

KHẢO SÁT VÀ ĐÁNH GIÁ CÁC PHƯƠNG PHÁP ỨNG DỤNG CỦA HỌC TĂNG CƯỜNG CHO ĐỊNH TUYẾN TRONG MẠNG KHÔNG DÂY HÌNH LƯỚI

Lê Hữu Bình^a, Nguyễn Ngọc Thủy^b

Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế

Email: ^alhbinh@hueuni.edu.vn, ^bnnthuy@hueuni.edu.vn

Ngày nhận bài: 5/5/2023; ngày hoàn thành phản biện: 16/5/2023; ngày duyệt đăng: 8/6/2023

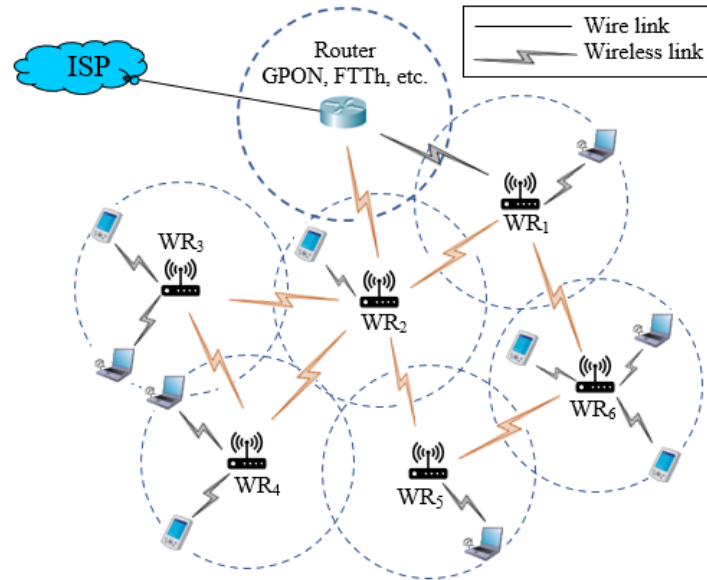
TÓM TẮT

Ứng dụng của học tăng cường vào các giao thức định tuyến trong mạng không dây là một chủ đề thu hút sự quan tâm của nhiều nhóm nghiên cứu trong thời gian gần đây. Đặc trưng cơ bản của các hệ thống mạng không dây là tải lưu lượng biến động lớn, tô-pô mạng thay đổi thường xuyên do các nút di chuyển. Vì vậy, việc ứng dụng học tăng cường để tính toán bảng định tuyến tại mỗi nút sao cho hiệu quả nhất là một thách thức lớn. Trong bài báo này, chúng tôi tập trung khảo sát và đánh giá các phương pháp ứng dụng của học tăng cường vào việc định tuyến trong mạng không dây hình lưới. Thông qua các kết quả mô phỏng trên OMNeT++, chúng tôi đưa ra các khuyến nghị phù hợp để cải thiện hiệu năng mạng khi sử dụng các giao thức định tuyến dựa trên học tăng cường.

Từ khóa: định tuyến, học tăng cường, mạng không dây hình lưới.

1. MỞ ĐẦU

Công nghệ mạng truy nhập không dây đã và đang được ứng dụng rộng rãi trong mạng nội bộ của các cơ quan, doanh nghiệp, trường học. Thời gian gần đây, giải pháp mạng truy nhập không dây tô-pô hình lưới (Mesh Wireless Networks - WMN) [1], [2] đã được nghiên cứu và triển khai trong thực tế. WMN có nhiều ưu điểm so với mạng truy nhập không dây sử dụng điểm truy nhập (Access Point) truyền thông, điển hình như giảm tình trạng tắc nghẽn do có khả năng cân bằng tải, thuận tiện trong việc triển khai hạ tầng do không cần phải kết nối đường truyền có dây đến tất cả các bộ định tuyến không dây. Hình 1 là một ví dụ của mạng WMN sử dụng 6 bộ định tuyến không dây (Wireless Router - WR). Các WR trong vùng phủ sóng của nhau được kết nối với nhau bằng một kết nối không dây tạo thành một tô-pô hình lưới. Ngoài ra, một số WR kết nối với bộ định tuyến của nhà cung cấp dịch vụ viễn thông để truy cập Internet. Các user kết nối đến hệ thống mạng thông qua các WR.



