



PHÂN TÍCH CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN HIỆU QUẢ CỦA VIỆC ÁP DỤNG HỌC TĂNG CƯỜNG VÀO ĐIỀU KHIỂN ĐỊNH TUYẾN TRONG MẠNG KHÔNG DÂY HÌNH LƯỚI

Lê Hữu Bình^{1*}, Lê Đức Huy²

¹ Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế, Việt Nam

² Khoa Công nghệ thông tin, Trường Đại học Kinh doanh và Công nghệ Hà Nội, Việt Nam

Tóm tắt. Việc ứng dụng học tăng cường vào điều khiển định tuyến trong mạng không dây hình lưới đang thu hút sự quan tâm của nhiều nhóm nghiên cứu. Các công trình nghiên cứu gần đây đã chỉ ra rằng định tuyến dựa trên học tăng cường mạng lại hiệu quả cao hơn so với các giao thức định tuyến truyền thống, đặc biệt là trong các mô hình mạng không dây hình lưới với tải lưu lượng cao, vùng diện tích rộng, sử dụng nhiều bộ định tuyến. Trong bài báo này, chúng tôi tập trung phân tích các yếu tố ảnh hưởng đến hiệu quả của việc áp dụng học tăng cường vào điều khiển định tuyến. Từ đó, đưa ra các khuyến nghị phù hợp nhằm nâng cao hiệu năng mạng khi sử dụng các giao thức định tuyến dựa trên học tăng cường.

Từ khóa: Mạng không dây hình lưới, định tuyến dựa trên học tăng cường, hiệu năng mạng

Analysis of factors influencing the efficiency of applying reinforcement learning to routing control in wireless mesh network

Le Huu Binh^{1*}, Le Duc Huy²

¹ Faculty of Information Technology, University of Sciences, Hue University, Hue City, Vietnam

² Faculty of Information Technology, Ha Noi University of Business and Technology, Vietnam

Abstract. The application of reinforcement learning to routing control in mesh wireless network is attracting the attention of many research groups. Recent works have shown that reinforcement learning-based routing is more efficient than traditional routing protocols, especially in mesh wireless network models with heavy traffic loads, wide area, and multiple routers. In this paper, we focus on investigating the factors that affect the effectiveness of applying reinforcement learning to routing control. Thence, we provide appropriate recommendations for improving network performance when employing reinforcement learning-based routing protocols.

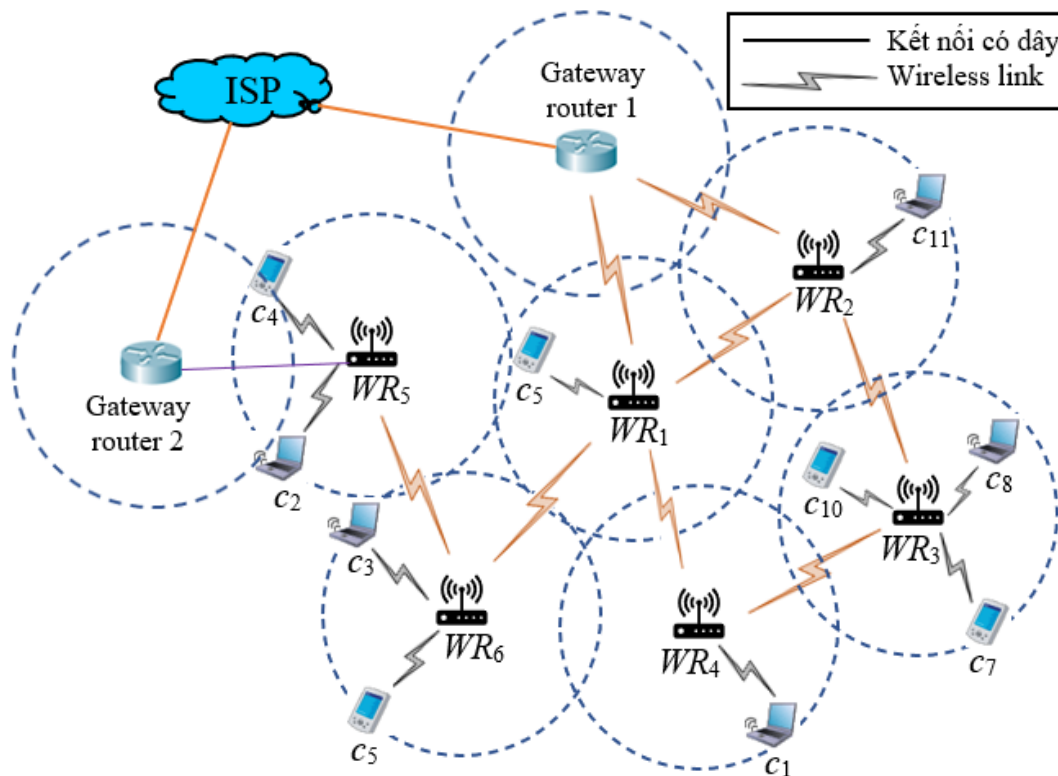
* Liên hệ: lhbinh@hueuni.edu.vn

Keywords: Wireless mesh network, reinforcement learning-based routing protocol, network performance.

1 Giới thiệu

Công nghệ mạng không dây hình lưới (Wireless Mesh Networks - WMN) đang ngày càng phát triển và dần thay thế mạng không dây sử dụng điểm truy nhập (Access Point) truyền thống. WMN đang nhận được sự quan tâm đặc biệt từ các nhà cung cấp dịch vụ cũng như người sử dụng mạng. Hiện nay, WMN đang là lựa chọn hàng đầu trong việc triển khai mạng cục bộ không dây (Wireless Local Area Network - WLAN) của các cơ quan, doanh nghiệp, trường học vì nó có nhiều ưu điểm. Trong công tác triển khai lắp đặt, WMN thuận tiện đối với các hệ thống mạng WLAN có vùng diện tích rộng và sử dụng nhiều bộ định tuyến. Về mặt cung cấp dịch vụ, WMN phù hợp với các hệ thống mạng WLAN mà tải lưu lượng phát sinh thường ngày lớn vì khả năng cân bằng tải của các bộ định tuyến. Về mặt quản trị hệ thống, WMN thường sử dụng cơ chế quản trị tập trung dựa trên giải pháp mạng điều khiển bằng phần mềm (Software Defined Networking - SDN) nên rất thuận tiện cho người quản trị trong công tác cài đặt, cấu hình và giám sát hệ thống. Hình 1. cho ta thấy một ví dụ điển hình của mạng WMN thường sử dụng trong WLAN. Thành phần chính của một mạng WMN là các bộ định tuyến không dây (Wireless Router - WR). Các WR trong vùng phủ sóng của nhau sẽ tạo một kết nối không dây để kết nối giữa chúng. Tập các WR và các kết nối không dây tạo thành một tô-pô hình lưới. Trong số các WR, chỉ cần một vài WR kết nối trực tiếp đến bộ định tuyến của nhà cung cấp dịch vụ mạng (Gateway Router - GR). Các kết nối giữa WR và GR có thể qua môi trường không dây hoặc có dây sử dụng cáp xoắn đôi CAT5/CAT6 hoặc cáp quang. Như ví dụ ở Hình 1, có ba WR kết nối trực tiếp đến GR là WR₁, WR₂ và WR₅, trong đó WR₁, WR₂ kết nối đến GR₁ qua môi trường không dây, còn WR₅ kết nối đến GR₂ bằng cáp xoắn đôi CAT6. Các WR còn lại kết nối đến GR thông qua ba WR này. Mỗi người sử dụng (client) kết nối đến một WR trong vùng phủ sóng để sử dụng các dịch vụ mạng. Nếu có nhiều WR phủ sóng đến một client, theo mặc định thì client đó sẽ kết nối đến WR có cường độ tín hiệu mạnh nhất.

