논문 2024-19-14

Optimizing Energy Efficiency in Mobile Ad Hoc Networks: An Intelligent Multi-Objective Routing Approach

Sun Beibei*

Abstract: Mobile ad hoc networks represent self-configuring networks of mobile devices that communicate without relying on a fixed infrastructure. However, traditional routing protocols in such networks encounter challenges in selecting efficient and reliable routes due to dynamic nature of these networks caused by unpredictable mobility of nodes. This often results in a failure to meet the low-delay and low-energy consumption requirements crucial for such networks. In order to overcome such challenges, our paper introduces a novel multi-objective and adaptive routing scheme based on the Q-learning reinforcement learning algorithm. The proposed routing scheme dynamically adjusts itself based on measured network states, such as traffic congestion and mobility. The proposed approach utilizes Q-learning to select routes in a decentralized manner, considering factors like energy consumption, load balancing, and the selection of stable links. We present a formulation of the multi-objective optimization problem and discuss adaptive adjustments of the Q-learning parameters to handle the dynamic nature of the network. To speed up the learning process, our scheme incorporates informative shaped rewards, providing additional guidance to the learning agents for better solutions. Implemented on the widely-used AODV routing protocol, our proposed approaches demonstrate better performance in terms of energy efficiency and improved message delivery delay, even in highly dynamic network environments, when compared to the traditional AODV. These findings show the potential of leveraging reinforcement learning for efficient routing in ad hoc networks, making the way for future advancements in the field of mobile ad hoc networking.

Keywords : Q-routing, ad-hoc networks, Multi-objective routing, Adaptive routing

Ⅰ. INTRODUCTION

Mobile Ad hoc Networks (MANETs) have garnered significant attention for their potential applications in dynamic and infrastructure-less wireless communication environments [1]. In such networks, mobile devices communicate without relying on centralized infrastructure, making them well-suited for scenarios like disaster response, military operations, and vehicular networks. However, routing in MANETs presents numerous challenges due to the dynamic nature of network topology, resource limitations, and the necessity to optimize multiple objectives [2].

One of the most efficient way to deal with such challenges is based on the routing algorithm used to forward data packets. Traditional routing protocols in MANETs, such as Ad hoc On-demand Distance Vector (AODV) [3] and Dynamic Source Routing (DSR) [4], have been widely employed but encounter limitations in efficient route selection and simultaneous achievement of multiple objectives [5]. These protocols typically prioritize metrics like hop count or finding the shortest path, often overlooking other critical considerations such as minimizing latency, maximizing network throughput, and achieving load balancing [6]. The shortcomings of these traditional approaches emphasize the need for more intelligent routing schemes in MANETs that can address the diverse and evolving challenges posed by the network's dynamic nature.

In MANETs, the high mobility and unpredictable nature of wireless links lead to the formation of an unstable topology, rendering it highly unreliable for efficient data transfer. To address these challenges, this paper proposes a novel multi-objective and adaptive routing scheme based on the popular reinforcement learning algorithm, Q-learning [7] for MANETs. Our Q-learning algorithm strategically selects optimal next-hops to construct the most efficient route, ensuring efficient data transfer regardless of the dynamic and

*Corresponding Author (cherrybaby0718@gmail.com)
Received: Nov. 30, 2023, Revised: Dec. 26, 2023, Accepted: Feb.23, 2024.
Sun Beibei: Semyung University (Ph.D. Student)

ⓒ IEMEK J. Embed. Sys. Appl. 2024 Apr. 19(2) 107-114
ISSN : 1975-5066
http://dx.doi.org/10.14372/IEMEK.2024.19.2.107

unpredictable changes of the network. The introduced routing scheme, named A-AODV is implemented on the widely-used AODV routing protocol to assess its efficiency and effectiveness. The proposed scheme takes into consideration feedbacks from the external enironment by considering factors such as the energy consumption of nodes, the degree of congestion among neighboring nodes, and the degree of mobility amongst moving nodes. Based on the received signal feedbacks, the algorithm enhances its decision-making process in response to the dynamic conditions of the network..