Hình 1. Một ví dụ của mạng truy nhập không dây hình lưới (WMN)

Đặc điểm vượt trội của công nghệ WMN là vùng phủ sóng rộng, băng thông cao, tốc độ dữ liệu lớn. Vì vậy, các giao thức điều khiển, mà điển hình là lớp giao thức định tuyến cần phải được thay đổi mới có thể khai thác hiệu quả tài nguyên mạng. Yêu cầu đặt ra với các giao thức định tuyến là tính toán nhanh, thích nghi với các mô hình mạng có tải lưu lượng cao, độ biến động lớn, có thể xét đến nhiều điều kiện ràng buộc để đảm bảo chất lượng truyền dẫn, chất lượng dịch vụ. Với các yêu cầu này, cần phải áp dụng các mô hình tính toán thông minh, có khả năng dự đoán được các độ đo hiệu năng tại các nút và trên các kết nối, nhằm cung cấp thông tin trạng thái một cách nhanh chóng, chính xác cho quá trình khám phá, duy trì và sử dụng lộ trình của giao thức định tuyến. Giải pháp hữu hiệu nhất để thực hiện điều này là sử dụng học tăng cường (Reinforcement learning) - một phương pháp học máy đã được ứng dụng hiệu quả vào bài toán định tuyến trong mạng WMN. Đây là một chủ đề thu hút sự quan tâm của nhiều nhóm nghiên cứu trong thời gian gần đây [3]-[8]. Nhóm tác giả trong [3] đã đề xuất một thuật toán định tuyến cho mạng WMN sử dụng học tăng cường, được đặt tên là RLBPR (Reinforcement Learning-based Best Path Routing). Thuật toán RLBPR sử dụng thông tin về tải lưu lượng tại các gateway để lựa chọn gateway tốt nhất cho việc truyền dữ liệu. Để lựa chọn lộ trình đến gateway, thuật toán RLBPR sử dụng phương pháp Q-Learning, với độ đo xác suất nghẽn gói dữ liệu được sử dụng như là "giá trị phản hồi" (reward) để lựa chọn bước truyền tiếp theo. Bằng phương pháp mô phỏng trên NS-2, các tác giả đã chứng minh rằng thuật toán RLBPR thực thi hiệu quả hơn các thuật toán khác nếu xét các độ đo hiệu năng về độ trễ và thông lượng mạng. Trong [6], các tác giả đã đề xuất hai cơ chế lựa chọn lộ trình dựa trên phương pháp học tăng cường để cải thiện hiệu năng của mạng vô tuyến đa chặng. Cả hai cơ chế sử dụng độ đo về thời gian kênh sẵn sàng tại các kết nối có tải lưu lượng cao làm giá trị phần thưởng của hàm Q- Learning. Độ đo

này cũng được sử dụng cho việc lựa chọn lộ trình giữa mỗi cặp nút nguồn - đích. Kết quả đánh giá bằng thực nghiệm đã chứng minh rằng, cả hai cơ chế được đề xuất thực thi hiệu quả hơn các cơ chế đã được công bố trước đó. Cũng với phương pháp học tăng cường, nhóm tác giả trong [9] đã đề xuất một giao thức định tuyến cân bằng năng lượng sử dụng thuật toán Q-learning cho mạng WMN, được đặt tên là QEBR. Giao thức này hoạt động theo nguyên lý của định tuyến phân tán. Trong giao thức QEBR, reward của hàm Q-learning là năng lượng của các nút láng giềng. Hiệu năng của giao thức QEBR cũng được đánh giá bằng phương pháp mô phỏng, sử dụng ngôn ngữ lập trình Python để so sánh với các giao thức định tuyến phổ biến trong mạng WMN.

Thông qua các công trình nghiên cứu đã được đề cập ở trên, chúng tôi có nhận xét rằng, việc ứng dụng kỹ thuật học tăng cường vào bài toán định tuyến trong mạng WMN là một giải pháp hữu hiệu nhằm nâng cao hiệu năng mạng. Trong bài báo này, chúng tôi tập trung khảo sát và đánh giá các phương pháp ứng dụng của học tăng cường cho bài toán định tuyến trong mạng WMN, để tìm ra giải pháp phù hợp tùy theo đặc trưng của hệ thống mạng.

Các phần còn lại của bài báo được tổ chức như sau: Phần 2 trình bày những vấn đề cơ bản của phương pháp học tăng cường và ứng dụng cho bài toán định tuyến trong mạng WMN. Phần 3 phân tích các kết quả mô phỏng được thực hiện trên OMNeT++. Cuối cùng là kết luận và hướng phát triển tiếp theo, được trình bày chi tiết trong phần 4.

