DOI: 10.15625/vap.2022.0206

VAQR: MỘT TIẾP CẬN HỌC TĂNG CƯỜNG TRONG ĐỊNH TUYẾN FANET

Mai Cường Thọ[1,2], Nguyễn Thị Hương Lý[1], Lê Hữu Bình[2], Võ Thanh Tú[2]

[1] Khoa Công nghệ thông tin, Trường Đại học Nha Trang

[3]Khoa Công nghệ thông tin, Trường Đại học khoa học, Đại học Huế

mctho@hueuni.edu.vn, lynth@ntu.edu.vn, lhbinh@hueuni.edu.vn, vttu@hueuni.edu.vn

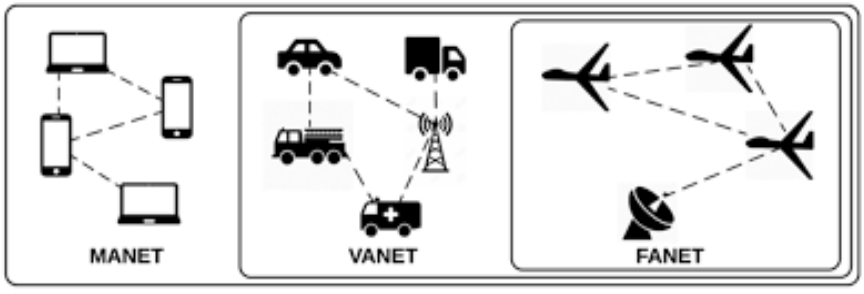
TÓM TẮT: *Bài báo này trình bày kết quả nghiên cứu về ứng dụng của học tăng cường cho các giao thức định tuyến trong mạng FANET. Một thuật toán định tuyến dựa trên vị trí (VAQR) được đề xuất với ý tưởng xây dựng một hàm phần thưởng sử dụng hai độ đo vận tốc tương đối và góc, bảng định tuyến của mỗi nút được cập nhật sử dụng thuật toán Q-learning với hàm phần thưởng được đề xuất. Kết quả mô phỏng trên OMNeT++ cho thấy rằng thuật toán được đề xuất đạt được hiệu năng cao về mặt tỉ lệ chuyển gói thành công và thông lượng so với giao thức GPSR, với trễ truyền tải trung bình cao hơn trong phạm vi chấp nhận được.*

Từ khóa: *FANET routing, Q-routing, GPSR.*

I. GIỚI THIỆU

Trong thời gian gần đây phương tiện bay không người lái (UAV- Unmanned Aerial Vehicle) với kích thước nhỏ và khả năng bay đã nổi lên như một kỹ thuật đầy hứa hẹn trong các ứng dụng quân sự và dân sự, bao gồm trinh sát quân sự, tìm kiếm và cứu hộ, giám sát thiên tai, an ninh đô thị,... [1-3]. So với việc sử dụng một UAV đơn lẻ, các hệ thống nhiều UAV hiệu quả hơn nhiều với khả năng đa nhiệm nhanh hơn, thời gian sử dụng mạng lâu hơn và khả năng mở rộng cao hơn. Tuy nhiên chúng cũng mang lại nhiều vấn đề thách thức do các đặc tính riêng biệt của UAV (ví dụ: vận tốc di chuyển cao và triển khai thưa thớt).

Một trong những vấn đề cơ bản quan trọng nhất là liên lạc hợp tác giữa các UAV. Một giao thức liên lạc hoặc định tuyến hiệu quả giữa các UAV đóng một vai trò quan trọng trong việc truyền dữ liệu và các ứng dụng thực tế khác nhau. Để truyền các gói một cách hiệu quả, một nhóm UAV giao tiếp và cộng tác với nhau để tự tổ chức thành một mạng, được gọi là UAVNet (UAV ad-hoc Network) hay FANET (Flying Ad-hoc NETwork). FANET thực tế là trường hợp đặc biệt của VANET, còn VANET là trường hợp đặc biệt của MANET. Trong đó mạng tùy biến di động (Mobile Adhoc Network - MANET) là một mạng không dây đặc biệt, với ưu điểm là khả năng hoạt động độc lập không phụ thuộc vào cơ sở hạ tầng mạng cố định, chi phí thấp, triển khai nhanh và tính di động cao. Các nút trong mạng MANET phối hợp với nhau để truyền thông nên vừa là một host, vừa đảm nhận chức năng của bộ định tuyến.



Hình 1. Mô hình FANET và mối quan hệ với VANET, MANET

Trong vài năm qua, số lượng nghiên cứu về giao thức định tuyến trong mạng Ad-hoc ngày càng tăng nhanh chóng, nhưng cơ bản chúng không thể được sử dụng trực tiếp cho FANET. Việc phát triển một giao thức định tuyến hiệu quả cho tính di động, đặc biệt là các UAV tốc độ cao đó là một nhiệm vụ đầy thách thức [4].

Các giao thức định tuyến cho UAV thường sử dụng định tuyến đa chặng, các gói tin được chuyển tiếp từng chặng một, tức là một nhóm UAV cộng tác với nhau để chuyển các gói thông qua các đường dẫn định tuyến đa chặng. Do đó, trong định tuyến đa chặng, việc lựa chọn một UAV tiếp theo phù hợp là bước cốt lõi. Dựa trên các chiến lược khác nhau để lựa chọn chặng tiếp theo, định tuyến đa chặng có thể được phân loại thêm thành các loại: định tuyến dựa trên cấu trúc liên kết (Topology-based) và định tuyến dựa trên vị trí (Position-based), định tuyến kết hợp topo-vị trí, định tuyến lấy cảm hứng từ sinh học sử dụng trí thông minh bầy đàn (Bio-Inspired), định tuyến dựa trên học tăng cường (Reinforcement Learning Approach). Về tổng quan các giao thức đề xuất cho FANET được đánh giá với mô hình di động chưa thiên hướng ứng dụng và di chuyển trong không gian 2 chiều, điều này chưa phù hợp với FANET.