Mặc dù WMN đã được triển khai trong thực tế, nhưng về các giao thức điều khiển việc truyền dữ liệu qua mạng còn nhiều vấn đề cần được nghiên cứu để giải quyết, đặc biệt là trong các hệ thống mạng có vùng diện tích rộng, tải lưu lượng lớn, băng thông rộng, mật độ nút cao, sử dụng nhiều WR. Một số hướng nghiên cứu điển hình về WMN đã được triển khai trong thời gian gần đây như bài toán quy hoạch mạng WMN [1, 2, 3], bài toán định tuyến [4, 7, 8, 9, 10, 11], chỉ định điểm truy nhập [12, 13]. Trong số các chủ đề này, định tuyến là chủ đề nhận được sự quan tâm nhiều nhất, vì các kỹ thuật định tuyến tối ưu cho phép cải thiện nhiều độ đo hiệu năng trong mạng WMN.



Hình 1. Một ví dụ của mạng không dây hình lưới [21]

Định tuyến trong mạng WMN thường sử dụng các giao thức cơ bản của mạng không dây tùy biến. Các giao thức được sử dụng nhiều nhất là AODV (Ad hoc On-demand Distance Vector) [15], DSDV (Destination Sequenced Distance Vector routing), DSR (Dynamic Source Routing) [16]. Đây là các giao thức chuẩn của mạng không dây tùy biến, sử dụng tiêu chí tổng số chặng (hop-count) để chọn lộ trình. Trong trường hợp mạng WMN có tải lưu lượng lớn và biến động nhiều, việc sử dụng các giao thức cơ bản nên trên bộ lộ một số nhược điểm, như tắc nghẽn cục bộ tại các nút xảy ra nhiều do các giao thức định tuyến này không có khả năng cân bằng tải. Trong trường hợp này, cần phải sử dụng các giao thức định tuyến thông minh hơn, có khả năng tính toán nhanh để thích ứng với sự thay đổi thường xuyên của trạng thái mạng. Ứng dụng học tăng cường (reinforcement learning) vào việc khám phá lộ trình của giao thức định tuyến là một giải pháp khả thi cho yêu cầu này. Điều này đã được kiểm nghiệm thông qua các công trình nghiên cứu đã công bố gần đây [7, 8, 9, 10, 11, 14]. Kết quả đạt được của một số công trình nghiên cứu tiêu biểu được tóm tắt như ở Bảng 1.

Bảng 1. Một số công trình nghiên cứu tiêu biểu về định tuyến dựa trên học tăng cường

Công trình	Phương pháp học	Độ đo sử dụng cho giá trị phản hồi (reward)	Kết quả đạt được
[7]	Học qua gói ACK, cải tiến từ giao thức AODV	Tỷ lệ tín hiệu trên nhiễu (SNR), tải lưu lượng trên các kết nối	Đề xuất thuật toán định tuyến cải tiến từ AODV, tăng thông lượng mạng, giảm độ trễ.
[9]	Học qua gói ACK	Năng lượng của nút láng giềng	Đề xuất thuật toán định tuyến cân bằng năng lượng dựa trên Q-Learning, tăng thời gian tồn tại của các nút, giảm tỷ lệ tổn thất gói dữ liệu.
[10]	Học qua gói ACK	Tải lưu lượng trên các kênh	Đề xuất thuật toán định tuyến dựa trên Q-Learning, cải thiện xác suất nghẽn mạng.
[11]	Học qua gói Hello, thay đổi linh hoạt hệ số tỷ lệ học để xác định điều kiện ràng buộc của QoS	Tỷ lệ truyền gói dữ liệu thành công	Đề xuất thuật toán định tuyến có tên QGIR, giảm tỷ lệ nghẽn gói dữ liệu, tăng thông lượng mạng.
[14]	Học qua gói ACK	Chất lượng đường truyền	Đề xuất thuật toán RLBPR, tăng thông lượng, giảm độ trễ.
[18]	Sử dụng kiến trúc SDN, học tập trung tại SDN controller	Xác suất nghẽn gói, SNR	Đề xuất thuật toán định tuyến IRSML, tăng SNR, thông lượng mạng và độ trễ.
[21]	Học qua gói ACK và Hello	Tổng số chặng (Hop count)	So sánh hiệu năng của phương pháp học qua gói ACK và gói Hello, xác định được ưu nhược điểm của các phương pháp

Nhóm tác giả trong [14] đã sử dụng học tăng cường để đề xuất một thuật toán định tuyến có tên RLBPR (Reinforcement Learning-based Best Path Routing) cho mạng WMN. Để lựa chọn lộ trình đến GR, thuật toán RLBPR sử dụng phương pháp Q-Learning với giá trị của hàm phản hồi (reward) là xác suất nghẽn gói dữ liệu. Độ đo này được lưu trữ trong bảng giá trị Q tại mỗi nút, làm cơ sở cho các bộ định tuyến không dây chọn trình để truyền dữ liệu đến đích. Các kết quả mô phỏng sử dụng NS-2 đã cho thấy rằng thuật toán RLBPR mang lại hiệu quả hơn các thuật toán khác về độ trễ và thông lượng mạng. Cũng với phương pháp sử dụng học tăng cường, nhóm tác giả trong [11] đã đề xuất một thuật toán định tuyến thông minh đảm bảo chất lượng dịch vụ (Quality of Sercive - QoS) cho mạng WMN với tải lưu lượng cao. Các tác giả đã xây dựng một hàm phản hồi với độ đo là tỷ lệ truyền gói dữ liệu thành công đến đích. Hàm phản hồi này được

sử dụng cho việc cập nhật bảng định tuyến tại mỗi nút thông qua thuật toán Q-Learning. Ngoài ra, hệ số tỷ lệ học của thuật toán Q-Learning được thay đổi một cách linh hoạt để xác định điều kiện ràng buộc về thời gian trễ từ nguồn đến đích. Bằng phương pháp mô phỏng, các tác giả đã cho thấy rõ thuật toán định tuyến được đề xuất mang lại hiệu năng mạng cao hơn các thuật toán phổ biến.