2. HỌC TĂNG CƯỜNG VÀ ỨNG DỤNG TRONG BÀI TOÁN ĐỊNH TUYẾN

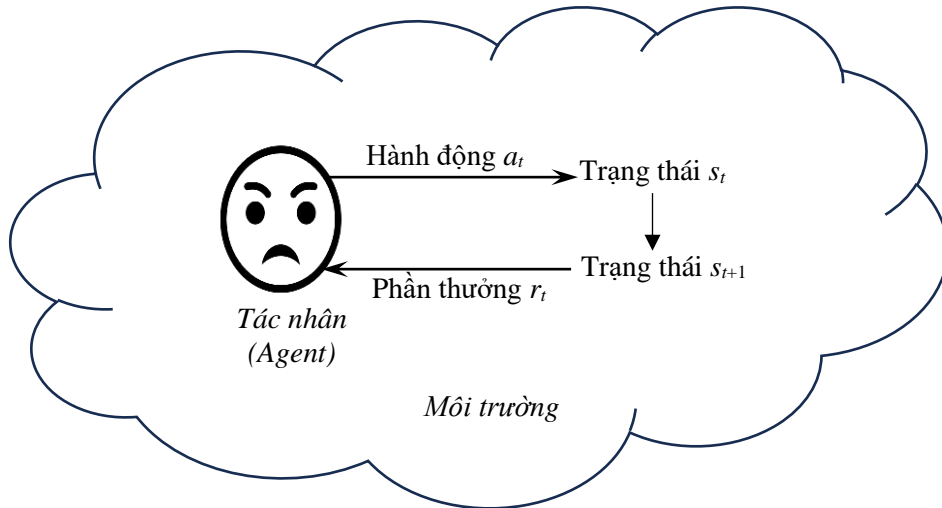
2.1. Cơ bản về học tăng cường

Học tăng cường là một dạng của học máy với nguyên lý cơ bản là hệ thống học từ các hành động trước đó để chọn hành động tốt hơn trong tương lai. Nguyên lý này được minh họa như ở Hình 2, trong đó các thực thể thực hiện nhiệm vụ học được gọi là các tác nhân (agent). Tại mỗi thời điểm t , tác nhân này 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 (reward) 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. Tổng giá trị giá trị phản hồi nhận được khi thực hiện hành động a_t tại trạng thái s_t là $Q(s_t, a_t)$, thường được xác định theo thuật toán Q-Learning như sau:

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

trong đó $\alpha \in [0, 1]$ là hệ số tỷ lệ học, xác định mức độ giá trị phản hồi mới được cập nhật cho giá trị phần thưởng cũ. Nếu $\alpha = 0$, tác nhân sẽ không học được gì, giá trị phản hồi

giữ nguyên như giá trị hiện tại. Nếu $\alpha = 1$, tác nhân chỉ xem xét thông tin mới nhất mà không quan tâm đến thông tin trước đó. Trong trường hợp các trạng thái của môi trường hoàn toàn xác định, hệ số học $\alpha = 1$ là tối ưu. Nếu tập trạng thái của môi trường phụ thuộc nhiều vào các yếu tố ngẫu nhiên, hệ số học α nên được thiết lập nhỏ hơn 1, thường là 0.5. $\gamma \in [0, 1]$ là hệ số chiết khấu, xác định mức độ ảnh hưởng của giá trị phản hồi trong tương lai. $R(s_t, a_t)$ là giá trị phản hồi nhận được khi tác nhân thực hiện hành động a_t tại trạng thái s_t .



Hình 2. Minh họa nguyên lý của học tăng cường

2.2. Ứng dụng của học tăng cường cho bài toán định tuyến

Học tăng cường đã được ứng dụng rộng rãi trong các bài toán kỹ thuật, điển hình như điều khiển rô bốt, trò chơi điện tử, điều khiển ô tô không người lái. Thời gian gần đây, phương pháp học máy này cũng đã được ứng dụng vào bài toán điều khiển định tuyến trong mạng không dây [1], [4], [5], [6], [7]. Để thực hiện điều này, quá trình định tuyến trong một hệ thống mạng không dây cần được mô hình hóa thành một mô hình học tăng cường với 5 yếu tố đặc trưng, đó là tác nhân, môi trường, trạng thái, hành động và giá trị phản hồi.

2.2.1. Tác nhân

Trong phương pháp học tăng cường, tác nhân là thực thể thực hiện việc học để đưa ra các hành động phù hợp cho hệ thống. Với bài toán định tuyến trong mạng không dây, tùy theo phương pháp định tuyến mà tác nhân có thể là các thành phần khác nhau. Cụ thể, nếu hệ thống mạng sử dụng kỹ thuật định tuyến phân tán, tác nhân chính là các nút cần phải thường xuyên cập nhật trạng thái mạng để tính toán bảng định tuyến. Nếu hệ thống mạng sử dụng kỹ thuật định tuyến tập trung, tác nhân chính là khối chức năng thực hiện việc định tuyến. Ví dụ, đối với các giao thức định tuyến tập trung

dựa trên mạng điều khiển bằng phần mềm (Software Defined Networking - SDN), tác nhân là bộ điều khiển SDN (SDN Controller).

2.2.2. Môi trường

Môi trường là những gì tồn tại xung quanh tác nhân, nơi mà tác nhân hoạt động và tương tác. Với bài toán định tuyến sử dụng học tăng cường, môi trường chính là hệ thống mạng, là tập hợp của tất cả các nút, các kết nối có dây và không dây, các yếu tố tác động lên hiệu năng của hệ thống mạng như nhiễu nhiệt, nhiễu tần số từ các thiết bị ngoài hệ thống mạng.