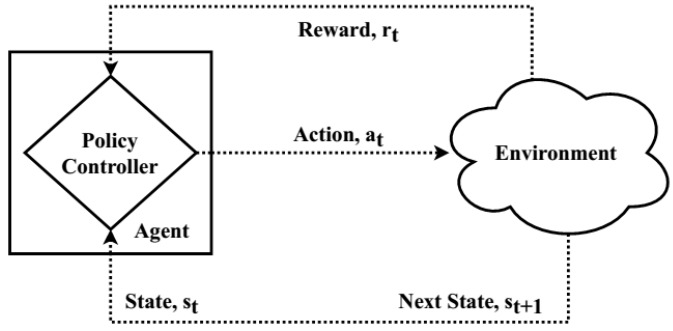
Mỗi loại định tuyến có những ưu điểm riêng và phù hợp với bối cảnh ứng dụng nhất định. Tuy vậy qua nghiên cứu các công trình đã công bố gần đây liên quan đến định tuyến FANET, chúng tôi nhận thấy hướng ứng dụng học tăng cường trong định tuyến cho mạng không dây nói chung, MANET/VANET/FANET nói riêng đã đem lại những kết quả tích cực. Đây cũng chính là động cơ nghiên cứu của chúng tôi trong bài bào này.

Các phần tiếp theo của bài báo được bố cục như sau. Phần II trình bày phương pháp học tăng cường và ứng dụng của nó trong bài toán định tuyến. Phần III khảo sát các công trình nghiên cứu liên quan đến định tuyến trong mạng FANET sử dụng học tăng cường. Phần IV trình bày thuật toán đề xuất của chúng tôi và các kết quả thử nghiệm bằng mô phỏng trên OMNeT++. Cuối cùng là các kết luận và hướng phát triển tiếp theo, được trình bày chi tiết trong phần V.

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

A. Học tăng cường

Học tăng cường là một hình thức học máy, ở đó hệ thống học từ các hành động trước đó của nó để chọn hành động tốt hơn trong tương lai. Hình 2 mô tả nguyên lý làm việc của học tăng cường, ở đó các tác nhân hoạt động như là một thực thể học tập tương tác với môi trường để lựa chọn một hành động sao cho nhận được phần thưởng tốt nhất. Tác nhân khám phá môi trường và quyết định hành động nào cần thực hiện bằng cách sử dụng quy trình quyết định Markov (MDP) [5]. MDP được biểu diễn bởi bộ 4 tham số *(S, A, p, r)*. Trong đó: S là tập các trạng thái (states), A là tập các hành động (actions), *p* là phân bố xác suất khi chuyển đổi từ trạng thái *s* sang trạng thái kế tiếp *s'* sau khi thực hiện hành động *a*, và *r* là phần thưởng (reward) nhận được tức thì khi chuyển trạng thái từ *s* sang *s'*.



Hình 2. Nguyên lý làm việc của học tăng cường

Theo Hình 2 trên, ở thời điểm *t*, tác nhân quan sát trạng thái hiện tại s_t của nó trong môi trường rồi thực hiện hành động a_t, sau đó tác nhân này nhận được phần thưởng r_t và trạng thái mới s_{t+1} từ môi trường. Mục tiêu chính của tác nhân là xác định chính sách π để tích lũy phần thưởng tối đa có thể từ môi trường. Về lâu dài, tác nhân cũng cố gắng tối đa hóa tổng phần thưởng chiết khấu dự kiến được xác định bằng $\max[\sum_{t=0}^{T} \delta . r_t(s_t, \pi(s_t))]$, với $\gamma \in [0,1]$ là hệ số chiết khấu. Sử dụng phần thưởng chiết khấu, một phương trình Bellman có tên là Q-function được xây dựng để thực hiện hành động tiếp theo a_t sử dụng MDP khi xác suất chuyển đổi trạng thái được biết trước. Q-function được biểu thị bằng công thức sau:

$Q(s_t, a_t) = (1 - \alpha) * Q(s_t, a_t) + \alpha[r + \gamma(\max Q(s_{t+1}, a_t))]$, với α là tốc độ học. γ là hệ số chiết khấu.

Học tăng cường với Q-function được gọi là Q-learning (QL). Ban đầu, thực thể học tập khám phá mọi trạng thái của môi trường thực hiện các hành động khác nhau và tạo Q-table bằng cách sử dụng QL cho từng cặp "trạng thái-hành động". Sau đó tác tử bắt đầu khai thác môi trường bằng cách lấy các hành động với Q-value tốt nhất từ Q-table. Chính sách này được gọi là chính sách ε-greedy, trong đó tác nhân bắt đầu khám phá hoặc khai thác môi trường tùy thuộc vào giá trị của xác suất ε.

B. Mô hình hóa bài toán định tuyến sử dụng học tăng cường

Trong bài toán định tuyến cho FANET nói riêng, MANET nói chung, thực thể học tập là các nút tương tác với môi trường chính là hệ thống mạng, hành động thực hiện là chọn nút láng giềng để chuyển tiếp gói tin tới đích.