Unlike previous studies, in the propsoed A-AODV routing scheme, valid neighbor information are maintained for a limited and a dynamically adjusted period of time called, the Estimated Link Duration (ELD) between two adjacent nodes. The ELD timer is updated each time a node receives an update periodic 'Hello' message from its neighbor. This strategy enable nodes to disregard expiring or soon-to-expire wireless links, promoting a more accurate representation of the current network state. Additionally, it assigns lower preference to nodes with links that are likely to expire soon, even if they might exhibit better performance in other aspects, such as energy efficiency or congested buffer.

Traditional Q-learning-based routing protocols often face limitations due to their incapacity to adapt to the dynamic nature of MANETs, caused by the use of fixed Q-learning parameters such as the discount factor and the learning rate. In study [8] the authors suggests that as the network topology changes more rapidly, a larger learning rate should be used to give higher priority to new information, while a smaller discount factor should be used to reflect unstable future expectations. In our study, we establish the relationship between the discount factor and the ELD parameter, as well as between the learning rate and the mobility factor (MF) of a node. Additionally, we incorporate the use of shaped rewards based on the physical distance between forwarding normal nodes and the destination nodes. This integration aims to discourage the selection of longer routes and speed up the learning process, leading to more efficient routing decisions.

To estimate the ELD of a wireless link between two nodes, we use the extended Hello messages to allow mobile nodes to exchange important mobility information such as GPS location and their moving speed. A node is allowed to adjust its interval of sending Hello messages based on its neighborhood measured mobility. Hello message interval is ranging between 1 to 2 seconds only. With the use of the adaptive Hello message interval, we expect to see a significant reduction in the number of Hello messages flooding the network. Using the network simulator tool ns-3.36 [9] we verify the usefulness of our proposed scheme by observing its performance on various parameters.

The rest of this paper is organized as follows. In section II we go into a summarized introduction to our work while reviewing some of the previous studies related to this paper. Section II discusses the system model and details on different algorithms and scheme applied when implementing the A-AODV routing scheme. Section III presents simulation results and discusses about simulation results. We conclude our study and discuss in brief the possible future direction of our research in Section IV.

II. SYSTEM MODEL AND METHODS

This section presents a brief introduction to the proposed routing scheme and went into more description on how the shaped rewards were incorporated to enhance its learning speed.

1. Q-learning Model for Routing in MANET

Reinforcement learning is a machine learning technique that enables a learning agent to learn optimal actions by interacting with an environment. RL algorithms consist of an agent, an environment, and a reward signal. The agent takes actions in the environment, receives feedback in the form of rewards or penalties, and learns to maximize the cumulative reward over time [10].

As it can be seen from Fig. 1 above, in Q-learning, the agent's learning process comprises of a sequence of stages, referred to as epochs (0, 1, ..., n, ...). At time t ,the agent in state $\boldsymbol{S_t}$, performs an action $\boldsymbol{a_t}$ and receives a reward $\boldsymbol{R_t}$ as it transitions to the new state $\boldsymbol{S_{t+1}}$. The action value is updated as follows:

$$Q_t(S_t, A_t) \leftarrow Q_{t-1}(S_t, A_t) \times (1-\alpha) + \alpha \times [R_t + \gamma \times \max_A Q_t(S_{t+1}, A_t)]. \quad (1)$$

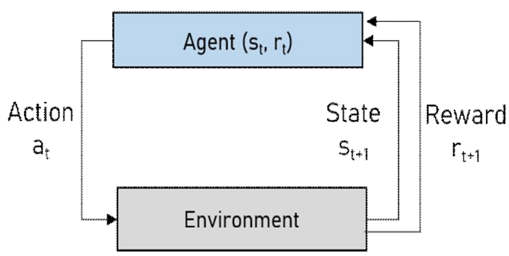


Fig. 1. Summary of a Q-learning process

The reward R_t is a scalar feedback signal indicating the agent's performance at step t. The agent's objective is to maximize the cumulative rewards obtained from the environment, based on its previous experiences. A fundamental aspect of the Q-learning algorithm as a routing protocol is the utilization of the so-called Q-table as the routing table. To guide the exploration behavior of the learning agent and integrate domain-specific heuristic knowledge, we introduce an additional reward F_{t+1}., to the reward received from the environment R_t. This is referred to as shaping reward function F, and when considered in the Q-learning update rule it is done as follows:

$$Q_t(S_t, A_t) \leftarrow Q_{t-1}(S_t, A_t) \times (1-\alpha) + \alpha \times [R_t + F_{t+1} + \gamma \times \max_A Q_t(S_{t+1}, A_t)]. \tag{2}$$

To ensure that these modifications do not alter the optimal policy, F is formulated as the difference of a potential value between the a particular state and the final state,

$$F_{t+1} = \gamma\phi(t+1) - \phi(t). \tag{3}$$

across the state space. Here, ϕ is a potential function providing insights into the states. In our study, we define the potential function ϕ for reward shaping as the Normalized Euclidean Distance between the current state s and the goal state G, expressed in the equation below with x being the furthest position to the goal state from any point within the simulation area.

$$\varphi = 1 - \frac{distance(s, G)}{distance(x, G)}. \tag{4}$$

In the context of routing in MANETs, RL algorithms have gained attention as promising approaches to overcome the limitations of traditional routing schemes. RL-based routing schemes enable agents to learn from network dynamics and make intelligent decisions based on the network's current state [11]. By taking into account network environmental factors like node mobility, link quality, and energy consumption of nodes, RL algorithms can adaptively select routes that optimize multiple objectives, such as energy efficiency and QoS-related parameters like delivery delay and network throughput [12].

2. The proposed RL based multi-objective routing model

In the proposed routing scheme, we consider a network topology consisting of a number of mobile nodes and a static destination node, of which all mobile nodes are aware of its position. Data packets are forwarded in multi-hop manner from a source node (src) to the destination node (dst). The entire network is considered as an environment of state S; thus, a data packet at node i is represented as state s_i. The process of forwarding a data packet indexed as $P\ (i,dst)$ to the next node j is seen as an action $a_{i,j}$ as the packet is moving to a new state s_j, node i receives a reward $r_{i,j}$.

2.1 Multi-objective optimization in MANET

Routing in Mobile Ad hoc Networks (MANETs) involves optimizing multiple objectives simultaneously. Traditional routing schemes often focus on a single objective, such as minimizing routing costs in terms of parameters like hop-count or delivery delay. However, in MANETs, it is crucial to consider multiple objectives, including network throughput, network lifetime, and load balancing. Multi-objective optimization (MOO) techniques offer a framework to manage trade-offs between these conflicting objectives. MOO algorithms aim to identify a set of solutions known as the Pareto front, where improving one objective comes at the expense of another. These solutions represent various trade-offs, allowing decision-makers to select the most appropriate solution based on their specific preferences.

Applying MOO to solve routing problems in MANETs enables the selection of routes while striking a balance among various objectives. By considering multiple metrics, such as energy consumption, load balancing, and link lifetime, the proposed routing scheme is capable of autonomously adapting to the dynamic nature of MANETs. It guarantees to optimize performance across multiple dimensions, even without full knowledge of the entire network topology.

2.2 Modified Q-learning based AODV routing protocol

To effectively track the unpredictable changes in the network, the proposed routing scheme consists of several autonomous mechanisms, including neighbor discovery, tracking the validity time of exchanged information, and Q-routing with adjustable parameters.

Furthermore, as illustrated in Fig. 2, we have extended the AODV protocol's Hello message header to carry additional information such as node internal state (Queue congestion), look-ahead state (Maximum Q-value among its neighbor), geographical and mobility information obtained through GPS. Such information is necessary when estimating the shaped reward and checking the validity of exchanged information.