2.2.3. Trạng thái

Trạng thái là những gì mà tác nhân cảm nhận được. Trong trường hợp tác nhân là các nút mạng, mỗi trạng thái được xác định bởi tổng số nút láng giềng, tình trạng của các kết nối từ nút đó đến các nút láng giềng, trạng thái của các bộ đệm trong các nút, v.v. Tùy theo mục tiêu của thuật toán định tuyến mà mỗi trạng thái có thể được xác định bởi các tham số khác nhau.

2.2.4. Hành động

Hành động là cách thức mà các tác nhân tương tác với môi trường để cập nhật trạng thái. Với bài toán định tuyến sử dụng học tăng cường, tác nhân là các nút mạng. Do vậy, hành động mà một tác nhân tương tác với môi trường có thể là gửi một gói dữ liệu đến nút láng giềng [1], gửi một gói xác nhận (ACK) khi một nút nhận thành công một gói dữ liệu, gửi quảng bá một gói Hello hoặc gói Beacon đến tất cả các nút láng giềng [4].

2.2.5. Giá trị phản hồi

Sau khi thực hiện mỗi hành động tương tác với môi trường, tác nhân sẽ nhận được một giá trị phản hồi. Giá trị phản hồi của các hành động trước là cơ sở để tác nhân lựa chọn hành động tiếp theo với mục tiêu dần dần thu được giá trị phản hồi tốt nhất. Tổng giá trị phản hồi nhận được khi một tác nhân thực hiện hành động a_t tại trạng thái s_t là $Q(s_t, a_t)$, được xác định bởi thuật toán Q-Learning theo phương trình (1). Giá trị $Q(s_t, a_t)$ chính là mục tiêu của các thuật toán định tuyến sử dụng học tăng cường.

2.3. Học bảng định tuyến thông qua gói dữ liệu và gói ACK

Với các thuật toán định tuyến sử dụng học tăng cường, bảng định tuyến của các nút được cập nhật thường xuyên thông qua hành động mà tác nhân tương tác với môi trường, cụ thể là hành động mà mỗi nút tương tác với hệ thống mạng. Một phương pháp phổ biến để thực thi hành động này là sử dụng gói dữ liệu và gói ACK [1], [8]. Quy trình thực thi thuật toán định tuyến trong mạng không dây sử dụng học tăng cường thông qua gói dữ liệu và gói ACK được trình bày như ở Thuật toán 1. Trong trường hợp này, thuật toán định tuyến sử dụng trọng số phổ biến nhất của các giao thức định tuyến trong

mạng không dây, đó là tổng số chặng (hopCount). Vì vậy, giá trị Q trong hàm cập nhật phần thưởng theo thuật toán Q-Learning ở bước (11) được tính toán theo tổng số chặng. Giá trị Q tốt nhất để tác nhân lựa chọn hàng động tương tác với môi trường trong trường hợp này là giá trị Q nhỏ nhất, nghĩa là lộ trình được chọn để truyền dữ liệu là lộ trình có tổng số chặng nhỏ nhất.

Thuật toán 1. Cập nhật bảng định tuyến tại nút I sử dụng học tăng cường thông qua gói dữ liệu và gói ACK

- //Tại nút I
- (1) Đọc bảng định tuyến, tìm lộ trình có trọng số nhỏ nhất để truyền dữ liệu đến nút đích (D)
 - (2) **if** (tìm được lộ trình $\{D, J, Q_{i,j,d}\}$) **then**
 // $\{D, J, Q_{i,j,d}\}$: D là nút đích, J là nút tiếp theo và $Q_{i,j,d}$ là trọng số
 - (3) Truyền gói dữ liệu đến nút J ;
 //Tại nút J
 - (4) **if** (J nhận được gói dữ liệu từ nút I) **then**
 - (5) Đọc bảng định tuyến, xác định trọng số nhỏ nhất từ J đến D theo phương trình: $Q_{j,d}^{\min} = \min_{\forall n \in N_j} (Q_{j,n,d})$;
 - (6) Tạo gói ACK, lưu $Q_{j,d}^{\min}$ vào gói ACK;
 - (7) Gửi gói ACK về nút I .
 - (8) **endif**
 - //Tại nút I
 - (9) **if** (nút I nhận được gói dữ ACK từ nút J) **then**
 - (10) Nút I đọc $Q_{j,d}^{\min}$ trong gói ACK;
 - (11) Cập nhật giá trị $Q_{i,j,d}$ của lộ trình $\{D, J, Q_{i,j,d}\}$ trong bảng định tuyến theo thuật toán Q-Learning:
$$Q_{i,j,d} = (1 - \alpha)Q_{i,j,d} + \alpha(R_{i,j} + \gamma Q_{j,d}^{\min})$$
 - (12) **endif**;
 - (13) **else**
 - (14) Chọn ngẫu nhiên nút $J \in N_i$
 - (15) Quay lại bước (3)
 - (16) **endif**;
-