- Tập các trạng thái: $S \in \{s_{u_i}(t), i = (1,2, \dots U)\}$ biểu thị cho vị trí của nút u_i ở thời điểm t, $u_i \in U$
- Tập các hành động: $A \in a_{u_{ij}}$ biểu diễn cho việc lựa chọn nút chuyển tiếp u_j từ các láng giềng của u_i
- $r_{u_{ij}}$ là giá trị thưởng mà u_i nhận được từ u_j khi u_i chọn u_j là nút chuyển tiếp. Giá trị $r_{u_{ij}}$ được tính toán thông qua một hàm thưởng được xây dựng trước để đánh giá chất lượng của hành động.
- Giá trị Q-value được cập nhật lặp đi lặp lại tại mỗi nút UAV theo phương trình sau:

$$Q\left(s_{u_i}, a_{u_{ij}}\right) \leftarrow Q\left(s_{u_i}, a_{u_{ij}}\right) + \alpha * \left[r_{u_{ij}} + \gamma * \max_{a'_{u_{ij}}} Q\left(s'_{u_i}, a'_{u_{ij}}\right) - Q\left(s_{u_i}, a_{u_{ij}}\right)\right] \quad (1)$$

$\max_{a'_{u_{ij}}} Q\left(s'_{u_i}, a'_{u_{ij}}\right)$ biểu diễn cho Q-value ước tính có được trong tương lai ở trạng thái s'_{u_i} sau khi thực hiện hành động tốt nhất $a_{u_{ij}}$. Ý nghĩa của α thể hiện mức độc học nhanh thuật toán QL và giá trị của nó xác định mức độ thông tin mới thu được ghi đè thông tin cũ và tham số này kiểm soát sự hội tụ của thủ tục học, còn hệ số γ xác định

mức độ mà thuật toán Q-Learning học được từ sai lầm của nó do vừa thực hiện hành động xấu và giá trị của nó kiểm soát tầm quan trọng của phần thưởng tương lai.

Phần tiếp theo đây chúng tôi trình bày một số nghiên cứu liên quan đến loại giao thức định tuyến sử dụng cơ chế chuyển tiếp dựa trên vị trí sử dụng học tăng cường để ra quyết định chọn đường ứng dụng cho FANET.

III. CÁC NGHIÊN CỨU ĐỊNH TUYẾN TRONG FANET SỦ DỤNG HỌC TĂNG CƯỜNG

Giao thức QGeo [6] được Jung và cộng sự đề xuất để ứng dụng cho FANET. Không như giao thức định tuyến theo vị trí, thay vì chỉ tiến hành tìm kiếm trong phạm vi truyền dẫn về phía nút đích, QGeo dựa vào vận tốc gói tin để chọn nút chuyển tiếp. Hàm thưởng được thiết kế dựa trên tham số vận tốc gói trên liên kết của cặp nút láng giềng gửi-nhận. Giá trị thưởng âm được sử dụng khi thiết kế hàm thưởng của QGeo giúp có thể tránh tối ưu cục bộ. QGeo không xem xét vấn đề tiêu thụ năng lượng, do vậy không thể đồng thời cung cấp việc đảm cải thiện đồng thời các tham số hiệu năng. QGeo sử dụng tốc độ học cố định và chưa xem xét sự cân bằng giữa chiến lược thăm dò và khai thác để khám phá các UAV chuyển tiếp tốt hơn.

Khắc phục một số yếu điểm ở QGeo, các tác giả [7] xây dựng giao thức RFL-QGeo sử dụng hàm thưởng mới cho QGeo giúp ít phải truyền lại hơn, trễ đầu cuối trung bình thấp hơn, và tỉ lệ chuyển gói cao hơn. RFL-QGeo sử dụng kỹ thuật "học tăng cường ngược" để thiết kế quyết định chọn tuyến bằng cách trao đổi gói "hello". Việc này giúp tăng tốc quá trình học tập với ít chi phí truyền thông hơn. Các gói "hello" giúp các UAV chia sẻ vị trí, tình trạng liên kết, mức năng lượng còn lại, lỗi liên kết, lỗi vị trí và Q-value với các UAV láng giềng. Việc cập nhật Q-value sử dụng hàm thưởng mới đề xuất có xem xét đến khoảng cách giữa hai UAV hướng về nút đích và giá trị thời gian truyền gói. Tuy vậy, việc bổ sung nhiều thông tin vào gói "hello" làm gia tăng kích thước gói, khi tần suất trao đổi lớn sẽ gây tốn băng thông mạng, làm giảm hiệu năng.

Giải quyết vấn đề cần tối ưu hóa đồng thời việc tiêu hao năng lượng và trễ đầu-cuối khi ra quyết định chọn tuyến, Jianmin Liu [8] đã ứng dụng học tăng cường trong xây dựng giao thức định tuyến đa mục tiêu (QMR). QMR áp dụng phương pháp điều chỉnh thức ứng các tham số học tăng cường và một cơ chế khai thác và khám phá mới. Để làm cho quá trình học tập hiệu quả và ổn định hơn, QMR thực hiện cập nhật Q-value bằng cách cập nhật một cách thích ứng tốc độ học tập và hệ số chiết khấu, sử dụng hàm mũ của độ trễ một chặng chuẩn hóa và sự thay đổi trong tập láng giềng ở hai thời điểm khác nhau một cách tương ứng. Trong quá trình khai thác, QMR giúp đưa ra quyết định tốt hơn vì Q-value được trọng số hóa theo chất lượng liên kết, với chất lượng liên kết được ước tính bằng cách sử dụng phương pháp ETX[9]. QMR xem xét chính sách thưởng nhỏ nhất khi xảy ra vòng lặp định tuyến, mức tối thiểu cục bộ và trạng thái lỗi của nút xảy ra trong quá trình lựa chọn nút chuyển tiếp. Tuy nhiên, QMR không kiểm soát tính di động và khoảng thời gian quảng bá gói "hello". Q-value được cập nhật mà không xem xét đến mức SINR của các liên kết.

Trong định tuyến học tăng cường điển hình, giá trị Q-value được cập nhật dựa vào phần thưởng nhận được từ các tập gần nhất, giao thức Q-FANET[10] khai thác thêm yếu tố ở tầng thấp bằng cách kết hợp hai mô đun QMR và Q-noise+[11] để ra quyết định chọn tuyến. Q-FANET cập nhật Q-value chính xác hơn bằng cách xem xét phần thưởng theo trọng số trên một lượng hữu hạn các tập cuối cùng và mức SINR của liên kết đã chọn. Môđun QMR được sử dụng để chọn nút chuyển tiếp theo vận tốc gói lớn nhất. Khi vấn đề hố định tuyến xảy ra, Q-FANET sử dụng cơ chế phạt của QMR và phân bổ phần thưởng nhỏ nhất cho nút chuyển tiếp đó. Trong trường hợp không gặp hố định tuyến, giao thức này thực hiện chiến lược cân bằng giữa thăm dò và khai thác sử dụng chính sách tham lam $\in$-greedy.