<table>
<tr><td>Type</td><td>R</td><td>A</td><td>Reserved</td><td>Prefix Size</td><td>Hop Count</td></tr>
<tr><td colspan="6">Destination IP Address</td></tr>
<tr><td colspan="6">Destination Sequence Number</td></tr>
<tr><td colspan="6">Geographical location (coordinates of x and y)</td></tr>
<tr><td colspan="3">Queue congestion</td><td colspan="3">Maximum Q-value</td></tr>
<tr><td colspan="3">Learning rate</td><td colspan="3">Discount factor</td></tr>
<tr><td colspan="6">Node Velocity(double)</td></tr>
</table>

Fig. 2. Extended Hello message header

1) Adaptive neighbor discovery

The traditional AODV routing protocol, through its neighbor discovery process, uses broadcast periodic Hello messages to establish communication links between nodes. This process is determined by two important timers: the Hello interval and neighbor timeout timer. The Hello timer accounts for the interval at which a node periodically broadcasts Hello messages at a fixed interval (default is 1 second). Meanwhile, the neighbor timeout timer accounts for the time a neighbor node is considered reachable and within communication range. These timers are necessary to ensure link status monitoring, neighbor discovery and maintenance, route validation, and network topology updates. In traditional AODV, the neighbor timeout timer is fixed and set to correspond to three consecutive missed 'Hello' messages (3 seconds).

However, the use of fixed settings for such important timers in AODV and MANETs, in general, presents several challenges, particularly due to the dynamic nature of these networks. These challenges include inflexibility in tracking rapid changes, unnecessary network overhead in stable conditions, and increased energy consumption. This study proposes an adaptive timer setting approach for both timers. We begin by estimating the valid Estimated Link Duration (ELD) between two nodes and assign this value as the neighbor timeout timer for a neighbor node. A node's 'Hello' interval is determined by the minimum value of its computed ELD values, termed as $eld_{\min}$. The ELD timer is bound within a specific range from $eld_{\min}$ (0.5 seconds) to $eld_{\max}$ (3 seconds). We assign the value of the estimated link lifetime to be equal to the estimated ELD as computed in study [13].

Each node based on its Hello interval periodically sends HELLO packets that include its geographic location, energy, mobility model (i.e., moving speed and direction of the neighbor), queue congestion, maximum Q-value, learning rate and Q-routing parameters. When a node receives HELLO packets, it extract such important to establish and maintain its neighbor table. Each node utilizes the information of its neighbor table to perceive the network condition. If the information of a neighbor is not refreshed within $timeout\ timer$ seconds it will be removed from the neighbor table.

2) Q-learning adaptive parameters

In Q-learning, the learning rate determines to what extent the newly acquired information overrides the old information. If the learning rate is higher, the Q-value is updated faster. Most of the existing Q-learning-based routing protocols have a fixed learning rate, as is the case with the discount factor, which determines the importance of future rewards compared to immediate rewards during routing decision-making. However, in MANETs, if the mobility of nodes is relatively high, then the speed of updating the Q-values should be faster. Also, in such cases, future rewards should be considered less. Therefore, we introduce a method of adjusting the learning rate and the discount factor to adapt to the speed of updating Q-value by assigning a corresponding learning rate and a discount factor for each link.

We utilize the Mobility Factor $\boldsymbol{MF}$ and the normalized $\boldsymbol{ELD}$ parameter of a node to adaptively adjust its learning rate and discount factor, respectively. The $\boldsymbol{MF}$, calculated for each node in relation to its neighbors, is determined based on the rate of connectivity changes observed over a period of time t as follows:

$$MF = \frac{Lost_{links} + New_{links}}{Maintained_{links}}. \tag{5}$$

Hence, the new learning rate (α') is computed as:

$$\alpha' = \alpha \times MF. \tag{6}$$

whereby $1 < \alpha < 0$. Also, the discount factor is further discounted as the measured mobility of a neighbor node get higher. Therefore the new discount factor γ' is estimated as:

$$\gamma' = \gamma \times \frac{eld_i}{eld_{\max}}. \tag{7}$$

whereby, eld_i is the estimated link lifetime of the next hop and $eld_{\max}$ is the maximum possible link lifetime between two nodes.