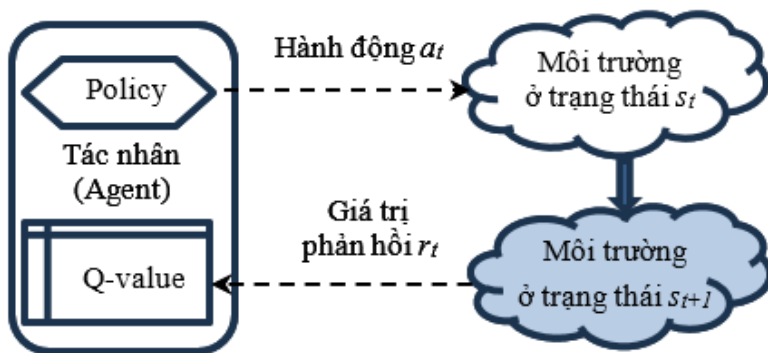
Kết quả của các công trình nghiên cứu được khảo sát ở trên đã cho thấy rõ hiệu quả của việc áp dụng học tăng cường vào điều khiển định tuyến trong mạng WMN. Để nâng cao hiệu quả của kỹ thuật định tuyến này, việc nghiên cứu các yếu tố ảnh hưởng đến hiệu năng của thuật toán định tuyến là điều cần thiết, nhằm đưa ra các khuyến nghị phù hợp để việc áp dụng thuật toán định tuyến dựa trên học tăng cường mang lại hiệu quả tốt nhất. Đây chính là mục tiêu nghiên cứu của bài báo này.

Các phần tiếp theo của bài báo được bố cục như sau: Phần 2 trình bày cách thức áp dụng học tăng cường vào điều khiển định tuyến trong mạng WMN. Các kết quả mô phỏng thực hiện trên hệ mô phỏng mã nguồn mở OMNeT++ và INET framework được phân tích chi tiết trong phần 3. Cuối cùng là kết luận và đề xuất hướng phát triển tiếp theo của chủ đề nghiên cứu này, được trình bày trong phần 4.

2 Ứng dụng của học tăng cường vào điều khiển định tuyến trong WMN

2.1 Cơ bản về học tăng cường

Nguyên lý cơ bản của học tăng cường được minh họa như ở Hình 2. Thành phần chính của một hệ thống học tăng cường là các thực thể thực hiện nhiệm vụ học được gọi là các tác nhân (agent). Các tác nhân thực hiện việc học thường xuyên bằng cách tương tác với môi trường thông qua các hành động để làm thay đổi môi trường, đồng thời thu được các giá trị phản hồi (reward), làm cơ sở cho việc lựa chọn hành động ở các lần học sau.



Hình 2. Nguyên lý cơ bản của học tăng cường

Xét một thời điểm t , giả sử môi trường lúc đó đang ở trạng thái s_t , khi cần thực hiện nhiệm vụ học, tác nhân tương tác với môi trường bằng hành động a_t . Với hành động này, môi trường chuyển từ trạng thái s_t thành trạng thái s_{t+1} , đồng thời tác nhân nhận được một giá trị phản hồi r_t . Ở các lần học tiếp theo, dựa trên các giá trị phản hồi thu được ở các lần học trước, tác nhân lựa chọn hành động sao cho giá trị phản hồi thu được là tốt nhất. Chính sách này được gọi là chính sách ε -greedy. Gọi $Q(s_t, a_t)$ là tổng giá trị giá trị phản hồi nhận được khi tác nhân thực hiện hành động a_t tại trạng thái s_t , bằng cách áp dụng thuật toán Q-Learning, giá trị $Q(s_t, a_t)$ được xác định như sau [7]:

$$Q(s_t, a_t) = (1 - \alpha)Q(s_t, a_t) + \alpha[R(s_t, a_t) + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})] \quad (1)$$

trong đó $\alpha \in [0, 1]$ là hệ số tỷ lệ học và $\gamma \in [0, 1]$ là hệ số chiết khấu.

Để sớm thu được giá trị Q tốt nhất, thuật toán Q-learning thường sử dụng chính sách ε -greedy để lựa chọn hành động tại mỗi thời điểm. Gọi $A(s_t)$ là tập hành động khả thi khi môi trường ở trạng thái s_t , $\pi(a_t, s_t)$ là xác suất mà tác nhân lựa chọn hành động a_t tại trạng thái s_t . Theo chính sách ε -greedy ta có [17]:

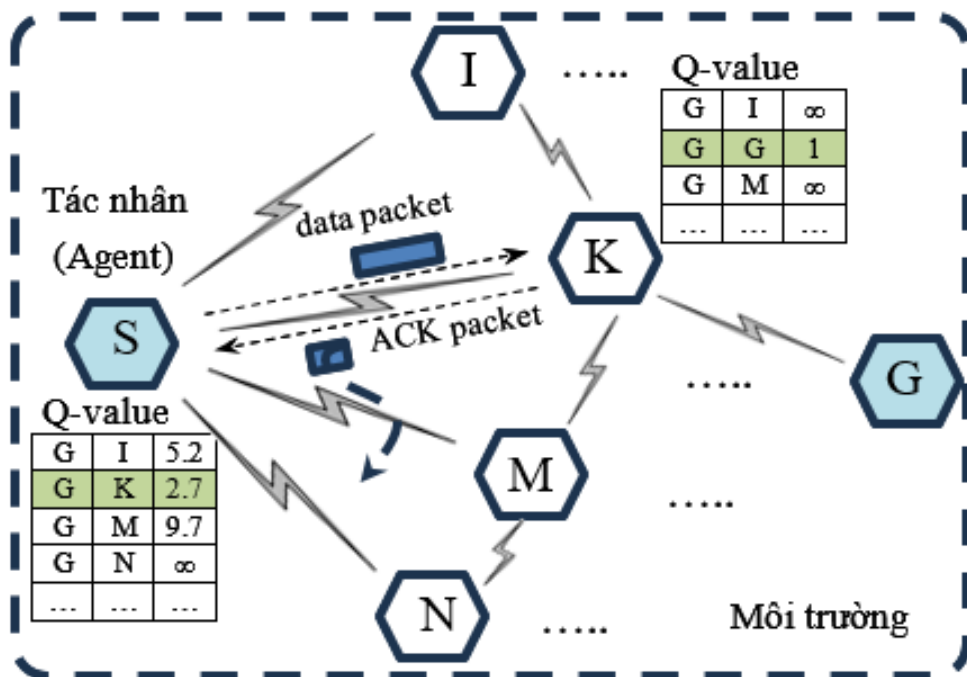
$$\pi(a_t, s_t) = \begin{cases} 1 - \frac{\varepsilon}{|A(s_t)|} & \text{nếu } a_t = a_t^* \\ \frac{\varepsilon}{|A(s_t)|} & \text{ngược lại} \end{cases} \quad (2)$$