Giao thức QTAR [12] có thể đưa ra quyết định chọn tuyến tốt hơn bằng cách mở rộng chế độ quan sát cục bộ của mỗi UAV bằng cách sử dụng thông tin láng giềng hai chặng. QTAR đã cập nhật các Q-value bằng cách điều chỉnh thích ứng tốc độ học dựa trên hệ số chiết khấu và độ trễ hai chặng chuẩn hóa theo hàm mũ dựa trên tập tương tự láng giềng một chặng hiện tại và láng giềng trước đó của mỗi nút UAV láng giềng. Thông qua tiến trình này, QTAR tạo ra Q-value ổn định hơn cho việc thăm dò tốt hơn. Hàm thưởng đa mục tiêu sử dụng thông tin láng giềng hai chặng tối ưu hóa độ trễ và tạo ra sự cân bằng tải thích hợp trong mức tiêu hao năng lượng trong quá trình định tuyến đa chặng. Cơ chế phạt trong hàm thưởng đảm bảo tránh các vòng lặp định tuyến, lỗ hổng định tuyến và tối ưu cục bộ.

IV. ĐỀ XUẤT GIAO THỨC VAQR

Phần này chúng tôi trình bày đề xuất của mình trong thiết kế và cài đặt một giao thức định tuyến cho FANET sử dụng học tăng cường với hàm thưởng 2 biến: vận tốc tương đối, góc.

A. Ý tưởng thuật toán

Đặc điểm của các nút UAV trong FANET là mức độ di động cao, ở thời điểm hiện tại chúng có thể là láng giềng của nhau nhưng nhanh chóng chúng có thể rời xa khỏi phạm vi truyền thông của nhau, gây đứt liên kết nếu đang truyền dữ liệu. Mức độ "nhanh chóng rời xa khỏi nhau" hay mức độ "cùng chiều" có thể được đánh giá qua vector vận tốc tương đối giữa 2 nút, liên kết giữa 2 nút sẽ bền vững hơn nếu vận tốc tương đối giữa chúng nhỏ hơn. Như vậy trong học tăng cường, nếu một nút thực hiện hành động chuyển gói cho một nút láng giềng có mức độ "cùng chiều" cao sẽ nhận được phần thưởng cao.

Tuy nhiên, thực hiện chiến lược trên mới đảm bảo khả năng nâng cao được việc chọn nút chuyển tiếp có độ bền liên kết, khai thác ý tưởng chuyển tiếp tham lam của GPSR[13], chúng tôi sử dụng thêm độ đo mức độ gần đích nhằm

cải thiện tham số hiệu năng trễ truyền đầu-cuối. Nếu chuyển gói cho nút gần đích hơn thì sẽ nhận được mức thưởng cao hơn. Tham số góc được chúng tôi sử dụng để đánh giá mức độ gần đích thay vì khoảng cách euclid.

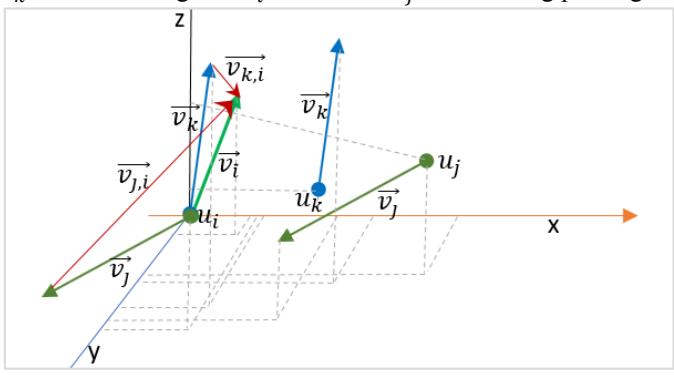
Việc học của một nút sẽ được thực hiện qua việc nhận được phần thưởng kết hợp hai độ đo (vận tốc tương đối và góc) từ nút được chọn chuyển tiếp gói dữ liệu theo một trọng số nhất định.

B. Thiết kế hàm thưởng

1. Vận tốc tương đối

Yếu tố vận tốc tương đối được xét đến nhằm giúp đánh giá độ ổn định của liên kết. Như minh họa ở Hình 3, u_j và u_k là láng giềng của u_i, các vector vận tốc tương ứng là $\overrightarrow{v_j}$, $\overrightarrow{v_k}$ và $\overrightarrow{v_i}$, vận tốc tương đối của các láng giềng u_j và u_k với u_i lần lượt là $\overrightarrow{v_{j,i}}$ và $\overrightarrow{v_{k,i}}$. Với $\overrightarrow{v_{j,i}} = \overrightarrow{v_i} - \overrightarrow{v_j}$ và $\overrightarrow{v_{k,i}} = \overrightarrow{v_i} - \overrightarrow{v_k}$.

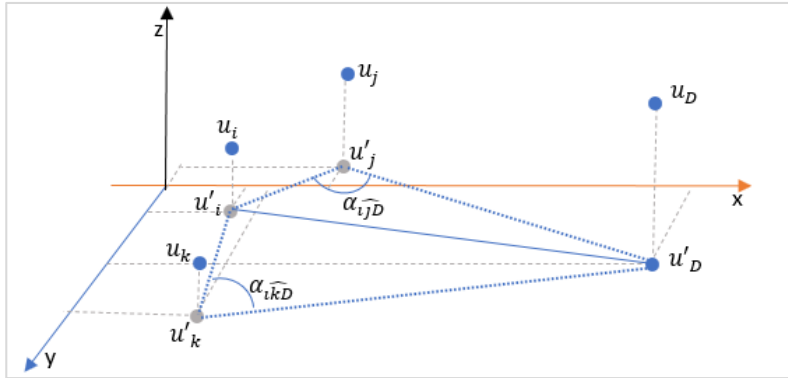
Vận tốc tương đối càng nhỏ cho thấy chúng càng chuyển động "cùng phương" và do vậy khả năng đứt liên kết càng ít. Như tại Hình 3, nút u_k nên trả thưởng cho u_i nhiều hơn u_j vì mức cùng phương của nó với nút gửi cao hơn.