3) Adaptive Q-learning in A-AODV

To implement our proposed scheme, we proposed an algorithm (**algorithm 1**) which computes and update the Q-value for a given source-destination node pair based on observed network state.

Algorithm 1: Q-value Update and Action Selection

Input: All (*src*, *dst*) pairs, α, γ, ε, Network graph, Node residual energy, queue length, and distances to dst node (s), number of episodes *n,* eld and MF.

for each (*src, dst*) ∈within topology **do**
  Initialize Q : A * S→R initialized with 0
  for episode ←1 to n **do**
    Start in state S_n = src or a node searching a path
      while $S_{n+1} \neq dst$ **do**
        Select A_n for S_n
          $R_{n+1} \leftarrow R(S_n, A_n)$ // *get the reward and observe the new state S_{n+1} using expression (5) to update Q-value.*
        $S_n \leftarrow S_{n+1}$ // *Move to the next state;*
      end
    end
    Using the newly update Q-table find the best path with state-action pairs that has the maximum Q-value..
  end
Store the set of all computed paths

4) The proposed Multi-objective reward function

The multi-objective Q-routing scheme simultaneously optimizes multiple objectives, such as energy consumption, load-balancing and selection of stable links. We define the reward function as observed by node i when selecting node j as its next hop as a directly proportional variable to the objective functions as shown below.

$$R_{ij} = \omega_1 \cdot \frac{E_j}{E_{node}} + \omega_2 \cdot \frac{Q_i}{Q_{\max}}. \quad (8)$$

Whereby, E_j and E_{node} are the remained energy and initial energy units of node j respectively measured in unit Joule. Q_j and $Q_{\max}$ represent the available queue size and the maximum queue size of node j respectively.

Ⅲ. RESULTS

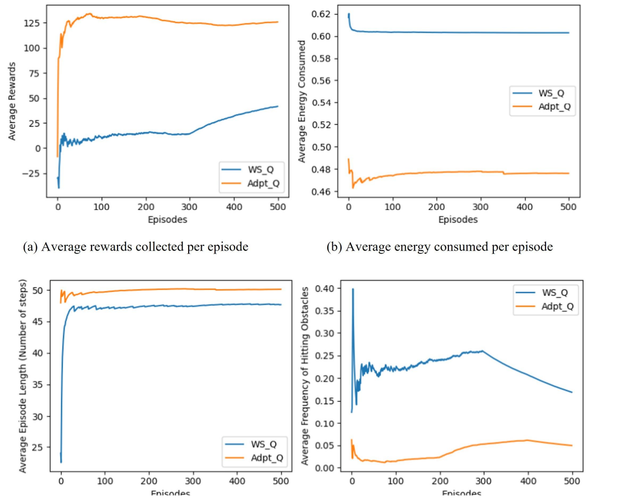
1. Simulation Settings

First, we illustrate our simulation settings (see. Table 1), followed by simulation results and discussions to evaluate the results. We evaluate the performance of our proposed scheme by comparing its performance against the traditional multi-objective AODV protocol with fixed weight on parameters considered. To demostrate the optimality of our proposed adaptive Q-routing scheme (Adp_Q), we test its initial performance on solving a pathfinding problem on a 5 by 5 dynamic Gridworld environment against the state of the art the traditional Q-routing which combines multiple objectives linearly based on the weighted sum approach (WS_Q). The WS_Q parameter weight is set at a fixed ratio of 0.1, 0.4 and 0.5 for the objectives related to reducing the path length, avoiding obstacles, and saving energy respectively. The aim is to observe the key performance indicators which includes their convergence speed by observing the average rewards collected, the average energy consumed by agents whle searching for a path, the average episode length and the average frequency of hitting obstacles.