2.4. Học bảng định tuyến thông qua gói Hello

Một phương pháp khác để các nút học bảng định tuyến với thuật toán học tăng cường là sử dụng gói Hello [4]. Với phương pháp này, tác nhân là các nút tương tác với môi trường là hệ thống mạng bằng cách gửi quảng bá gói Hello định kỳ đến tất cả các nút láng giềng. Với mỗi nút nhận được gói Hello, nó sẽ thực hiện cập nhật lại bảng định tuyến theo thuật toán Q-Learning. Quy trình thực thi thuật toán định tuyến trong mạng không dây sử dụng học tăng cường thông qua gói Hello được trình bày như ở Thuật toán 2.

Thuật toán 2. Cập nhật bảng định tuyến tại sử dụng học tăng cường thông qua gói Hello

- (1) **while (true)**
 - (2) **wait**(IntervalTime);
 - (3) Đọc bảng định tuyến của nút hiện hành (nút K), xác định trọng số nhỏ nhất từ K đến nút đích (D) theo phương trình:

$$Q_{k,d}^{\min} = \min_{\forall i \in N_k} (Q_{k,i,d});$$
 - (6) Tạo gói Hello, lưu $Q_{k,d}^{\min}$ vào gói Hello;
 - (7) Gửi quảng bá gói Hello đến tất cả các nút láng giềng của K ;
 - (8) **for** (mỗi nút nhận được gói Hello (nút I))
 - (9) Đọc giá trị $Q_{k,d}^{\min}$ trong gói Hello;
 - (10) Cập nhật giá trị $Q_{i,k,d}$ của lộ trình $\{D, K, Q_{i,k,d}\}$ trong bảng định tuyến theo thuật toán Q-Learning:

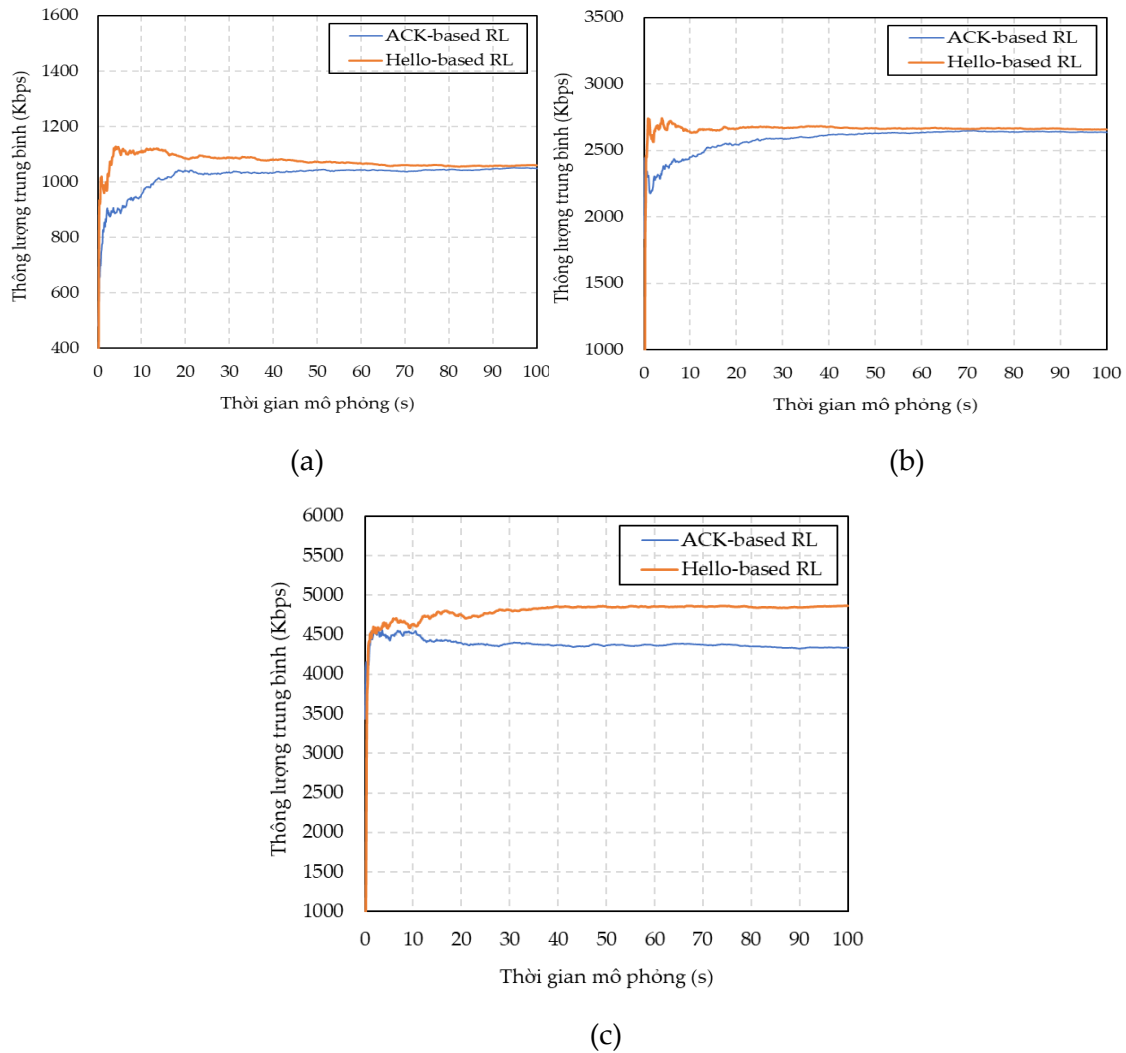
$$Q_{i,k,d} = (1-\alpha)Q_{i,k,d} + \alpha(R_{i,k} + \gamma Q_{k,d}^{\min})$$
 - (11) **endfor**;
 - (12) **endwhile**;
-

