

A. Rajaram
drrajaram@egspec.org

1 Department of Electronics and Communication Engineering, SCSVMV University, Kanchipuram 631 561, India

2 Er. Perumal Manimekalai College of Engineering, Koneripalli, Hosur 635117, India

3 Ross Techsys, Bangalore 562129, India

4 Department of ECE, Karpagam College of Engineering, Otthakalmandapam, Coimbatore, India

5 Department of Electronics and Communication Engineering, E.G.S Pillay Engineering College, Nagapattinam 611002, India

1 Introduction

In today's rapidly evolving landscape, the focus on Mobile Ad hoc Networks (MANETs) has intensified, especially with the widespread adoption of 5G in mobile devices. The spotlight is now on mobile nodes, emphasizing their significance and the need for heightened attention. This is a common approach as mobile networks seek scalability and adaptability, adapting seamlessly as mobile devices join, move, or exit the network. This exploration revolves around creating adaptable and decentralized setups for mobile networks. A Mobile Ad hoc Network (MANET) is a network composed of wireless mobile nodes that autonomously unite to establish a network without the need for pre-existing infrastructure. In this network, each node serves as both a sender and receiver, facilitating the exchange of data packets among all nodes within the same communication cluster. The visual representation of this MANET structure can be seen in Fig. 1.

In the realm of communication, MANETs are on the rise, driven by the need for faster data transmission, improved quality of service, and reduced latency. To address these demands, various parameters come into play, enhancing data transmission across the network. Centered on mobile nodes, MANETs provide a pathway for adaptable and scalable communication. Unlike traditional networks relying on fixed structures, MANETs are built upon wireless mobile nodes that establish connections on the go, without the constraints of a predefined network setup. To attain low latency and high throughput, Deep Reinforcement Learning (DRL) is employed in the context of data center networks within cloud manufacturing. For conventional load balancing struggles with dynamic traffic, the Dynamic Load Balancing (DLB) approach, driven by DRL, dynamically monitors and adapts parallel paths, optimizing link utilization and fine-tunes path weights, aiding switches in selecting optimal paths for load balancing, which effectively reduces average flow completion time in asymmetric network topology (Wang et al. 2023). The reinforcement learning-based optimization algorithm enhances satellite communication systems by leveraging Inter satellite links, which minimize reliance on ground networks, decrease complexities, and focus on improving network performance. Unlike traditional approaches, this algorithm treats link assignment as a sequence decision problem, considering factors like inter-satellite visibility and network connectivity. Through a series of selection actions, the agent chooses links to optimize link delay (Ren et al. 2022). The deep reinforcement learning method is utilized to handle the complex state spaces to improve Internet-of-vehicle networks using D2D cellular systems with distributed caching for video streaming. It optimizes decisions for selecting cache-enabled vehicles, power allocation, and cellular vehicles' power allocation. By concurrently optimizing these choices, the objective is to improve video quality, ensure seamless playback, and guarantee data rate performance (Choi et al. 2021). The RL implies an application in the field of Internet congestion control. Congestion control involves efficiently regulating data transmission rates to optimize network capacity, which is particularly crucial for services like live video and Internet-of-Things. By employing RL, the deep network policies excel in capturing complex data traffic and network conditions (Jay et al. 2019).

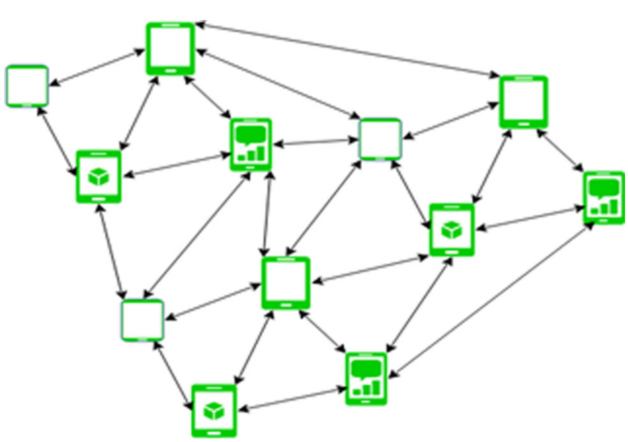

Fig. 1 MANET (Mobile Ad Hoc Network) structure

The existing research negatively impacts data quality during transmissions, leading to delays, reduced throughput, and increased latency. To overcome these issues, the study introduces the Dynamic Link Utilization with Reinforcement Learning (DLU-RL) method. This approach aims to optimize storage allocation for communication data packets dynamically using reinforcement learning algorithms. By doing so, it seeks to enhance network performance by efficiently allocating storage resources based on real-time demand and priority, thus improving throughput and reducing end-to-end delays. Additionally, the study pioneers the integration of reinforcement learning into wireless communication network management, offering adaptability and real-time optimization to address the evolving challenges in wireless communication.

The paper is structured as follows: Sect. 2 covers related works, Sect. 3 details the proposed Dynamic Link Utilization with Reinforcement Learning (DLU-RL) model, Sect. 4 discusses the results obtained through various metrics, and finally, Sect. 5 concludes by summarizing the findings and presenting future perspectives.

2 Literature review

Li et al. (2020) introduced a new approach using reinforcement learning (RL) to create a smart controller. This controller can adapt to changing conditions in a system and make decisions in real time. It's all about making sure the system meets its speed goals while also making the most of its servers. They tested it on a real system, like a search engine, and found that it reduced delays while making servers work up to 70% better than if they were only used for interactive tasks. This approach is especially useful when the system has to handle a mix of jobs, like in machine learning. This study deals with important ideas like data centers, self-organizing computing, cloud systems, and using RL to make decisions in a sequence.

Yang et al (2023) focuses on dynamic link prediction in complex evolving networks. Existing models face challenges like vulnerability to attacks, low accuracy, and instability. To address this, a new approach called Self-adaptive Stable Gate-incorporated Graph Convolutional Network (SAGE-GCN) is proposed. SAGE-GCN includes a state encoding network and a policy network. It captures local topology using a multi-power adjacency matrix, distinguishing features across snapshots. A stable gate ensures spatiotemporal dependencies, proven to be stable under network changes. A self-adaptive strategy and policy network then learn optimal features for dynamic link prediction, showing improved accuracy and robustness against attacks in experiments on real-world graphs.