Table 1. Dynamic gridWorld test parameter settings

Parameters	Value/ Range
Learning rate	0.3
Discount factor	0.85
Exploration rate	0.4
Minimum exploration rate	0.01
Exploration decay rate	0.95

The reward function were designed to include three objectives: reducing the number of steps taken to reach the goal state, avoiding obstacles, and optimizing energy consumption. Our proposed adaptive Q-routing algorithm incorporates the additional shaping reward function based on the Manhattan distance between the current state and the goal state. The experiment environment was designed to include various states such as; the goal state (4, 4), the initial state (0, 0), obstacle states, high energy consumption states, recharging point states, and normal states. Initially starting with 5 units of energy, the path finding learning agent loses energy by 0.1 units at each step during state transition. The environment at any given time has up to a maximum of three obscles located at different randomly selected states changing after every 100 learning episodes. The learning agent is pernalized by 10 units of energy whenever it hit one of the obstcles. Meanwhile awarded 10 units of energy whenever it goes through the recharging point. Finally, whenver an agent runs out of energy while navigating it is pernalized by an additional 20 units of energy. Settings for the key parameters such of this experiments were set as summarized below in Table 1 and the obtained results are given in Fig. 3.



(a) Average rewards collected per episode (b) Average energy consumed per episode

(c) Average episode length per episode (d) Average frequency of hitting obstacles per episode

Fig. 3. Performance the proposed Q-routing on a dynamic GridWorld environment

From the learning curve, Fig. 3 a above, it shows that, the proposed scheme scheme reaches optimal point relatively faster than the traditional Q-routing by averagely collecting more rewards. Eventhough, the proposed scheme performs slightly less by tending to take longer routes, it is effective in avoiding energy expensive and obstacle states.

Testing a virtual network environment, we compare the performance of the two schemes in dynamic network conditions with varied number of data flow and the node density. We compare their performance by observing important network parameters such as the packet delivery ratio, end-to-end delay, and network lifetime parameters. We simulate our network environment 20 times to obtain the averaged results. Parameter settings is summarized in Table 2 below.

Table 2. Simulation parameters

Parameters	Value/ Range
Simulation time (s)	200
Traffic type	UDP
Packet size (Bits)	512
Number of traffic sources	2, 4, 6, 8, and 10
Number of source/destination node (s)	5
Node maximum speed	20 m/s
Initial energy of nodes	10 J
Mobility model	Random Waypoint model
Packet generation rate (pkts/sec)	5
Discount factor, learning rate	0.8, 0.85

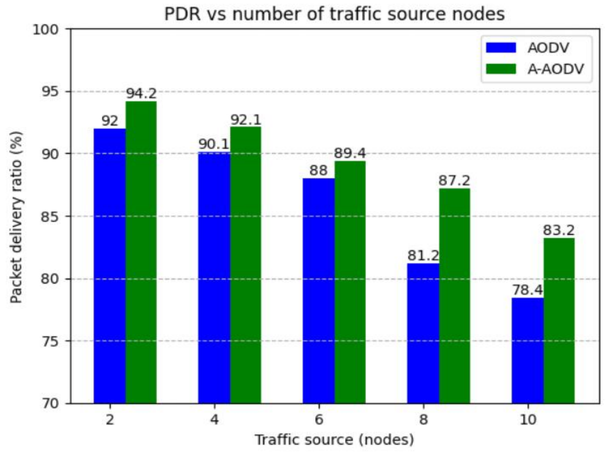


Fig. 4. PDR vs Number of nodes

2. Results Analysis

2.1 Packet Delivery Ratio of Routing Protocols

First, we present and discuss the performance of the 2 routing protocol schemes in a network environment of randomly distributed 20 nodes and observe how their PDR is affected when the number of traffic sources is varied. Shown in Fig. 4, the results show that the proposed scheme performs relatively better by almost 10% to 20% during low and higher traffic flow respectively compared to the traditional AODV routing protocol.

This is because, unlike the traditional AODV which prefers shorter paths without considering real network conditions, The A-AODV can select routing paths and make intelligent decisions because of its adaptability and capability to distinguish the constantly changing network environment of the ad-hoc wireless nodes and make confident decisions through the proposed algorithm. It can update its routing decision based on the newly learnt network dynamics and adjust accordingly, resulting to an improved network performance compared to the traditional AODV routing protocol.