3. KẾT QUẢ MÔ PHỎNG VÀ THẢO LUẬN

Trong phần này, chúng tôi sử dụng phương pháp mô phỏng để so sánh, đánh giá hiệu quả thực thi của các thuật toán định tuyến sử dụng học tăng cường bằng hai phương pháp như được trình bày ở phần trước, một là học bảng định tuyến thông qua các gói dữ liệu và ACK, hai là học bảng định tuyến thông qua các gói Hello. Mô phỏng được thực thi trên OMNeT++ [10] và INET Framwork [11]. Kịch bản mô phỏng được thiết lập như trong Bảng 1. Chúng tôi sử dụng kịch bản ứng dụng điển hình của mạng WMN gồm có 17 bộ định tuyến không dây (WR), 50 nút truy nhập chính là các user. Mỗi nút truy nhập kết nối với một WR với cường độ sóng mạnh nhất để truy cập các dịch vụ mạng.

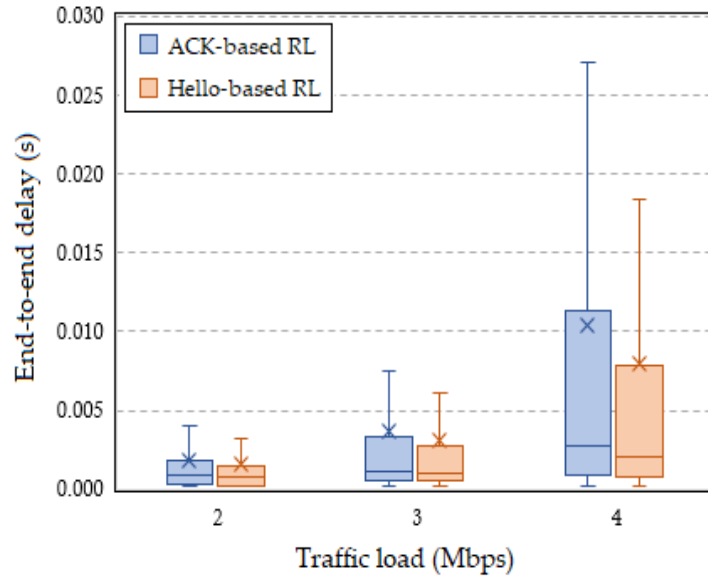
Bảng 1. Kịch bản mô phỏng

TT	Tham số	Giá trị
1	Vùng diện tích mô phỏng	1000 × 1000 [m ²]
2	Tổng số WR	17
3	Tổng số nút truy nhập	50 [nút]
5	Vùng truyền dẫn của mỗi WR	250 [m]
6	Giao thức MAC	IEEE 802.11ac
7	Tốc độ dữ liệu	54 [Mbps]
8	Công suất phát	12 [dBm]
9	Độ nhạy thu	-76 [dBm]
10	Hệ số tỷ lệ học trong thuật toán Q-Learning (α)	0.5
11	Hệ số giảm trong thuật toán Q-Learning (γ)	0.5
12	Loại lưu lượng	UDP
13	Kịch bản di chuyển của các nút truy nhập	Random Waypoint
14	Thời gian mô phỏng	500 giây



Hình 3. So sánh thông lượng của thuật toán định tuyến sử dụng học tăng cường thông qua gói ACK và gói Hello trong các trường hợp tải lưu lượng (a) 1.5 Mbps, (b) 3 Mbps và (c) 5 Mbps