Feng et al. (2019) focuses on the significance of link quality in WSN to ensure stable communication. To address this, we introduce the SUBXBFCM algorithm, which categorizes link quality based on how well packets are received. We also examine the connections between hardware characteristics and reception rates using the Pearson coefficient. Additionally, we employ XGB_LQE, which considers factors like received signal strength, link quality, and signal-to-noise ratio to assess the current link quality grade. This approach leverages XGBoost's classification capabilities to provide a more accurate evaluation. Building on this, XGB_LQP employs XGBoost regression for predicting future link quality based on historical data. Experimental results validate SUBXBFCM's effectiveness in grade division and highlight XGB_LQP's superior predictive performance compared to Support Vector Regression and 4C methods in single-hop wireless sensor networks (WSN).

Guan et al. (2022) address spectrum scarcity in vehicle-mounted communication networks for the Internet of Vehicles (IoV). It optimizes a metric balancing V2V and V2I link capacity while reducing interference for high-priority links. Using deep reinforcement learning, a spectrum allocation strategy is proposed. Simulation results show improved V2V and V2I link rates compared to random allocation. The approach reduces interference for priority links by 14.2 dB, prioritizing services and remaining robust against communication noise.

Rajaram and Baskar (2023) introduces a Hybrid Optimization method for MANETs. This method divides the network into segments using Dual Constraint Clustering (DCC). It then chooses Cluster Heads (CHs) employing a fuzzy approach. For routing, it employs a combination of Hybrid Cellular Automata and African Buffalo Optimization (HCA2BO). Simulation confirms improved network metrics and demonstrates its efficacy in air pollution monitoring.

Prem Anand and Rajaram (2020) address the challenge of unstable mobile nodes in wireless networks leading to inaccurate data transmission, packet loss, and energy inefficiency.

The author introduced the Enhanced Data Accuracy-based Path Discovery (EAPD) technique intending to ensure high data accuracy during the data transmission. EAPD verifies the data accuracy of nodes in the routing path, selectively choosing nodes with the highest accuracy while excluding those with lower accuracy. Furthermore, a selection method is employed for selecting a route to prevent intrusion during mobile communication, resulting in reduced energy consumption and fewer packet drop rates. The author proposed the Enhanced Data Accuracy-based Path Discovery (EAPD) technique, aiming to ensure high data accuracy in transmission. EAPD verifies the data accuracy of nodes in the routing path and selects nodes with maximum accuracy, rejecting those with lower accuracy. Additionally, a route selection algorithm is introduced to avoid intrusion during communication, reducing energy consumption and packet drop rates. This research offers a promising approach to enhance data accuracy and network efficiency in the presence of unstable mobile nodes.

Rahamatbasha et al. (2022) introduce the Reliability Antecedent Packet Forwarding (RAF) technique to ensure reliable routing from source to destination mobile ad hoc networks. RAF avoids flooding nodes, backs up routing information, and retrieves it in case of interference. A straddling path recovery algorithm monitors node packet flow rates, providing interference-free routes with more nodes for communication. These nodes possess higher resources and serve as data backups, enhancing network longevity and reducing packet loss.

Abbasi et al. (2023a, b) addresses the function of home care services during the COVID-19 pandemic by introducing an integrated location-allocation-routing model for home healthcare supply chain corporations. The model combines metric and constraint-based methods to optimize

Authors	Methodology	Metrics	Limitations
Li et al. (2020)	Reinforcement learning	Latency	Evaluation is limited to a physical prototype, and may not generalize to all environments
Yang et al. (2023)	Self-adaptive stable gate-incorporated graph convolutional network	Link prediction accuracy	Evaluation is limited to a physical prototype, and may not generalize to all environments
Feng et al. (2019)	SUBXBFCM algorithm	Packet reception rate	Evaluation is limited to a physical prototype, and may not generalize to all environments
Guan et al. (2022)	Deep reinforcement learning	Capacity, inference reduction	Results based on simulations, real-world deployment challenges not addressed
Rajaram and Baskar (2023)	Hybrid cellular automata and African Buffalo optimization	Network metrics improvement	Limited to MANETs, and may not apply to other network types
Prem Anand and Rajaram (2020)	Enhanced data Accuracy based Path Discovery	Energy consumption, packet drop rate	Effectiveness in highly dynamic environments not fully explored
Rahamatbasha et al. (2022)	Reliability Antecedent Packet Forwarding	Packet loss rate	Specific to mobile ad hoc networks, may not generalize to other network types

pharmacy locations, patient allocation, and nurse routing, ultimately recommending the LP-metric approach for efficient decision-making and enhanced employee performance in the home healthcare sector.

2.1 Case study

The proposed DLU-RL method offers invaluable solutions for real-life engineering scenarios. In a smart manufacturing setup, these techniques could optimize data storage in IoT devices, enhancing production efficiency by reducing data transmission delays and improving overall system throughput. Similarly, in a smart city deployment, the algorithms could streamline data exchange in traffic management systems, minimizing congestion and bolstering real-time decision-making capabilities. Implementing these methods might require some adaptation to existing infrastructure but offers a clear pathway to tangible benefits through heightened operational efficiency and improved data flow in complex engineering environments.