2.2 End-to-End Delay of Routing Protocols

Next, we look at the E2E simulation results between the two schemes as shown in Fig. 5. From the results, it can be seen that the traditional AODV performs relatively low compared to our proposed scheme. This is due to the fact that, initial routes selected by the traditional AODV routing scheme do not change fast enough because in AODV link change occurs when the current link expires (no updates received through exchanged Hello or control messages). This mechanism often fails to relate to the real-network conditions and hence leads to overused links. As a result, nodes along such paths are congested.

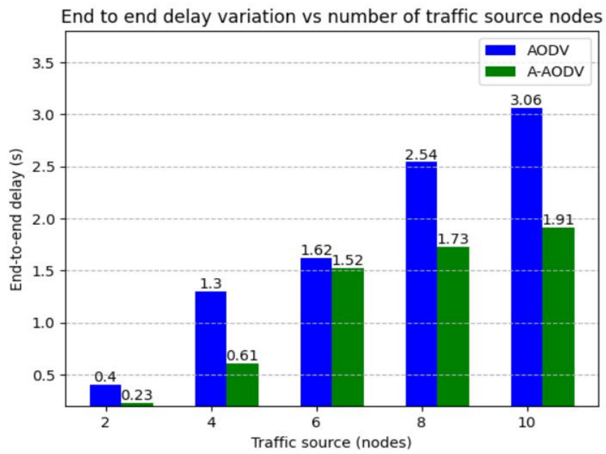


Fig. 5. Delay vs Number of nodes

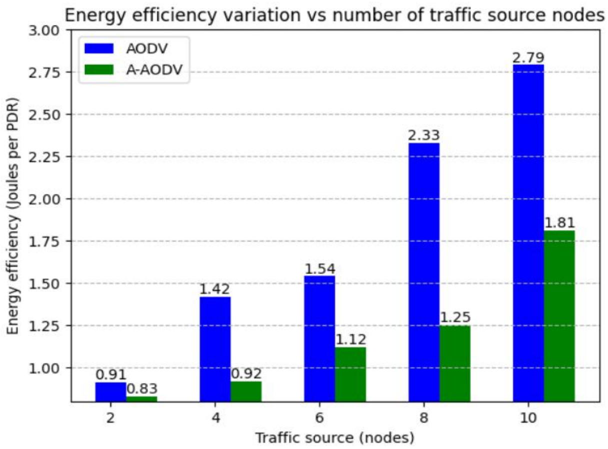


Fig. 6. Energy efficiency vs Number of nodes

Our proposed routing scheme is capable of constantly changing routes when necessary according to the current network conditions. The A-AODV algorithm delivers packets relatively faster due to its load-balance aware features which guarantees fair traffic distribution across multiple paths in the network.

2.3 . Energy Efficiency of Routing Protocols

Measured as the ratio of the percentage of packets delivered to the average of total energy consumed during communication time, improving this parameter is of our most concern. As it can be seen in Fig. 6, the performance of both routing protocols gradually drops as the number of traffic increases in the network. This is because, as traffic flow increases PDR drops and the average energy consumed during transmission increases.

Regardless, our proposed scheme seems to perform relatively better by maintaining good PDR and distributed low energy consumption across multiple nodes participating in routing.

With the A-AODV, ad-hoc nodes require less average energy to successfully transmit data packets to the specified destination nodes. The incorporated shaped rewards provide additional information when selecting routes, thereby allowing for the use of paths that a relatively shorter and energy efficient unlike those selected by the AODV scheme. The A-AODV scheme can monitor the energy levels of nodes, detect energy depletion or bottlenecks, and reroute traffic to more energy-efficient paths. This adaptability allows it to achieve energy efficient routing with multi-objective optimization, surpassing the generic energy-AODV routing scheme.