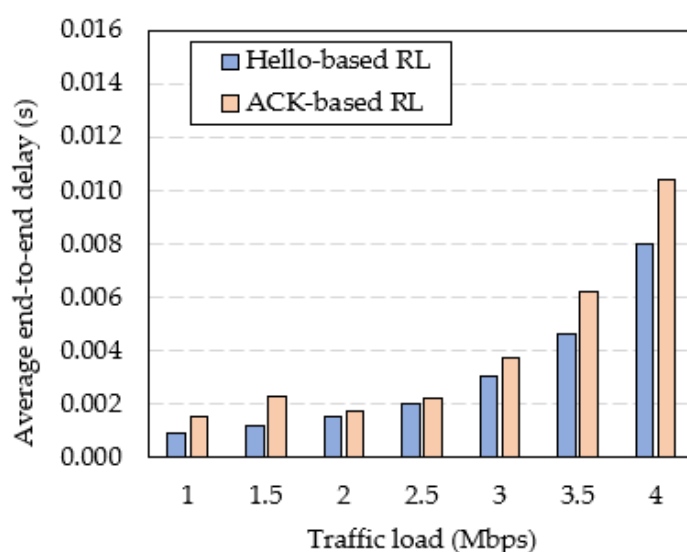
Các kết quả mô phỏng trong Hình 3 cho thấy thông lượng trung bình trên toàn mạng với các trường hợp tải lưu lượng là 1.5, 3 và 5 Mbps. Ta thấy rằng, với trường hợp học bảng định tuyến thông qua gói Hello, thông lượng thu được luôn luôn lớn hơn so với trường hợp sử dụng gói ACK. Đặc biệt, tải lưu lượng càng lớn thì chênh lệch về thông lượng giữa hai phương pháp này càng lớn. Điều này cho thấy phương pháp sử dụng gói Hello để học bảng định tuyến mang lại hiệu quả cao trong trường hợp tải lưu lượng lớn.



Hình 4. So sánh độ trễ của thuật toán định tuyến sử dụng học tăng cường thông qua gói ACK và gói Hello

Tiếp theo, chúng tôi phân tích độ trễ của hai phương pháp sử dụng gói ACK và gói Hello. Kết quả thu được như cho thấy ở Hình 4. Ta thấy rằng, phương pháp sử dụng gói Hello mang lại độ trễ thấp hơn so với phương pháp sử dụng gói ACK. Xét trường hợp tải lưu lượng 4 Mbps, phân bố thời gian trễ của các gói dữ liệu đối với phương pháp sử dụng gói ACK trong khoảng từ 0.186 ms đến 27.045 ms, giá trị trung vị là 2.78 ms, trung bình là 10.442 ms. Đối với trường hợp sử dụng gói Hello, thời gian trễ của các gói dữ liệu phân bố trong khoảng từ 0.186 ms đến 18.354 ms, giá trị trung vị là 2.087 ms, trung bình là 7.99 ms. Như vậy, phương pháp sử dụng gói Hello mang lại độ trễ thấp hơn phương pháp sử dụng gói ACK cả giá trị lớn nhất, trung vị và trung bình. Độ trễ trung bình giảm 2.452 ms.

Sự phụ thuộc của độ trễ trung bình theo tải lưu lượng cũng đã được khảo sát. Kết quả thu được như cho thấy ở Hình 5. Khi tải lưu lượng tăng từ 1 Mbps đến 4 Mbps, độ trễ trung bình của cả hai phương pháp sử dụng gói ACK và gói Hello tăng dần. Tuy nhiên, phương pháp sử dụng gói Hello luôn mang lại độ trễ trung bình thấp hơn so với phương pháp sử dụng gói ACK. Đặc biệt, tải lưu lượng càng cao thì phương pháp sử dụng gói Hello càng mang lại hiệu quả.



Hình 5. So sánh độ trễ trung bình theo tải lưu lượng của thuật toán định tuyến sử dụng học tăng cường thông qua gói ACK và gói Hello

4. KẾT LUẬN

Định tuyến trong mạng WMN sử dụng học tăng cường là một chủ đề thu hút sự quan tâm của nhiều nhóm nghiên cứu trong thời gian gần đây. Một số kết quả nghiên cứu đã công bố cho thấy rằng đây là một giải pháp mang lại hiệu quả cao trong việc cải thiện hiệu năng mạng. Trong bài báo này, chúng tôi tập trung khảo sát và đánh giá các phương pháp ứng dụng của học tăng cường vào việc định tuyến trong mạng không dây hình lưới. Cụ thể là tập trung vào hai phương pháp, học bảng định tuyến thông qua gói ACK và học bảng định tuyến thông qua gói Hello. Các kết quả mô phỏng trên OMNeT++ đã cho thấy rằng phương pháp sử dụng gói Hello mang lại hiệu quả cao hơn về mặt thông lượng và độ trễ, đặc biệt là trong trường hợp tải lưu lượng lớn.