3 Proposed methodology

3.1 Problem identification

In the landscape of modern wireless networks, particularly (Abbasi and Choukolaei 2023) in Mobile Ad-hoc Networks (MANETs), a significant challenge emerges from the constrained storage capacity within individual clusters of nodes. This limitation underscores a critical issue affecting network performance. As mobile nodes traverse the network, their restricted storage capabilities can lead to a multitude of adverse convergences. Among these challenges, data quality during transmission is compromised due to the inability to efficiently store and transmit data. Furthermore, the waiting time for storage allocation creates a delay in data transmission, subsequently causing a reduction in overall network throughput and an increase in end-to-end latency. The repercussions extend to data loss and packet drop when storage constraints prevent temporary data storage, thereby eroding the reliability of communication. Such storage-related bottlenecks have cascading effects on the network's ability to support real-time applications and data-intensive tasks, underlining the critical need for innovative solutions. This context set the stage for the proposed Dynamic Link Utilization with Reinforcement Learning (DLU-RL) method, which addresses these challenges through dynamic resource allocation guided by reinforcement learning algorithms, aimed at optimizing storage allocation, enhancing resource utilization, and ultimately, elevating network performance within the confines of limited storage capacity. In essence, the convergence of limited storage capacity and wireless networks necessitates adaptive solutions. The DLU-RL method presents a holistic approach to proactively tackle storage limitations and enhance network performance dynamically.

3.2 Reinforcement learning framework setup

Reinforcement Learning serves as a cornerstone of the Dynamic Link Utilization with Reinforcement Learning (DLU-RL) method, enabling the dynamic optimization of storage allocation within the network. This framework equips the system with the ability to learn and adapt storage allocation strategies based on the observed network conditions.

3.2.1 RL framework components

State (S): It encapsulates the present network configuration, encompassing critical parameters such as storage capacities, link strengths, ongoing data traffic, and other relevant metrics. This state representation acts as an input for the agent's decision-making process. It is represented as follows:

$$S = [S_1, S_2 \ldots S_n] \tag{1}$$

where S_k encompasses storage capacity, link quality, and other data for node k.

Action (A): In this paradigm, the action denotes the choice of storage allocation for each node as $A = [A_1, A_2, \ldots, A_n]$. In the context of our system, this translates to determining the amount of storage space assigned to a specific node or cluster and also signifies storage allocation decisions for an individual node.

Reward (R): It constitutes a numerical evaluation of the chosen action's impact on network performance. Crafted a function of defined metrics, the reward function gauges the desirability of storage allocation decisions and shapes the learning process. It is expressed as below:

$$R = \mathrm{Func}(S, A) \tag{2}$$

Policy (π): It embodies the strategic approach adopted by the RL agent to map observed states to optimal actions. It forms the core of the RL agent's decision-making strategy for storage allocation. The policy π maps state to actions, expressed as $\pi: S \rightarrow A$.

3.3 Data collection and preprocessing

In the initial phase of our proposed framework, comprehensive data collection and preprocessing set the stage for informed decision-making. Real-time and historical data about critical parameters such as storage capacities, network conditions, and traffic patterns are gathered. This data encompasses the dynamic nature of the wireless environment, capturing fluctuations in resource availability and utilization. Storage capacities across network nodes are recorded to establish a baseline for efficient resource allocation. Concurrently, network conditions and traffic patterns are scrutinized, illuminating the intricate interplay between data transmission and network dynamics.

Subsequently, the collected data undergoes a crucial preprocessing phase. This step is vital to render the data suitable for input into the reinforcement learning model. Data cleansing techniques are employed to eliminate outliers, inaccuracies, and anomalies that could skew the model's learning process. Normalization or scaling is applied to ensure that disparate data types and ranges are harmonized, facilitating coherent interpretation by the model. Furthermore, categorical data might be encoded for compatibility with the reinforcement learning framework. Ultimately, the data preprocessing phase culminates in a refined dataset, primed to fuel the reinforcement learning algorithm's understanding and learning process. This meticulous data curation lays the foundation for the subsequent stages of the proposed framework, empowering the reinforcement learning model to make informed decisions that optimize storage allocation dynamically and enhance overall network performance.

3.4 Link quality prediction

The overall process involves the compression and utilization of local observation data from each Mobile-to-Mobile (M2M) link within a Mobile Ad-hoc Network (MANET). Every M2M link carries out observations to collect vital data, such as its transmission power (TP), the cumulative interference power (IP) originating from other links, and the interference channel gain extending from the M2M link to all neighboring mobile nodes. Each M2M link performs observations to gather essential information, including its transmitting power (TP), combined interference power (IP) from other links, and the interference channel gain from the M2M link to all neighboring mobile nodes. This information is crucial to assess the impact of M2M links on each other within the MANET. To reduce signaling overhead, the interference (I) channel gain information is estimated at each mobile node and shared with its neighboring nodes. Additionally, the transmitter of each M2M link can acquire the power gain of the current channel through timely feedback. This observed data ($\mathrm{OD_k}$) is as follows,

$$\mathrm{OD}_k = \{L_k, I_{k,b}, IP_k, TP_k\} \tag{3}$$

The observed data consists of its transmitting power, interference power, interference channel gain to neighboring nodes, and power gain of the current channel. This observed data, represented as a vector, is then compressed by using Deep Neural Networks (DNN) on each M2M link. The DNN outputs a compressed data vector V_k, which includes elements representing the compressed observations from the M2M link. The compressed data vector D_k referred to as the feedback vector, is shared with neighboring mobile nodes. This feedback vector, denoted as F_k, serves a dual role. It acts as input to a Deep Q-Network (DQN) to facilitate the reinforcement learning process and simultaneously guides the DQN's decision-making process. The DQN takes into consideration the compressed observations from all M2M links and employs reinforcement learning techniques to optimize resource allocation decisions within the MANET as shown in Fig. 2. This approach reduces the need for excessive signaling while enabling intelligent decision-making in a dynamic wireless environment where mobile nodes collaborate for efficient network performance. This predictive insight enhances the adaptability and responsiveness of the network to changing conditions, ultimately leading to an optimized utilization of resources and improved network performance. The anticipatory nature of this approach aligns seamlessly with the reinforcement learning framework, further enhancing the overall efficacy of our methodology.

3.5 Reinforcement learning training

The Reinforcement Learning Training process plays a pivotal role in achieving long-term optimization within our proposed approach. Reinforcement learning serves as the driving force for dynamic decision-making in complex scenarios. Our innovative Deep Q Network (DQN) design, inspired by end-to-end learning principles, empowers agents to extract effective policies directly from intricate outputs. This tailored approach equips our system to handle the intricacies of spectrum allocation tasks in wireless networks. By seamlessly integrating reinforcement learning with our specialized DQN, our agents learn and adapt, leading to optimized spectrum utilization and enhanced network performance in dynamic environments.