Hình 3. Minh họa vector vận tốc và vector vận tương đối giữa các láng giềng u_j và u_k với u_i

2. Góc

Một nút láng giềng u_j hoặc u_k khi nhận được gói từ u_i sẽ đánh giá mức độ xa hay gần đích của nó để trả về cho u_i một mức thưởng nhất định. Ký hiệu u'_j , u'_k , u'_i, u'_D tương ứng là hình chiếu của u_j, u_k, u_i, u_D (u_D là nút đích) lên mặt đất (mặt phẳng Oxy), khi đó độ lớn của các góc $\alpha_{i\widehat{jD}}$, $\alpha_{i\widehat{kD}}$ sẽ cho biết mức độ gần đích của các nút u_j, u_k tương ứng. Minh họa tại Hình 4 như sau:



Hình 4. Minh họa góc tạo bởi hình chiếu các láng giềng u_j , u_k với u_i và u_D

Như trình bày trên Hình 4, nút u_j nên thưởng cho u_i nhiều hơn u_k vì rằng $\alpha_{i\widehat{jD}} > \alpha_{i\widehat{kD}}$.

3. Hàm thưởng đề xuất

Khai thác ý tưởng thiết kế hàm thưởng từ Q-FANET, với những phân tích trên, chúng tôi đề xuất tính toán giá trị thưởng nhờ vào kết hợp đánh giá góc và vận tốc tương đối để thưởng dựa trên độ ổn định liên kết và mức độ gần đích. Như vậy giá trị thưởng mà u_i nhận được từ u_j sau khi thực hiện hành động chọn nút chuyển tiếp là u_j được tính như hàm thưởng sau:

$$reward(u_i, u_j) = \begin{cases} 100, & nếu\ u_{j+1}\ là\ nút\ đích \\ -100, & nếu\ u_j\ là\ cực\ đại\ địa\ phương \\ \omega * (1 - Rv(u_i, u_j)) + (1 - \omega) * Ag(u_i, u_j, u_D) & cho\ trường\ hợp\ còn\ lại \end{cases} \quad (2)$$

Với $Rv(u_i, u_j) = |\overrightarrow{v_{j,i}}|/2 * v_{MAX}$ là vận tốc tương đối

và $Ag(u_i, u_j, u_D) = \alpha_{i\widehat{j}D}/180$ là góc được chuẩn hóa về miền giá trị [0, 1]

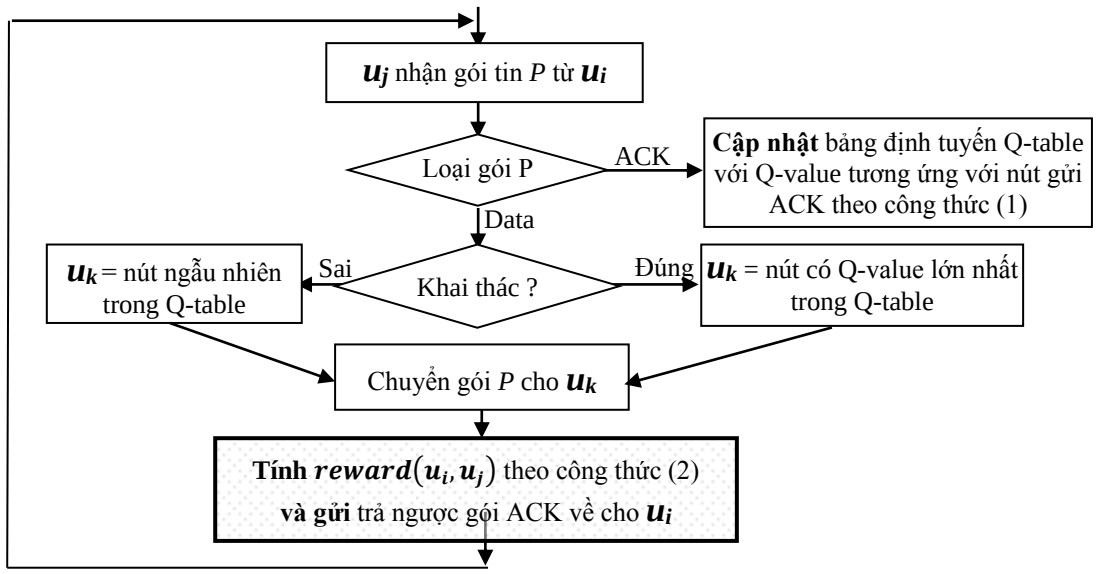
Theo công thức tính thưởng trên, giá trị thưởng cho u_i là cực đại (100) nếu u_j là láng giềng của nút đích. Trường hợp u_j không tìm được một láng giềng nào khác u_i để chuyển gói (đây là trường hợp gặp hố định tuyến, tức u_j là cực đại địa phương) thì giá trị thưởng (-100) ở đây được chọn mang tính "phạt" u_i cho hành động đã chọn nhầm u_j làm nút chuyển tiếp. Cho trường hợp còn lại, tham số ω được sử dụng để định mức độ tham gia của 2 tham số là vận tốc tương đối và góc trong tính toán giá trị thưởng.