trong đó ε là một giá trị đủ nhỏ, thường thiết lập bằng 0.1, a_t^* được xác định như sau:

$$a_t^* = \max_{A(s_t)} Q(a_t, s_t) \quad (3)$$

2.2 Áp dụng vào điều khiển định tuyến

Để áp dụng học tăng cường vào điều khiển định tuyến, chúng tôi mô hình hóa hệ thống mạng thành mô hình học tăng cường như cho thấy ở Hình 3. dựa trên ý tưởng như trong [14]. Trong trường hợp này, tác nhân là các nút mạng, cụ thể là các bộ định tuyến không dây (WR). Môi trường là toàn bộ hệ thống mạng. Tập trạng thái đối với tác nhân là tập các nút láng giềng tại thời điểm xem xét, tình trạng của các kết nối đến các nút láng giềng đó. Tập hành động là gửi gói dữ liệu đến một nút láng giềng và nhận gói ACK. Bảng giá trị Q tương ứng với bảng định tuyến của mỗi nút, giá trị phản hồi (reward) chính là trọng số của lộ trình tương ứng.



Hình 3. Mô hình hóa bài toán định tuyến thành mô hình học tăng cường

Thuật toán 1 là mã giả lập của thuật toán cập nhật bảng định tuyến tại mỗi nút. Tại thời điểm khởi tạo, mỗi nút tạo một bảng Q-value với mỗi bản ghi có cấu trúc $\{G_x, N, Q\}$, trong đó, G_x là GR đích, N là nút tiếp theo dọc theo lộ trình từ nút hiện tại đến G_x và Q là trọng số của lộ trình tương ứng. Vì độ đo định tuyến được sử dụng trong thuật toán này là tổng số chặng (hop-count), nên ở thời điểm khởi tạo, trọng số Q của mỗi lộ trình $\{G_x, N, Q\}$ được thiết lập bằng 1 nếu nút hiện tại có kết nối trực tiếp đến G_x , bằng ∞ trong trường hợp ngược lại (các bước từ 1 đến 9 trong thuật toán 1). Xét ví dụ như Hình 3, bảng Q-value của nút K ở thời điểm khởi tạo gồm có 3 bản ghi, vì K có 3 nút láng giềng, trong đó nút G là GR nên bản ghi tương ứng với nút láng giềng này có giá trị Q bằng 1. Hai nút láng giềng còn lại không phải là GR nên giá trị Q bằng ∞ . Bảng Q-value của mỗi nút được cập nhật thường xuyên trong quá trình truyền dữ liệu thông qua các gói dữ liệu và gói ACK theo các bước từ 10 đến 26 của thuật toán 1. Trở lại ví dụ ở Hình 3, xét bảng Q-value của nút S, ở thời điểm khởi tạo, tất cả các bản ghi đến nút G đều có giá trị Q là ∞ vì tất cả các nút láng giềng của nút S ở thời điểm hiện tại đều không phải là GR. Sau một số lần cập nhật giá trị Q theo các bước 10 đến 26 của thuật toán 1, giá trị Q sẽ hội tụ dần về số hop-count nhỏ nhất để đi đến G. Ở trạng thái hiện tại, lộ trình tốt nhất để đi đến G là $S \rightarrow K \rightarrow G$.

Thuật toán 1. Cập nhật bảng định tuyến tại mỗi nút sử dụng Q-learning

Đầu vào: Một tô-pô mạng WMN và các tham số kỹ thuật của hệ thống mạng, hệ số tỷ lệ học (α), hệ số chiết khấu (γ).

Đầu ra: Bảng Q-values của tất cả các WR.

Các bước thực hiện:

```

1:  For (mỗi nút  $I$  trong mạng)           //Khởi tạo bảng Q-value cho mỗi node
2:    For (mỗi nút  $J$  là láng giềng của  $I$ )
3:      If ( $J$  là GR) //  $J \equiv G$ 
4:        Thêm bản ghi  $\{G, G, 1\}$  vào bảng Q-value của nút  $I$ ;
5:      else
6:        Thêm bản ghi  $\{G, J, \infty\}$  vào bảng Q-value của nút  $I$ ;
7:      End if;
8:    End for;
9:  End for;
    //Cập nhật bảng Q-value
10: For (mỗi nút  $I$  trong mạng)
11:   If ( $I$  nhận được một gói dữ liệu)
12:     Chọn một bản ghi trong bảng Q-value theo chính sách  $\varepsilon$ -greedy sử dụng (2);
13:     Truyền gói dữ liệu đến nút láng giềng được chọn ở bước 11 (giả sử nút  $J$ );
14:     Chuyển sang bước 18;
15:     Chờ nhận gói ACK;
16:   End if;
17: End for;
    //Tại nút  $J$ 
18: If ( $J$  nhận được gói dữ liệu từ  $I$ )
19:   Tìm giá trị tốt nhất ( $Q_{\min}$ ) từ  $J$  đến GR trong bảng Q-value;
20:   Tạo gói ACK, lưu giá trị  $Q_{\min}$  vào gói ACK;
21:   Gửi phản hồi gói ACK về nút  $I$ ;
22: End if
    //Tại nút  $I$ 
23: If ( $I$  nhận được gói ACK từ  $J$ )
24:   Đọc giá trị  $Q_{\min}$  trong gói ACK;
25:   Cập nhật bảng Q-value theo phương trình (1);
26: End if;

```

3 Các yếu tố ảnh hưởng đến hiệu năng của thuật toán định tuyến dựa trên học tăng cường

Hiệu năng của thuật toán định tuyến dựa trên học tăng cường phụ thuộc chủ yếu vào độ hội tụ của thuật toán, cụ thể là thời gian mà các WR học được bảng định tuyến. Từ nguyên lý hoạt động của thuật toán định tuyến như được trình bày ở phần trên ta thấy rằng, độ hội tụ của thuật toán phụ thuộc chủ yếu vào tải lưu lượng phân phối đến các WR, tổng số WR trong mạng. Ngoài ra, hệ số tỷ lệ học (α) và hệ số chiến lược (γ) trong phương trình cập nhật Q-value (phương trình (1)) cũng có ảnh hưởng đến độ hội tụ của thuật toán.