The training process is initiated by utilizing both collected data and predicted link qualities obtained from the Deep Neural Network (DNN) model. The primary goal is to facilitate the RL model's learning of optimal storage allocation strategies based on historical and real-time data. To achieve efficient storage allocation decisions, we employ the Q-learning algorithm. The RL agent starts with an understanding of the state and action spaces, encompassing factors such as the storage capacities of nodes and the range of available actions for storage allocation. The reward function quantifies the desirability of actions in terms of network performance. Throughout the learning process, the RL agent navigates between exploration and exploitation. During exploration, the agent ventures into unexplored storage allocation strategies, enriching its experiences and refining its grasp of the network dynamics. Conversely, during exploitation, the agent leverages its acquired knowledge to optimally allocate storage resources based on past observations.

$$A_t = \{A_{\text{explore}} \text{withprobability} \epsilon A_{\text{exploit}} \text{withprobability} 1 - \epsilon \quad (4)$$

The RL agent remains in continuous interaction with the network environment, gathering feedback in the form of rewards and subsequently updating its policy. This learning process is encapsulated through the Q-learning equation, where Q-values are updated to balance immediate rewards

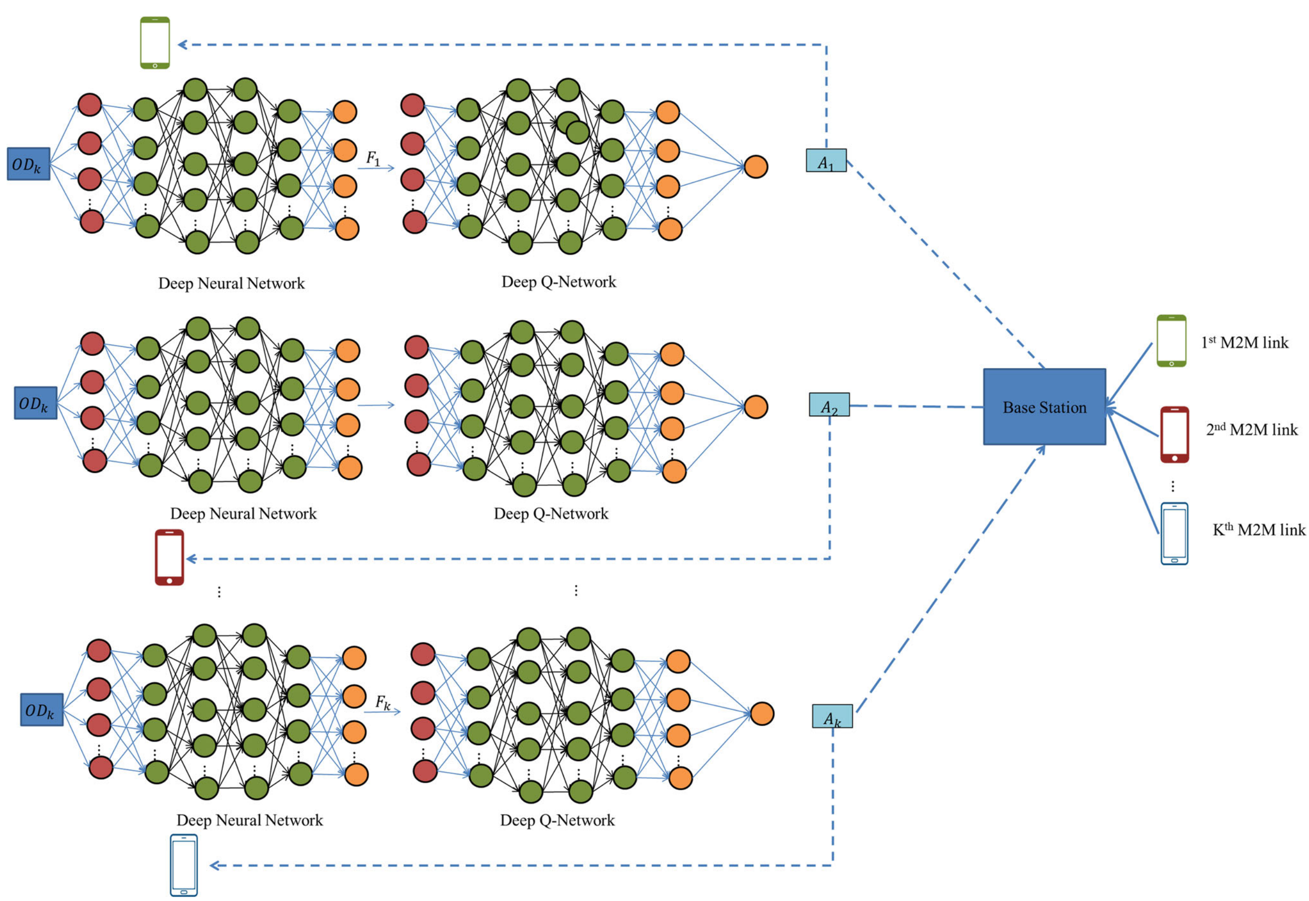

Fig. 2 Schematic structure for Link Quality Prediction process

and potential future gains. Through multiple iterations of network interactions, the RL agent's policy evolves and converges towards actions that yield maximum cumulative rewards. This iterative convergence process involves consistent updates based on observed experiences and expressed as below,

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \beta(R_{(t)} + \alpha \cdot \max_{A'} Q(S_{t+1}, A') - Q(S_t, A_t)) \quad (5)$$

Over multiple iterations of interactions with the network, the RL agent's policy refines and converges towards optimal actions that yield maximum cumulative rewards. This convergence process involves continually updating the values based on observed experiences.

$$Q(S_t, A_t) \leftarrow Q^*(S_t, A_t) \quad (6)$$

Through the infusion of the RL network into the proposed MANET system, the network gains the capability to dynamically optimize storage allocation decisions. This adaptability enhances storage utilization, leading to improved network performance and responsiveness in the face of evolving network conditions and demands. The exploration–exploitation trade-off, Q-value updates, and eventual convergence form the core of this learning process, driving the intelligent allocation of storage resources. Figure 3 illustrates the structure of the RL framework.