$|\overrightarrow{v_{j,i}}|$ cho biết độ lớn vận tốc tương đối của nút UAV u_j với u_i, v_{MAX} là vận tốc di chuyển cực đại của các UAV trong quá trình mô phỏng. Do $Rv(u_i, u_j)$ nhỏ thì mức thưởng lớn hơn và vì vậy (1- $Rv(u_i, u_j)$) càng lớn thì mức thưởng càng lớn theo.

$\alpha_{i\widehat{j}D}$ cho biết góc rộng tại là u_j được tạo bởi là u_i, là u_j và là $u_{Đích}$. $\alpha_{i\widehat{j}D}$ cũng được chuẩn hóa thành $Ag(u_i, u_j)$ nhằm cùng kết hợp với vận tốc tương đối trong hàm thưởng.

Việc đưa về miền giá trị [0, 1] giúp kiểm soát giá trị thưởng trong phạm vi nhỏ và không vượt quá giá trị thưởng tối đã khi gói tin tới được đích.

4. Lưu đồ thuật toán xử lý gói tin và học từ gói ACK tại nút u_j



Hình 5. Lưu đồ khối thuật toán xử lý gói tin và học từ gói phản hồi ACK tại nút u_j

Sơ đồ khối Hình 5 là sơ đồ chung cơ bản cho định tuyến sử dụng học máy tăng cường học từ gói phản hồi [14]. Đóng góp của chúng tôi ở bài báo này tập trung ở thiết kế hàm thưởng (như đã đề xuất ở công thức 2). Cấu trúc gói ACK nhỏ và đơn giản, là gói UDP với payload là giá trị thưởng tính được qua hàm thưởng. Bảng định tuyến Q-table với mỗi bản ghi là bộ ba <nút đích, chặng kế tiếp, Q-value>.

C. Đánh giá kết quả bằng mô phỏng

Ở phần này, chúng tôi trình bày thiết kế kịch bản mô phỏng và phân tích kết quả mô phỏng của giao thức đề xuất VAQR (relative Velocity and Angle aware Q-Routing). Các tham số hiệu năng được so sánh với GPSR [13] - một giao thức định tuyến theo vị trí điển hình cho MANET sử dụng hai chế độ chuyển tiếp: chuyển tiếp tham lam (chọn nút láng giềng gần đích nhất làm nút chuyển tiếp) và chuyển tiếp theo chu vi khi chuyển tiếp tham lam thất bại (khi không tồn tại láng giềng nào gần đích hơn nó).

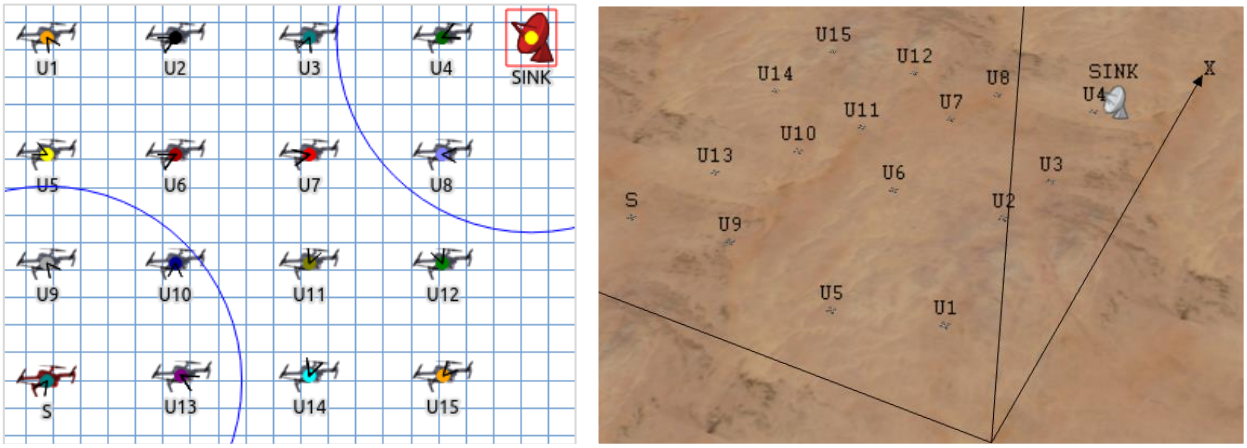
1. Thông số và kịch bản mô phỏng

Ở các nghiên cứu liên quan chỉ ra ở trên, kịch bản mô phỏng chỉ thực hiện trên diện tích 500 m^2, không đề cập đến cao độ, với lượng nút 25 UAV và phạm vi giao tiếp 180 m thì mật độ tương đối dày, tốc độ di chuyển thấp 0-15 m/s, mô hình di động RandomWayPoint hoặc Gaussian Markov.

Nhắm tới mục đích ứng dụng FANET cho hoạt động tìm kiếm cứu nạn [15], với sự hỗ trợ của INET/OMNet [16], chúng tôi thực hiện mô phỏng trên không gian địa lý rộng hơn với diện tích 5 km^2 và cao độ 800 m. Các UAV di chuyển theo mô hình di động MassMobility [17], các nút di chuyển thẳng theo hướng ngẫu nhiên và thực hiện thay đổi sau mỗi thời gian 5 giây, góc thay đổi theo mặt phẳng nằm và cao độ thay đổi ngẫu nhiên nhiêu trong khoảng -30 độ đến 30 độ.

Nút gửi S và nút đích SINK. Ở đây SINK đóng vai trò như là một trạm mặt đất. Các UAV ban đầu được phân bố đều trên lưới cách nhau gần 1 km, ăng-ten đẳng hướng, kênh truyền Ieee80211ScalarRadio với mô hình lan truyền không dây là mô hình không gian tự do.

Do đặc tính di động của mô hình chuyển động các nút UAV nên mỗi chúng tôi thực việc lặp lại mô phỏng 5 lần ở các tốc độ di động 30 m/s, 25 m/s, 20 m/s, 15 m/s, 10 m/s, 5 m/s để đánh giá hiệu quả.