3.1 Ảnh hưởng của tải lưu lượng

Theo nguyên lý hoạt động của thuật toán định tuyến dựa trên học tăng cường (thuật toán 1), các WR cập nhật bảng Q-values mỗi khi nó nhận được một gói phản hồi ACK từ một nút láng giềng. Để một nút nhận được gói ACK, nút đó phải thực hiện việc truyền dữ liệu. Vì vậy, số lần cập nhật bảng Q-values phụ thuộc vào tải lưu lượng phân phối đến mỗi nút. Có hai đặc trưng của của lưu lượng trong mạng ảnh hưởng đến độ hội tụ của thuật toán. Một là mật độ lưu lượng, đặc trưng này xác định tần suất cập nhật bảng Q-values của các WR. Hai là đặc tính phân bố của tải lưu lượng. Tải lưu lượng phân bố càng đồng đều cho tất cả các WR trong mạng thì thuật toán càng nhanh hội tụ.

3.2 Ảnh hưởng của tổng số WR

Tổng số WR trong mạng là một yếu tố có ảnh hưởng lớn đến độ hội tụ của thuật toán định tuyến dựa trên học tăng cường. Để các WR có thể học được lộ trình tốt nhất đến các nút đích, mỗi WR cần phải có thông tin về Q-value của tất cả các lộ trình từ WR đó đến đích. Điều này chỉ có thể đạt được nhờ việc trao đổi thông tin giữa các WR. Với thuật toán định tuyến dựa trên học tăng cường, việc trao đổi thông tin về Q-value được thực hiện thông qua việc truyền gói dữ liệu và gói ACK giữa các WR. Vì vậy, tổng số WR trong mạng WMN có ảnh hưởng đến độ hội tụ của thuật toán định tuyến. Số WR càng lớn thì thời gian hội tụ càng chậm do mỗi WR cần phải thực hiện nhiều lần cập nhật Q-value hơn để tìm được lộ trình tốt nhất đến đích.

4 Kết quả mô phỏng và thảo luận

Trong phần này, chúng tôi sử dụng phương pháp mô phỏng để phân tích các yếu tố ảnh hưởng đến hiệu năng của thuật toán định tuyến sử dụng học tăng cường được trình bày ở trên.

Các mô hình mô phỏng được cài đặt trên hệ điều hành Ubuntu 22.04, sử dụng phần mềm mã nguồn mở OMNeT++ 6.0.3 [19] và INET framework 4.5.2 [20].

4.1 Kịch bản mô phỏng

Các tham số trong Bảng 1 mô tả các kịch bản mô phỏng được thiết lập. Các mô hình mạng WMN được cài đặt trong vùng diện tích 1000×1000 [m²], tổng số bộ định tuyến không dây (WR) thay đổi trong khoảng từ 30 đến 50 tùy theo kịch bản mô phỏng, bán kính phủ sóng của mỗi WR là 250 [m]. Giao thức MAC được sử dụng là IEEE 802.11ac với tần số sóng mang là 2.4 [GHz], tốc độ dữ liệu của mỗi kênh là 54 [Mbps]. Tải lưu lượng phát sinh trung bình tại mỗi nút từ 1 đến 10 [Mbps], sử dụng các luồng UDP. Vì tính ngẫu nhiên của thuật toán, nên để đảm bảo độ chính xác của các kết quả, mỗi kịch bản mô phỏng được thực thi lặp lại 10 lần, các kết quả được trình bày trong phần này là giá trị trung bình của 10 lần chạy.

Bảng 2. Các tham số của kịch bản mô phỏng

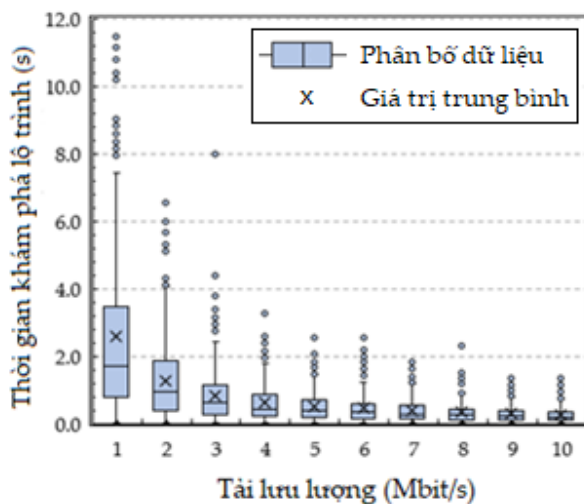
Tham số	Thiết lập
Tổng số WR	Từ 30 đến 50
Giao thức MAC	IEEE 802.11ac
Tốc độ dữ liệu	54 [Mbps]
Tần số sóng mang	2.4 [GHz]
Bán kính phủ sóng của các WR	250 [m]
Công suất phát của các WR	12 [dBm]
Độ nhạy thu của các WR	-76 [dBm]
Mô hình suy hao	Free space
Hệ số tỷ lệ học (α)	0.7
Hệ số chiết khấu (γ)	1.0
Tải lưu lượng	Từ 1 đến 10 [Mbps]
Loại lưu lượng	UDP
Vùng diện tích mạng WMN	1000×1000 [m ²]
Số lần chạy/mỗi kịch bản	10

4.2 Kết quả mô phỏng và thảo luận

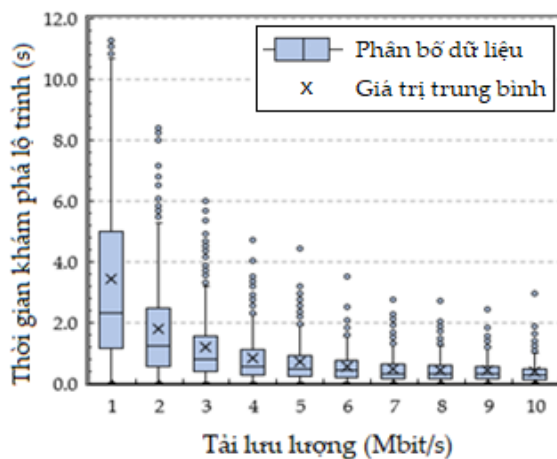
Đánh giá ảnh hưởng của tải lưu lượng

Trước tiên, chúng tôi phân tích ảnh hưởng của tải lưu lượng đến hiệu quả của việc áp dụng học tăng cường vào điều khiển định tuyến. Kết quả thu được ở Hình 4 cho thấy rõ ảnh hưởng của tải lưu lượng đến thời gian khám phá lộ trình trong trường hợp tổng số WR là 40. Độ đo về thời gian khám phá lộ trình trong kịch bản này được xác định từ khi mạng bắt đầu hoạt động cho đến khi tất cả các WR tìm được lộ trình đến GR.