3.6 Process flow

In the context of our proposed methodology, the Algorithm serves as the cornerstone of the decision-making process within the Mobile Ad-hoc Network (MANET). In each discrete time step t, this belongs to the set $\{1, 2, \ldots T\}$, we establish the observed value of the kth Mobile-to-Mobile (M2M) link denoted as OD_k^t. These individual link observations collectively shape the overall observed values for all M2M links at that specific time step. This comprehensive representation encapsulates the state of the entire network's links at that instant. we derive the approximate target value of return, Y^t, by incorporating the reward at the subsequent time step $R(t + 1)$, along with a discount factor β and the maximization of the action-value function Q that depends on observations at the next time step. This pivotal equation is as follows,

$$Y_t = R_{(t+1)} + \beta \max_A Q(OD_{t+1}, A; \phi) \quad (7)$$

The iterative update process of the Deep Q Network (DQN) situated at the Base Station (BS) is orchestrated by the equation as follows,

$$\phi \leftarrow \phi + \gamma \sum_{t \epsilon E} \frac{\partial Q(OD_t, A_t; \phi)}{\partial \phi} [Y_t - Q(OD_t, A_t; \phi)] \quad (8)$$

where ϕ signifies the parameter set of the neural network employed for Q-value approximation. The coefficient γ accounts for the temporal step length during the strategy gradient iteration, while E designates the collection of episodes. Within each episode t, every individual M2M link engages in local data observation, denoted as OD_k^t. which subsequently serves as input to a Deep Neural Network (DNN). This DNN processes the input data to generate a specific feedback vector. Following this, the feedback is transmitted to the BS. At the BS, the amalgamated feedback vectors from all M2M links collectively feed into the DQN. This DQN then generates decisive actions based on the amalgamated feedback and disseminates these actions across all M2M links. Following this, each M2M link independently selects its spectrum allocation based on the received decisions. This intricate process, facilitated through the harmonious integration of DNN and DQN, empowers the MANET with the capability to dynamically allocate spectrum resources intelligently. This intelligent adaptation leads to enhanced utilization of resources, ultimately bolstering the overall performance of the network in response to the ever-evolving conditions and requirements.

4 Result and discussion

Our proposed Dynamic Link Utilization with Reinforcement Learning (DLU-RL) method was assessed utilizing the NS-2.34 simulator. The simulation was conducted in a network comprising 250 nodes, spread over an area of 1060 m by 980 m. The MAC layer protocol employed was 802.11 g, and the radio range was set at 250 m, enabling wireless communication within this range. The simulation time spanned 60 ms, during which various scenarios were evaluated. For traffic generation, the Constant Bit Rate (CBR) traffic type was utilized, and each packet was sized at 150 bytes. The mobility pattern was based on the Random Way Point model, ensuring diverse movement of nodes throughout the network. The communication protocol chosen for the simulation was the Ad hoc On-Demand Distance Vector (AODV) routing protocol, facilitating dynamic route establishment and maintenance. This simulation setup aimed to emulate a real-world scenario and provide insights into the performance of our DLU-RL method. By simulating within the NS-2.34 environment, we effectively assessed our proposed methodology to enhance dynamic link utilization and resource allocation within the given network conditions (Table 1).

The proposed DLU-RL method compared with the existing system such as TCSSR (Prem Anand and Rajaram 2022), ETPA (Prem Anand and Rajaram 2023a), and MDS

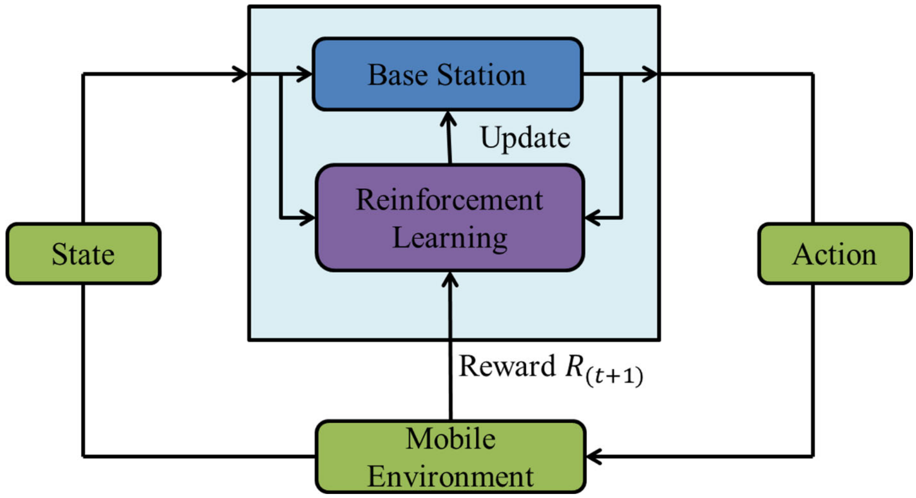

Fig. 3 Proposed DLU-RL block diagram

(Prem Anand and Rajaram 2023b) insight of performance metrics comparison analysis.