Hình 6. Phân bố các UAV nhìn theo phương trục Z (trái) và phối cảnh 3 chiều (phải)

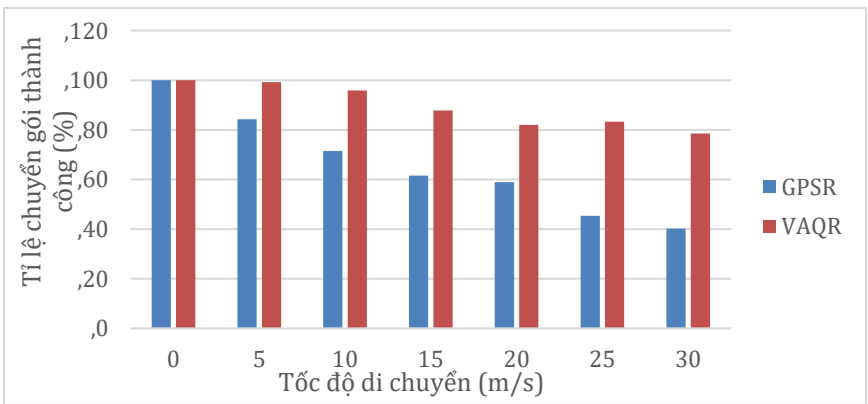
Bảng 1. Giá trị một số tham số trong mô phỏng

Thông số	**Giá trị**
Khu vực địa lý	5 km x 5 km x 800 m (cao độ)
Tổng số nút UAV	16
Công suất phát	100 mW
Tốc độ di chuyển	0, 5, 10, 15, 20, 25, 30 m/s
Mô hình chuyển động	Mass mobility
Nguồn phát + thu	1 + 1
Kích thước gói	512 bytes
Bitrate	2 Mbps
Thời gian mô phỏng	220 s
Thời điểm bắt đầu truyền	20 s
Hello interval	5 s
Tốc độ học	0,7
Hệ số chiết khấu	0,2
Trọng số vận tốc tương đối: ω	0,4
Trọng số góc: 1-ω	0,6

Theo Bảng 1 trên, chúng tôi dành 20 s mô phỏng đầu tiên cho các nút trao đổi thông điệp "hello" để xây dựng bảng láng giềng trước khi thực hiện truyền dữ liệu. Tốc độ học và hệ số chiết khấu tham khảo từ các bài nghiên cứu liên quan, trọng số vận tốc tương đối và góc tham gia ở hàm thưởng được chúng tôi chọn từ thực nghiệm.

2. Kết quả mô phỏng và đánh giá

a) Tỉ lệ chuyển gói thành công

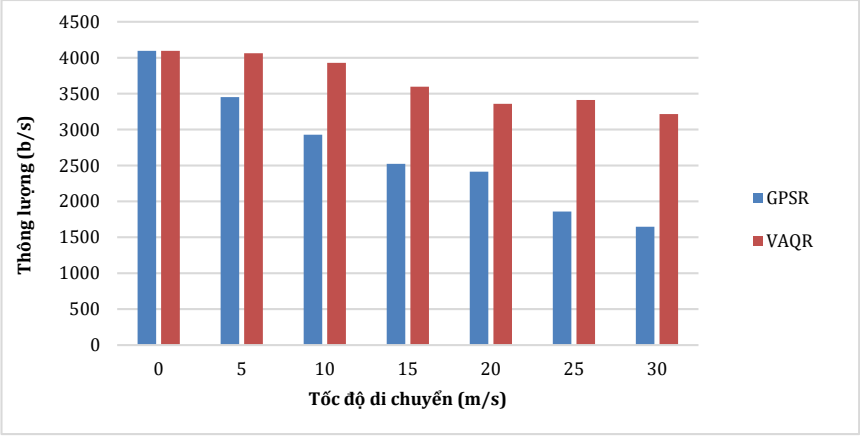


Hình 7. Tỉ lệ chuyển gói thành công theo tốc độ di chuyển

Như kết quả trình bày trên Hình 7, giao thức định tuyến đề xuất (VAQR) luôn đạt được mức chuyển gói thành công cao hơn đáng kể so với giao thức GPSR. Khi UAV di chuyển ở tốc độ cao 25 m/s và 30 m/s GPSR thể hiện rõ yếu điểm khi chỉ đạt mức chuyển gói ở mức 43 % và 40 %, trong khi đó VAQR đạt được mức độ chuyển gói gấp đôi.

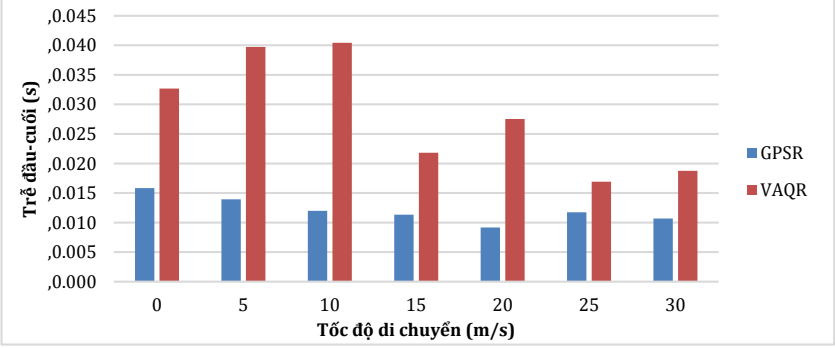
b) Thông lượng

Tương ứng với mức chuyển gói thành công, Hình 8 cũng cho thấy về mặt thông lượng VAQR cũng đạt được mức thông lượng tốt hơn so với GPSR ở hầu hết các tốc độ di chuyển của các UAV, đặc biệt khi di chuyển ở tốc độ cao 25 m/s và 30 m/s.