Các biểu đồ hộp trên Hình 4 cho thấy rằng, tải lưu lượng càng cao thì thời gian khám phá lộ trình càng nhỏ, nghĩa là hiệu quả của thuật toán định tuyến càng cao. Xét trường hợp tải lưu lượng là 1 Mbps, thời gian khám phá lộ trình từ tất cả các WR đến GR trong khoảng từ 0,77 đến 11,47 giây, giá trị trung vị là 1,69 giây, trung bình là 2.60 giây. Khi tải lưu lượng tăng lên 10 Mbps, thời gian khám phá lộ trình giảm một cách đáng kể, chỉ trong khoảng từ 0,19 đến 1,36 giây, giá trị trung vị là 0,19 giây và trung bình là 0,27 giây. Nguyên nhân là do việc cập nhật bảng Q-value bằng thuật toán Q-learning ở các WR phụ thuộc vào mật độ gói dữ liệu đến tại mỗi nút. Theo nguyên lý hoạt động của thuật toán định tuyến dựa trên học tăng cường như đã trình bày ở phần trước, việc cập nhật bảng Q-value được thực hiện thông qua các gói dữ liệu và gói ACK. Vì vậy, khi mật độ gói dữ liệu đến càng cao, số lần cập nhật lại bảng Q-value bởi các tác nhân (WR) càng nhiều, dẫn đến thuật toán càng sớm hội tụ. Kết quả cũng hoàn toàn tương tự cho kịch bản 50 WR, như cho thấy ở Hình 5, thời gian khám phá lộ trình càng nhỏ khi tải lưu lượng càng lớn.

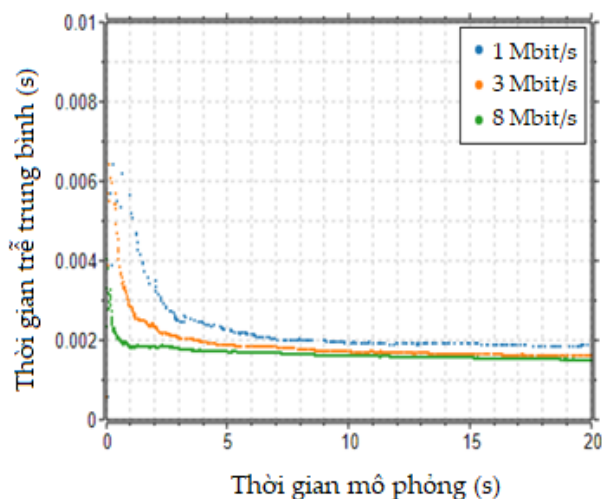


Hình 4. Ảnh hưởng của tải lưu lượng đến thời gian khám phá lộ trình khi tổng số WR là 40

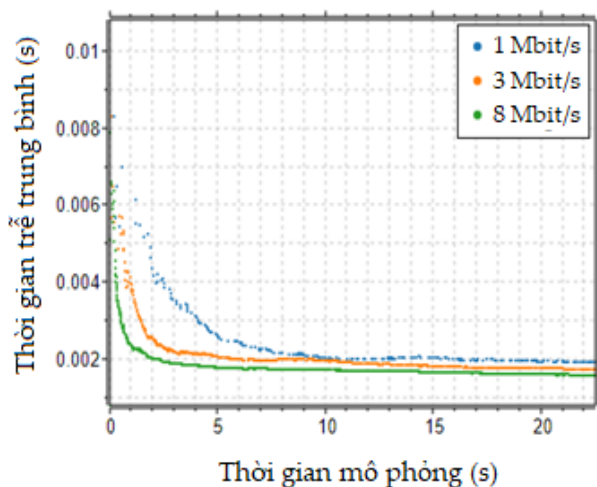


Hình 5. Ảnh hưởng của tải lưu lượng đến thời gian khám phá lộ trình khi tổng số WR là 50

Tiếp theo, chúng tôi phân tích ảnh hưởng của tải lưu lượng đến thời gian trễ từ nguồn đến đích (end-to-end delay - EED) của các gói dữ liệu. Các kết quả thu được như cho thấy ở Hình 6 và Hình 7, tương ứng với các kịch bản mô phỏng 40 WR và 50 WR. Ta thấy rằng, tải lưu lượng càng cao thì EED càng nhanh tiến đến giá trị ổn định. Xét trường hợp tô-pô mạng sử dụng 40 WR (kết quả ở Hình 6), nếu tải lưu lượng 1 Mbps thì cần đến gần 7 giây EED mới tiến đến giá trị ổn định. Trong khi đó, nếu tải lưu lượng 8 Mbps thì EED tiến đến giá trị ổn định chỉ sau hơn 1 giây. Nguyên nhân là do tải lưu lượng càng cao thì thuật toán định tuyến càng nhanh hội tụ như đã phân tích ở các kết quả trước. Điều này đồng nghĩa với việc các WR sớm có lộ trình tốt nhất để truyền dữ liệu, dẫn đến giảm EED.



Hình 6. Ảnh hưởng của tải lưu lượng đến thời gian trễ trung bình khi tổng số WR là 40

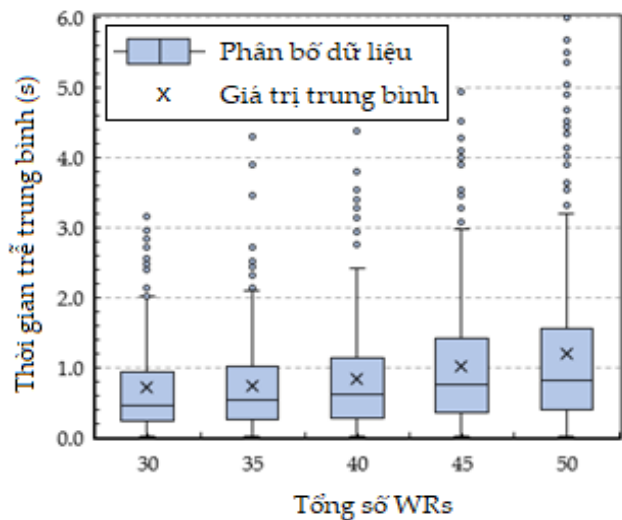


Hình 7. Ảnh hưởng của tải lưu lượng đến thời gian trễ trung bình khi tổng số WR là 50