4.1 End-to-end delay

This metric quantifies the time required for a data packet to traverse from one node to another, encompassing all network-related delays. In our proposed system, we aim to minimize end-to-end delay to ensure efficient data delivery. It is calculated as follows:

$$\text{End-to-end-Delay} = \frac{\sum(\text{Transmission delay} + \text{Propagation delay} + \text{Queueing Delay})}{\text{No. of Data packets}}$$

Figure 4 presents a comparative analysis of the end-to-end delay among different scenarios: Existing TCSSR CBS, ETPA, MDS, and Proposed DLU-RL. The number of nodes in the network varies across rows, ranging from 50 to 250 nodes. As observed from the figure, the end-to-end delay decreases consistently as we transition from Existing TCSSR, ETPA, and MDS to the Proposed DLU-RL method. For instance, with 50 nodes, the end-to-end delay varies: 14 for TCSSR, 12 for ETPA, 10 for MDS, and notably drops to 8 for Proposed DLU-RL. This pattern persists across node counts, highlighting the effectiveness of Proposed DLU-RL in reducing end-to-end delay. Even at 250 nodes, the end-to-end delay is 12 for TCSSR, 9 for ETPA, 6 for MDS, and significantly decreases to 3 for Proposed DLU-RL. These findings emphasize the Proposed DLU-RL's substantial impact on minimizing end-to-end delay across different network scales compared to existing methods.

Table 1 Simulation setup

Simulator	NS 2.34
Count of nodes	250
Network area	1060 × 980
MAC layer(mac)	802.11 g
Transmission distance	250 m
Duration of simulation	60 ms
Data flow type	Constant bit rate
Byte size of packets	150
Node mobility	Random way point
Data transfer protocol	Ad hoc on-demand distance vector

4.2 Communication overhead

It metric reflects the additional data and signaling introduced due to protocol and control information, beyond the actual payload. In our proposed system, we strive to minimize communication overhead to enhance spectral efficiency and conserve resources. The communication overhead can be calculated using the following equation:

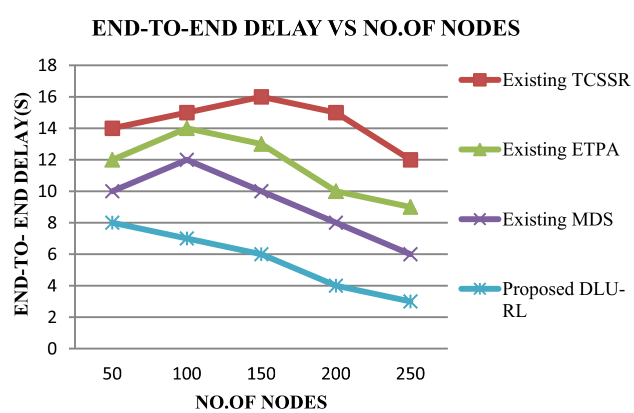

Fig. 4 Analysis of the proposed method corresponds to the End-to-End Delay

$$\text{Communication Overhead} = \frac{\text{Overhead Data packets}}{\text{Total Data packets Transmitted}} * 100\%$$

Figure 5 establishes the performance analysis of various methods over pause time ranges from 20–100. For instance, at a pause time of 20, the Proposed DLU-RL exhibits a communication overhead of 18, while Existing TCSSR, ETPA, and MDS show values of 30, 28, and 25, respectively. As the pause time extends to 100, the advantages of the Proposed DLU-RL become even more increased insight of 11, whereas Existing TCSSR, ETPA, and MDS have communication overhead values of 19, 16, and 14, respectively. It consistently demonstrates the lowest communication overhead values compared to the existing TCSSR, ETPA, and MDS methods across various pause times. These results underline the efficiency and effectiveness of the Proposed DLU-RL scheme in minimizing communication overhead and optimizing resource utilization across various operational scenarios. The data showcases its superiority over existing methodologies and its ability to adapt to different pause times, ultimately leading to improved network performance.

Packet delivery ratio: This metric calculates the proportion of packets successfully delivered out of the total number of packets sent. In our proposed system, achieving a high packet delivery ratio is crucial to ensure reliable data transmission.

$$\text{Packet Delivery Ratio} = \frac{\text{No. of data packets delivered}}{\text{Total number of sent packets}} * 100\%$$

Figure 6 depicts the packet delivery ratio of the existing and proposed method over increasing the count of nodes in the network. In the case of Existing TCSSR, For instance, with 50 nodes, the packet delivery ratio stands at 69, and for 250 nodes, the ratio increases to 81. Similarly, for ETPA With 50 nodes, the ratio is 81, and as the nodes increase to 250, the ratio decreases to 85. MDS follows a similar pattern, displaying diminishing packet delivery ratios as the node count increases. The packet delivery ratio starts at 82 with 50 nodes and drops to 87 with 250 nodes. However, the Proposed DLU-RL scheme stands out with a distinct pattern. The packet delivery ratio remains consistently high, showcasing its effectiveness in maintaining efficient packet delivery. With 50 nodes, the ratio is 86, and even with 250 nodes, the ratio only increases to 92.5. This highlights the robustness and superiority of the Proposed DLU-RL scheme in sustaining a high packet delivery ratio across varying node counts. It outperforms the existing TCSSR, ETPA, and MDS methods, making it a promising approach for enhancing network performance and ensuring reliable packet delivery in dynamic environments.

Network lifetime: It reflects the duration for which the network can operate efficiently before depleting its energy resources. It can be calculated as follows,

$$\text{Network lifetime} = \frac{\text{Total Initial Energy}}{\text{Total energy consumption rate}}$$

Figure 7 shows the network lifetime of existing and proposed schemes while increasing the count of nodes. The TCSSR experiences a network lifetime of marginal increase with a node count rise. In the case of ETPA, the network lifetime exhibits consistent growth with the expansion of nodes. MDS also showcases a pattern of improved network lifetime as the number of nodes rises. Remarkably, the Proposed DLU-RL scheme illustrates a notably higher network lifetime, regardless of the number of nodes. This scheme consistently outperforms the existing methods. This result shows the effectiveness of the Proposed DLU-RL scheme in significantly prolonging network lifetime. It surpasses the performance of the existing TCSSR, ETPA, and MDS methods, making it a highly promising approach for enhancing network sustainability and overall operational lifespan in dynamic environments.