Trong hướng nghiên cứu tiếp theo, chúng tôi tiếp tục phát triển các thuật toán định tuyến sử dụng học tăng cường để xét đến các điều kiện ràng buộc về chất lượng dịch vụ, chất lượng tín hiệu truyền dẫn nhằm cải thiện hơn nữa hiệu năng của mạng WMN.

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] I. F. Akyildiz and Xudong Wang, *Wireless Mesh Networks*, John Wiley & Sons Ltd, 2009.
- [2] Y. Zhang, J. Luo, and H. Hu, *Wireless Mesh Networking - Architectures, Protocols and Standards*, Taylor & Francis Group, LLC, 2007.
- [3] 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.
- [4] Thuy-Van T. Duong, L. H. Binh, and V. M. Ngo (2022), "Reinforcement learning for QoS-guaranteed intelligent routing in Wireless Mesh Networks with heavy traffic load", *ICT Express*, Vol. 8, No. 1, pp. 18-24.
- [5] Z. Mammeri (2019), "Reinforcement learning based routing in networks: Review and classification of approaches", *IEEE Access*, Vol. 7, pp. 55916–55950.
- [6] A. R. Syed, K. A. Yau, J. Qadir, H. Mohamad, N. Ramli, and S. L. Keoh (2016), "Route selection for multi-hop cognitive radio networks using reinforcement learning: An experimental study", *IEEE Access*, Vol. 4, pp. 6304-6324.
- [7] Thuy-Van T. Duong, L. H. Binh (2022), "IRSML: An intelligent routing algorithm based on machine learning in software defined wireless networking", *ETRI Journal*, Vol. 44, pp. 733–745.
- [8] Mai Cường Thọ, Nguyễn Thị Hương Lý, Lê Hữu Bình, Võ Thanh Tú (2022), "VAQR: Một tiếp cận học tăng cường trong định tuyến FANET", *Kỷ yếu Hội nghị Khoa học Quốc gia lần thứ XV về Nghiên cứu cơ bản và Ứng dụng Công nghệ thông tin - FAIR'2022*, NXB Khoa học Tự nhiên và Công nghệ, trang 51-58.
- [9] M. Yin, J. Chen, X. Duan, B. Jiao and Y. Lei (2018), "QEBR: Q-Learning Based Routing Protocol for Energy Balance in Wireless Mesh Networks," *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2018, pp. 280-284.
- [10] András Varga and OpenSim Ltd, *OMNeT++ Simulation Manual - Version 6.x*, Copyright ©1992-2021. [Online]. Available: <http://www.omnetpp.org/documentation/>.
- [11] A. Virdis and M. Kirsche, *Recent advances in network simulation - The OMNeT++ environment and its ecosystem*, Springer Nature Switzerland AG, 2019.

SURVEY AND EVALUATION OF APPLICATION METHODS OF REINFORCEMENT LEARNING FOR ROUTING IN WIRELESS MESH NETWORKS

Le Huu Binh^a, Nguyen Ngoc Thuy^b

Faculty of Information Technology, University of Sciences, Hue University

Email: ^alhbinh@hueuni.edu.vn, ^bnnthuy@hueuni.edu.vn

ABSTRACT

The use of reinforcement learning to routing algorithms in wireless mesh networks has recently piqued the interest of numerous research groups. The fundamental aspect of wireless mesh network systems is that traffic load varies substantially, and network topology changes often owing to node movement. As a result, using reinforcement learning to efficiently compute the routing table at each node is a significant problem. In this paper, we survey and evaluate reinforcement learning application methods for routing in mesh wireless networks. We propose relevant recommendations based on simulation results using OMNeT++ to increase network performance while employing reinforcement learning-based routing algorithms.

Keywords: Routing, reinforcement learning, wireless mesh network.



Lê Hữu Bình sinh năm 1978 tại Quảng Trị. Ông tốt nghiệp Kỹ sư ngành Điện tử Viễn thông năm 2001 tại Trường Đại học Bách khoa, Đại học Đà Nẵng và Thạc sĩ ngành Khoa học máy tính năm 2007 tại Trường Đại học Khoa học, Đại học Huế. Ông nhận học vị Tiến sĩ ngành Hệ thống thông tin năm 2020 tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Hiện nay, ông là giảng viên Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Công nghệ mạng thế hệ mới, ứng dụng của học máy và trí tuệ nhân tạo trong công nghệ mạng, mô phỏng và thiết kế mạng.



Nguyễn Ngọc Thủy sinh năm 1976 tại Thừa Thiên Huế. Ông tốt nghiệp Cử nhân ngành Công nghệ thông tin năm 1998 và Thạc sĩ ngành Khoa học máy tính năm 2005 tại Trường Đại học Khoa học, Đại học Huế. Hiện nay, ông là giảng viên Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Công nghệ mạng máy tính, mô phỏng và thiết kế mạng, an ninh mạng.