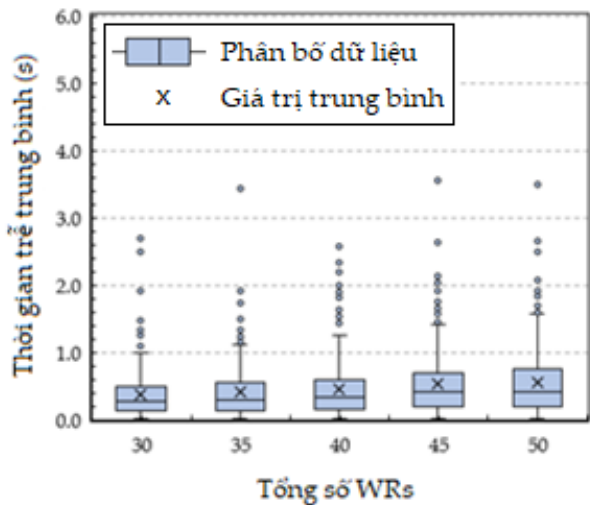
Đánh giá ảnh hưởng của tổng số WR

Một yếu tố khác cũng có ảnh hưởng lớn đến hiệu năng của kỹ thuật định tuyến dựa trên học tăng cường là tổng số WR. Chúng tôi cũng đã phân tích vấn đề này thông qua nhiều kịch bản mô phỏng khác nhau. Hình 8 là kết quả thu được khi thực thi 5 kịch bản mô phỏng với tổng số WR lần lượt là 30, 35, 40, 45 và 50, tải lưu lượng trung bình tại các WR là 3 Mbps. Các biểu đồ hộp trong Hình 8 cho thấy rằng khi tổng số WR càng nhiều thì thời gian khám phá lộ trình càng lớn. Cụ thể, xét kịch bản 30 WR, thời gian khám phá lộ trình phân bố trong khoảng từ 0.02 đến 3.13 giây, trung vị là 0.46 giây và trung bình là 0.71 giây. Với kịch bản 50 WR, thời gian khám phá lộ trình tăng lên khá lớn, phân bố trong khoảng từ 0.02 đến 5.98 giây, trung vị là 0.81 giây và trung bình là 1.19 giây. Nguyên nhân dẫn đến điều này là do khi tổng số WR nhiều, có một số lộ trình phải đi qua nhiều bước truyền và nhiều WR trung gian, muốn tìm thấy các lộ trình này thì các tác nhân (các WR) cần phải thực hiện nhiều lần học để cập nhật bảng Q-value, dẫn đến tăng thời gian tìm ra lộ trình.

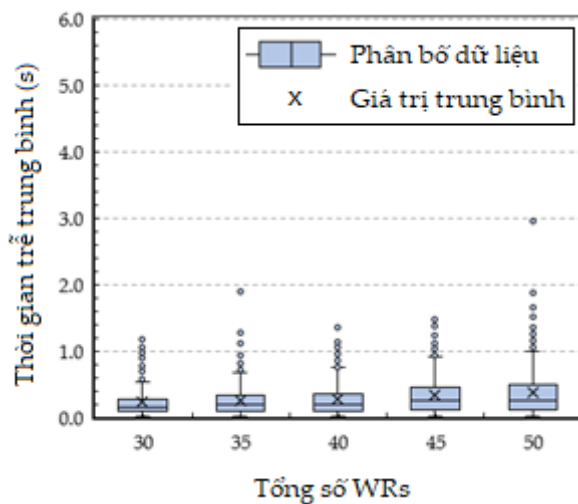
Trong trường hợp tải lưu lượng lớn hơn, thời gian khám phá lộ trình có giảm, tuy nhiên trường hợp hệ thống mạng có nhiều WR hơn thì vẫn có thời gian khám phá lộ trình lớn hơn. Điều này thể hiện rõ như các kết quả được trình bày trong Hình 9 và Hình 10, so sánh thời gian khám phá lộ trình của 5 kịch bản mô phỏng với tổng số WR lần lượt là 30, 35, 40, 45 và 50, tải lưu lượng trung bình tại các WR là 6 Mbps (Hình 9) và 10 Mbps (Hình 10).



Hình 8. Ảnh hưởng của tổng số WR đến thời gian khám phát lộ trình khi tải lưu lượng 3 Mbps

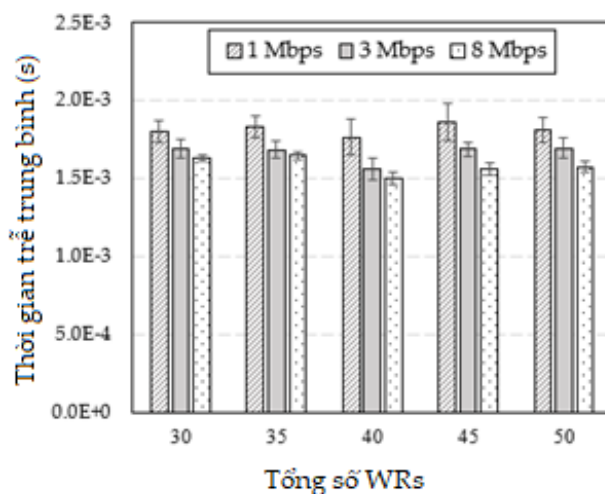


Hình 9. Ảnh hưởng của tổng số WR đến thời gian khám phát lộ trình khi tải lưu lượng 6 Mbps



Hình 10. Ảnh hưởng của tổng số WR đến thời gian khám phát lộ trình khi tải lưu lượng 10 Mbps

Ngoài việc ảnh hưởng đến thời gian khám phá lộ trình, tổng số WR cũng có ảnh hưởng đến thời gian trễ trung bình của các gói dữ liệu (EED), tuy nhiên ảnh hưởng là không nhiều. Các kết quả trên Hình 11 cho thấy rõ điều này. Chúng tôi đã phân tích các kịch bản với tổng số WR là 30, 35, 40, 45 và 50, tải lưu lượng trung bình là 1 Mbps, 3 Mbps và 8 Mbps. Khi tổng số WR thay đổi, sự khác nhau về EED là không nhiều. Chỉ khi tải lưu lượng thay đổi thì EED mới có sự khác biệt rõ ràng. Cụ thể là tải lưu lượng càng cao thì EED càng giảm, do thuật toán định tuyến hội tụ nhanh hơn khi tải lưu lượng cao hơn.



Hình 11. Thời gian trễ trung bình theo tổng số WR khi tải lưu lượng là 1 Mbps, 3 Mbps và 8 Mbps

Qua các kết quả mô phỏng được trình bày trong phần này, chúng tôi kết luận rằng, hai yếu tố có ảnh hưởng lớn đến hiệu quả của thuật toán định tuyến dựa trên học tăng cường là tải lưu lượng và tổng số WR. Cụ thể, tải lưu lượng càng cao thì thuật toán định tuyến hội tụ càng nhanh. Với yếu tố về tổng số WR, thuật toán hội tụ chậm hơn khi hệ thống mạng WMN có tổng số WR nhiều hơn.