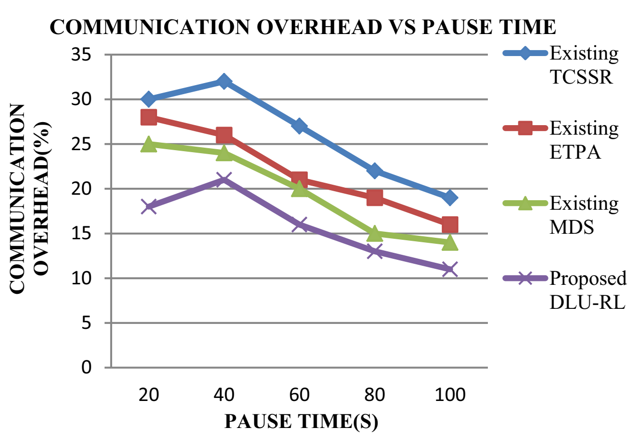

Fig. 5 Communication Overhead Comparison

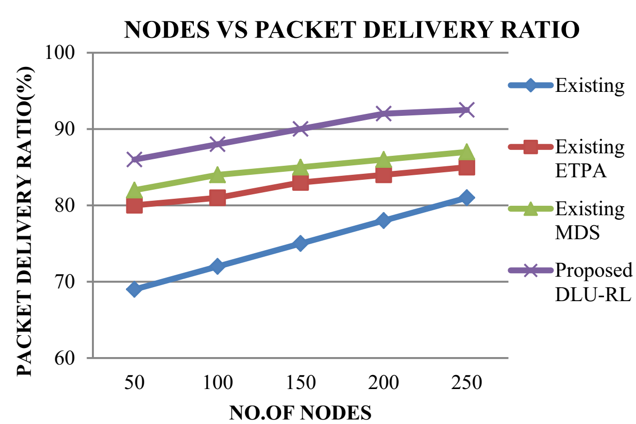

Fig. 6 Comparison between existing and proposed system based on the Packet Delivery Ratio measure

Energy consumption: It quantifies the amount of energy consumed for various network operations. Figure 8 demonstrates the comparison between various methods in light of energy consumption metrics. From the observation of the figure, the proposed DLU-RL attains a minimum energy consumption value of 102 J at 50 nodes, which is 2.04%, 3.721%, and 6.68% less than MDS, ETPA and TCSSR, respectively. The Proposed DLU-RL scheme demonstrates a substantial reduction in energy consumption than the other existing methods. This underscores the effectiveness of the Proposed DLU-RL scheme in minimizing energy consumption. It consistently surpasses the performance of the existing TCSSR, ETPA, and MDS methods, making it a highly promising approach for achieving efficient energy utilization and prolonged network operation in dynamic scenarios.

Packet drop rate: It measures the percentage of data packets that are dropped or lost during the data transmission. It can be computed as follows,

$$\text{Packet Delivery Ratio} = \frac{\text{Number of data packets dropped}}{\text{Total number of sent data packets}} * 100\%$$

Figure 9 presents a comprehensive analysis of packet drop rates about the varying density of network nodes. As we observe the changes in node density, intriguing trends emerge in the context of packet drop rates. For the Existing TCSSR method, the packet drop rates demonstrate a gradual upward trajectory as the number of network nodes increases. Similarly, the ETPA approach follows a similar trend, showing an incremental rise in packet drop rates with expanding node density. The MDS method exhibits a modest increase in packet drop rates within the network. In contrast, our Proposed DLU-RL scheme stands out by consistently maintaining lower packet drop rates across various node densities. This ability to withstand packet drops stands out, especially when contrasted with the established TCSSR, ETPA, and MDS approaches. The

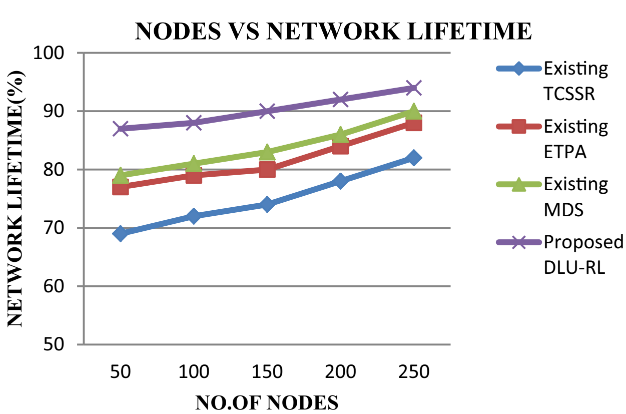

Fig. 7 Network lifetime analysis

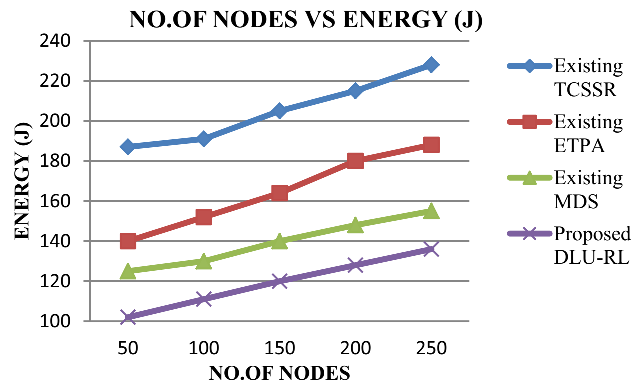

Fig. 8 Energy consumption comparison between existing and proposed system

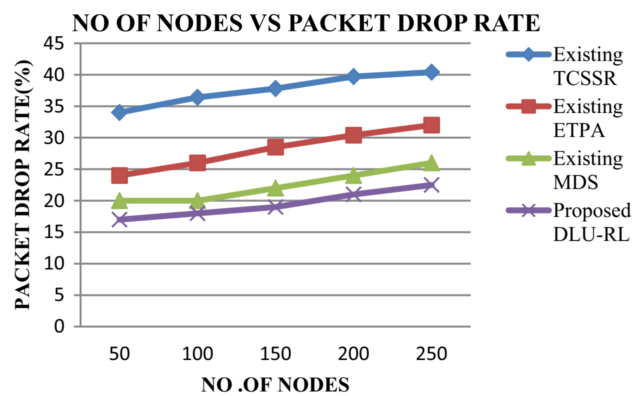

Fig. 9 Packet drop rate analysis

Proposed DLU-RL approach demonstrates the potential to effectively mitigate packet drop rates, thereby strengthening the reliability of data transmission and positioning itself as a promising solution to enhance the overall robustness of data delivery within dynamic network environments.