Ⅳ. CONCLUSIONS

In summary, we have proposed a novel routing protocol named the A-AODV to dynamically select routes from with primary goal of optimizing energy consumption of MANETs. The algorithm is capable of speeding up the learning process by incorporating additional shaping rewards unlike the traditional AODV algorithm. By employing the RL algorithm with shaped rewards in intelligent routing, our A-AODV is capable of finding reliable and energy-efficient routing policies with competitive performance. Following this, in our future work, we plan to introduce and study the effect of different parameters and test other methods to apply to the shaped reward with multi-objective optimization in routing for better performance in networking.

References

[1] M. A. Elsadig, Y. A. Fadlalla, “Mobile ad hoc Network Routing Protocols: Performance Evaluation and Assessment,” International Journal of Computing and Digital Systems, Vol. 7, No.1, pp. 59-66, 2018.

[2] S. K. Das, S. Tripathi, "Intelligent Energy-aware Efficient Routing for MANET. Wireless Networks,” Vol. 24, pp. 1139-1159, 2018.

[3] C. Perkins, E. Belding-Royer, S. Das, "RFC3561: Ad hoc on-demand Distance Vector (AODV) Routing.," 2003.

[4] D. B. Johnson., D. A. Maltz, J. Broch, "DSR: The Dynamic Source Routing Protocol for Multi-hop Wireless ad hoc Networks,” Ad hoc Networking. Vol. 5, No. 1, pp. 139-172, 2001.2

[5] A. H. Mohsin, "Optimize Routing Protocol Overheads in MANETs: Challenges and Solutions: A Review Paper,” Wireless Personal Communications. Vol. 126, No. 4, pp. 2871-2910, 2022.

[6] R. A. Diab, N. Bastaki, A. Abdrabou, "A Survey on Routing Protocols for Delay and Energy-constrained Cognitive Radio Networks," IEEE Access, Vol. 8, pp. 198779-198800, 2020.
[7] J. Boyan, M. Littman, "Packet Routing in Dynamically Changing Networks: A Reinforcement Learning Approach," Advances in Neural Information Processing Systems, Vol. 6 1993.
[8] J. Liu, Q. Wang, C. T. He, K. Jaffrès-Runser, Y. Xu, Z. Li, Y. J. Xu, "QMR: Q-learning Based Multi-objective Optimization Routing Protocol for Flying ad hoc Networks." Computer Communications, Vol. 150, pp. 304-316, 2020.
[9] L. Campanile, M. Gribaudo, M. Iacono, F. Marulli, M. Mastroianni, "Computer Network Simulation with ns-3: A Systematic Literature Review," Electronics. Vol. 9, No. 2, pp. 272, 2020.
[10] S. Padakandla, "A Survey of Reinforcement Learning Algorithms for Dynamically Varying Environments," ACM Computing Surveys (CSUR). Vol. 54, No. 6, pp. 1-25, 2021.
[11] B. Dai, Y. Cao, Z. Wu, Z. Dai, R. Yao, Y. Xu, "Routing Optimization Meets Machine Intelligence: A Perspective for the Future Network," Neurocomputing, Vol. 459, pp. 44-58, 2021.
[12] A. Serhani, N. Naja, A. Jamali, "A Q-Routing: Mobility-, Stability-aware Adaptive Routing Protocol for Data Routing in MANET - IoT Systems," Cluster Computing, Vol. 23, No. 1, pp. 13-27, 2020.
[13] D. G. Zhang, L. Chen, J. Zhang, J. Chen, T. Zhang, Y. M. Tang, J. N. Qiu, "A Multi-path Routing Protocol Based on Link Lifetime and Energy Consumption Prediction for Mobile Edge Computing," IEEE Access, Vol. 8, pp. 69058-69071, 2020.

Sun Beibei

2006 Computer Engineering from Shandong University (B.S.)
2012 Software Engineering from Shandong University (M.S.)
2022~Intelligent Communication from Semyung Univsersity (PhD)

Field of Interests: Wireless network, IoT, Machine Learning, Computer networks

Email: cherrybaby0718@gmail.com