5 Kết luận

Việc áp dụng học tăng cường vào điều khiển định tuyến trong mạng không dây hình lưới (WMN) là một giải pháp khả thi và hiệu quả để nâng cao hiệu quả của các giao thức định tuyến. Để có phương án sử dụng kỹ thuật định tuyến này một cách hiệu quả, việc nghiên cứu để các định rõ các yếu tố ảnh hưởng đến hiệu quả của kỹ thuật định tuyến tiên tiến này là điều cần thiết. Bài báo đã tập trung thực hiện công việc này. Kết quả nghiên cứu của bài báo đã xác định được hai yếu tố có ảnh hưởng lớn đến hiệu năng của kỹ thuật định tuyến dựa trên học tăng cường là tải lưu lượng và tổng số WR trong mạng. Tải lưu lượng càng cao thì thuật toán định tuyến hội tụ càng nhanh. Nhưng ngược lại, tổng số WR càng nhiều thì thuật toán hội tụ càng chậm. Vì vậy, kỹ thuật định tuyến dựa trên học tăng cường nên áp dụng cho các hệ thống mạng WMN có tải lưu lượng lớn. Trong trường hợp mạng WMN sử dụng nhiều WR, nếu tải lưu lượng lớn thì sử dụng kỹ thuật định tuyến dựa trên học tăng cường vẫn mang lại hiệu quả cao.

Trong hướng nghiên cứu tiếp theo, chúng tôi tiếp tục khảo sát việc ứng dụng của học tăng cường cho các hệ thống mạng không dây khác với độ biến động của trạng thái mạng nhiều hơn, như mạng không dây tùy biến (MANET), mạng không dây tùy biến với các thiết bị bay không người lái (Flying Ad-hoc NETwork - FANET).

Lời cảm ơn

Công trình nghiên cứu này được tài trợ bởi Đề tài Khoa học và công nghệ cấp Đại học Huế, mã số: DHH2023-01-204.

Tài liệu tham khảo

1. N. N. Abdelkader, A. Zibouda, A. Naouri, and H. Soufiene, "An efficient mesh router nodes placement in wireless mesh networks based on mothflame optimization algorithm," *International Journal of Communication Systems*, vol. 36, 2023.
2. L. H. Binh and T.-V. T. Duong, "A Novel and Effective Method for Solving the Router Nodes Placement in Wireless Mesh Networks using Reinforcement Learning," *PLOS ONE*, vol. 19, no. 4, e0301073, 2024.

3. T. S.Mekhmoukh, M. Yassine, G. A. Benmessaoud, M. Seyedali, Z. Atef, and R.-C. Amar, "Solving the Mesh Router Nodes Placement in Wireless Mesh Networks using Coyote Optimization Algorithm," *IEEE Access*, vol. 10, pp. 52 744–52 759, 2022.
4. E. O. Steven and E. Kamwesigye, "A Deep Learning-Based Routing Approach for Wireless Mesh Backbone Networks," *IEEE Access*, vol. 11, pp. 49 509–49 518, 2023.
5. C. Perkins, E. B. Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," RFC 3561.
6. D. Johnson, Y. Hu, and D. Maltz, "The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4," RFC4728.
7. L. H. Binh and T.-V. T. Duong, "An improved method of AODV routing protocol using reinforcement learning for ensuring QoS in 5G-based mobile ad-hoc networks," *ICT Express*, vol. 10, no. 1, pp. 97–103, 2024.
8. M. Boushaba, A. Hafid, and A. Belbekkouche, "Reinforcement learning-based best path to best gateway scheme for wireless mesh networks," in *2011 IEEE 7th Int. Conf. on Wireless and Mobile Computing, Netw. and Comm.*, 2011, pp. 373–379.
9. M. Yin, J. Chen, X. Duan, B. Jiao, and Y. Lei, "QEBR: Learning Based Routing Protocol for Energy Balance in Wireless Mesh Networks," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018, pp. 280–284.
10. A. R. Syed, K. A. Yau, J. Qadir, H. Mohamad, N. Ramli, and S. L. Keoh, "Route Selection for Multi-Hop Cognitive Radio Networks Using Reinforcement Learning: An Experimental Study," *IEEE Access*, vol. 4, pp. 6304–6324, 2016.
11. T.-V. T. Duong, L. H. Binh, and V. M. Ngo, "Reinforcement learning for QoS-guaranteed intelligent routing in Wireless Mesh Networks with heavy traffic load," *ICT Express*, vol. 8, no. 1, pp. 18–24, 2022.
12. G Vijaya Kumar, C Shoba Bindu, "A Study on Access Point Selection Algorithms in Wireless Mesh Networks", *Int. J. Advanced Networking and Applications*, vol. 6, no.1, pp. 2158-2167, 2014.
13. Alessandro Raschellà, Faycal Bouhafs, Michael Mackay, Qi Shi, Jorge Ortín, José Ramón Gállego, "Maria Canales, A dynamic access point allocation algorithm for dense wireless LANs using potential game", *Computer Networks*, vol.167, 106991, 2020.
14. M. Boushaba, A. Hafid, A. Belbekkouche, and M. Gendreau (2013), "Reinforcement learning based routing in wireless mesh networks", *Wireless Networks*, Vol. 19, No.8, pp. 2079-2091.
15. C. Perkins, E. B. Royer, and S. Das, "Ad hoc On- Demand Distance Vector (AODV) Routing," RFC 3561.
16. D. Johnson, Y. Hu, and D. Maltz, "The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4," RFC4728.
17. Thien T. T. Le and Sangman Moh, "An Energy-Efficient Topology Control Algorithm Based on Reinforcement Learning for Wireless Sensor Networks", *International Journal of Control and Automation*, Vol. 10, No. 5, pp.233-244, 2017.
18. T.-V. T. Duong and L. H. Binh, IRSML: An intelligent routing algorithm based on machine learning in software defined wireless networking, *ETRI Journal*, vol. 44, pp. 733–745, 2022.

19. András Varga and OpenSim Ltd, *OMNeT++ Simulation Manual - Version 6.x*, Copyright ©1992-2021. [Online]. Available: <http://www.omnetpp.org>.
20. A. Virdis and M. Kirsche, *Recent advances in network simulation - The OMNeT++ environment and its ecosystem*, Springer Nature Switzerland AG, 2019.
21. Lê Hữu Bình, Nguyễn Ngọc Thủy, “Khảo sát và đánh giá các phương pháp ứng dụng của học tăng cường cho định tuyến trong mạng không dây hình lưới”, *Tạp chí Khoa học và Công nghệ, Trường Đại học Khoa học, Đại học Huế*, Tập 22, số 1, trang 1-14, 2023.