Overall, the proposed DLU-RL method outperforms existing approaches (TCSSR, ETPA, MDS) in various aspects. It achieves minimized end-to-end delay and reduced communication overhead, enhancing overall network efficiency. Notably, the DLU-RL method significantly improves the packet delivery ratio, extends network lifetime, and lowers energy consumption through the innovative Position Middle storage space allocation algorithm.

5 Conclusion

We introduced the Dynamic Link Utilization with Reinforcement Learning (DLU-RL) method as a robust solution to the inherent challenges of constrained storage capacity within Mobile Ad-hoc Networks (MANETs). The constrained storage capacity of individual nodes in MANETs often leads to compromised data quality during transmission, delays in data allocation, increased end-to-end latency, data loss, and packet drops, severely impacting network performance. Our findings consistently highlight the superiority of DLU-RL across critical performance metrics. It excelled in minimizing end-to-end delays, ensuring efficient and timely data delivery even in scenarios involving 250 nodes, where the delay was as low as 3 units. DLU-RL also proved its capacity to enhance spectral efficiency and resource conservation by exhibiting a communication overhead of just 11% with a pause time of 100 units. Furthermore, it demonstrated exceptional reliability with a packet delivery ratio of 92.5% in 250-node scenarios. DLU-RL showcased its potential to extend network lifetimes significantly, facilitating prolonged and efficient network operation while maintaining energy-efficient behavior, with energy consumption as low as 102 J for 50 nodes. Remarkably, it consistently outperformed existing methods in mitigating packet drops, enhancing data transmission reliability, and overall network robustness. Collectively, these results position DLU-RL as a promising solution for MANETs, offering dynamic storage allocation, minimized delays, reduced communication overhead, efficient energy utilization, and resilient data transmission. Beyond addressing current network limitations, this flexible framework exhibits adaptability to dynamic network environments, holding substantial promise for real-world applications.

Acknowledgements There is no acknowledgement involved in this work.

Author contributions All authors are contributed equally to this work.

Funding No funding is involved in this work.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest Conflict of interest is not applicable in this work.

Ethics approval and consent to participate No participation of humans takes place in this implementation process.

Human and animal rights No violation of Human and Animal Rights is involved.

References

Abbasi S, Choukolaei HA (2023) A systematic review of green supply chain network design literature focusing on carbon policy. Dec Anal J. https://doi.org/10.1016/j.dajour.2023.100189

Abbasi S, Daneshmand-Mehr M, Ghane Kanafi A (2022) Designing sustainable recovery network of end-of-life product during the COVID-19 pandemic: a real and applied case study. Discr Dyn Nat Soc 2022:1–21

Abbasi S, Sıcakyüz Ç, Erdebilli B (2023a) Designing the home healthcare supply chain during a health crisis. J Eng Res. https://doi.org/10.1016/j.jer.2023.100098

Abbasi S, Daneshmand-Mehr M, Ghane Kanafi A (2023b) Green closed-loop supply chain network design during the coronavirus (COVID-19) pandemic: a case study in the Iranian Automotive Industry. Environ Model Assess 28(1):69–103

Botta A, Canonico R, Navarro A, Stanco G, Ventre G (2023) Scalable reinforcement learning for dynamic overlay selection in SD-WANs. In: 2023 IFIP networking conference (IFIP networking). IEEE, pp 1–9

Choi M, Shin M, Kim J (2021) Dynamic video delivery using deep reinforcement learning for device-to-device underlaid cache-enabled Internet-of-vehicle networks. J Commun Netw 23(2):117–128

Feng Y, Liu L, Shu J (2019) A link quality prediction method for wireless sensor networks based on XGBoost. IEEE Access 7:155229–155241

Guan Z, Wang Y, He M (2022) Deep reinforcement learning-based spectrum allocation algorithm in the internet of vehicles discriminating services. Appl Sci 12(3):1764

Jay N, Rotman N, Godfrey B, Schapira M, Tamar A (2019) A deep reinforcement learning perspective on internet congestion control. In: International conference on machine learning. PMLR, pp 3050–3059

Li Y, Sun D, Lee BC (2020) Dynamic colocation policies with reinforcement learning. ACM Trans Architect Code Optim (TACO) 17(1):1–25

Prem Anand RP, Rajaram A (2022) Effective timer count scheduling with spectator routing using stifle restriction algorithm in manet. IOP Conf Ser Mater Sci Eng 994(1):012031

Prem Anand RP, Rajaram A (2023a) Efficient path identification in 5g-based cellular networks for environmental impact analysis. J Environ Prot Ecol 24(5):1760–1772

Prem Anand RP, Rajaram A (2023b) Decision making in maximum data storage space allocation in wireless communication. J Environ Prot Ecol 24(5):1746–1759

Prem Anand RP, Rajaram A (2020) Enhanced data accuracy based path discovery using backing route selection algorithm in MANET. Peer-to-Peer Netw Appl 13(6):2089

Rahamatbasha AYS, Sharma C, Sayeed F et al (2022) Implementation of reliability antecedent forwarding technique using straddling path recovery in MANET. Wirel Commun Mob Comput. https://doi.org/10.1155/2022/6489185

Rajaram A, Baskar A (2023) Hybrid optimization-based multi-path routing for dynamic cluster-based MANET. Cybern Syst

Ren W, Zhu J, Qi H, Cong L, Di X (2022) Dynamic optimization of inter-satellite link assignment based on reinforcement learning. Int J Distrib Sens Netw 18(2):15501477211070202

Wang J, Luo W, He Y, Rao S, Hu J (2023) A novel deep reinforcement learning based dynamic load balancing mechanism in datacenter networks. Available at SSRN 4327073

Yang L, Liu H, Sun D, Liu K, Chen CL (2023) Robust temporal link prediction in dynamic complex networks via stable gated networks with reinforcement learning. https://doi.org/10.36227/techrxiv.22658380.v3