Hình 8. Mức thông lượng theo tốc độ di chuyển

c) Trễ đầu-cuối



Hình 9. Trễ đầu-cuối theo tốc độ di chuyển

Trên Hình 9 cho thấy, VAQR có trễ đầu-cuối nhiều hơn khi so sánh với GPSR. Tuy vậy mức trễ đã giảm dần ở các kịch bản có UAV chuyển động theo tốc độ cao hơn. Như trên hình ta thấy giao thức đề xuất gây trễ đáng kể ở các kịch bản UAV chuyển động với tốc độ nhỏ (0 m/s, 5 m/s, 10 m/s) so với GPSR, nhưng ở tốc độ 25 m/s, 30 m/s mức trễ đã được cải thiện giảm đáng kể.

V. KẾT LUẬN

Qua việc nghiên cứu về định tuyến trong FANET sử dụng học tăng cường, sử dụng hệ mô phỏng OMNET++ chúng tôi đã thực hiện cài đặt và đánh giá được hiệu quả giao thức đề xuất VAQR so với giao thức định tuyến theo vị trí GPSR. Kết quả mô phỏng cho thấy VAQR đã đạt được hiệu năng định tuyến cao ở chỉ số về tỉ lệ chuyển gói thành công và thông lượng. Ở chỉ số trễ đầu-cuối VAQR gây trễ hơn so với GPSR, điều này do thiết kế hàm thưởng yêu cầu nhiều xử lý hơn tại mỗi nút, cũng như vấn đề giao thức thực hiện cả phải pháp thăm dò thay vì chỉ khai thác. Thiết kế hàm thưởng tốt hơn, sử dụng phù hợp tỉ lệ thăm dò và khai thác, thay đổi thích ứng hệ số chiết khấu, sử dụng mô hình di động phù hợp cho kịch bản FANET ứng dụng trong tìm kiếm cứu nạn là định hướng mà chúng tôi sẽ tiếp tục nghiên cứu trong thời gian tới.

TÀI LIỆU THAM KHẢO

[1] S. Hayat, E. Yanmaz, and R. Muzaffar, "Survey on Unmanned Aerial Vehicle Networks for Civil Applications: A Communications Viewpoint," *IEEE Commun. Surv. Tutorials*, Vol. 18, No. 4, pp. 2624-2661, 2016, doi: 10.1109/COMST.2016.2560343.

[2] A. Rovira-Sugranes, A. Razi, F. Afghah, and J. Chakareski, "A review of AI-enabled routing protocols for UAV networks: Trends, challenges, and future outlook," *Ad Hoc Networks*, Vol. 130, No. 2008784, pp. 102790, 2022, doi: 10.1016/j.adhoc.2022.102790.

[3] A. Mukherjee, N. Dey, S. Satapathy, and E. Engineers, "Flying Ad-hoc Networks : A Comprehensive Survey Flying Ad-hoc Networks : A Comprehensive Survey," November, 2016.

[4] M. Y. Arafat and S. Moh, "A survey: Routing protocols for UAV networks," *IEEE Access*, Vol. 7, pp. 99694-99720, 2019, doi: 10.1109/ACCESS.2019.2930813.

[5] M. Ponsen, M. Taylor, and K. Tuyls, "Abstraction and Generalization in Reinforcement Learning: A Summary and Framework," in *Adaptive and Learning Agents*, 2010.

[6] W. S. Jung, J. Yim, and Y. B. Ko, "QGeo: Q-Learning-Based Geographic Ad Hoc Routing Protocol for Unmanned Robotic Networks," *IEEE Commun. Lett.*, Vol. 21, No. 10, pp. 2258-2261, 2017, doi: 10.1109/LCOMM.2017.2656879.

[7] W. Jin, R. Gu, and Y. Ji, "Reward Function Learning for Q-learning-Based Geographic Routing Protocol," *IEEE Commun. Lett.*, Vol. 23, No. 7, pp. 1236-1239, 2019, doi: 10.1109/LCOMM.2019.2913360.

[8] J. Liu *et al.*, "QMR:Q-learning based Multi-objective optimization Routing protocol for Flying Ad Hoc Networks," *Comput. Commun.*, Vol. 150, pp. 304-316, 2020, doi: 10.1016/j.comcom.2019.11.011.

[9] S. Rosati, K. Kruzelecki, L. Traynard, and B. Rimoldi, "Speed-aware routing for UAV ad-hoc networks," *2013 IEEE Globecom Work. GC Wkshps 2013*, pp. 1367-1373, 2013, doi: 10.1109/GLOCOMW.2013.6825185.

[10] L. A. L. F. da Costa, R. Kunst, and E. Pignaton de Freitas, "Q-FANET: Improved Q-learning based routing protocol for FANETs," *Comput. Networks*, Vol. 198, No. September 2020, p. 108379, 2021, doi: 10.1016/j.comnet.2021.108379.

[11] L. R. Faganello, R. Kunst, C. B. Both, L. Z. Granville, and J. Rochol, "Improving reinforcement learning algorithms for dynamic spectrum allocation in cognitive sensor networks," *IEEE Wirel. Commun. Netw. Conf. WCNC*, pp. 35-40, 2013, doi: 10.1109/WCNC.2013.6554535.

[12] M. Y. Arafat and S. Moh, "A Q-Learning-Based Topology-Aware Routing Protocol for Flying Ad Hoc Networks," *IEEE Internet Things J.*, Vol. 9, No. 3, pp. 1985-2000, 2022, doi: 10.1109/JIOT.2021.3089759.

[13] B. Karp and H. T. Kung, "GPSR: Greedy Perimeter Stateless Routing for wireless networks," *Proc. Annu. Int. Conf. Mob. Comput. Networking, MOBICOM*, pp. 243-254, 2000.

[14] M. M. Alam and S. Moh, "Q-Learning-Based Routing in Flying Ad Hoc Networks: A Survey," *Proc. 10th Int. Conf. Smart Media Appl. (SMA 2021)*, September, 2021.

[15] C. Bouras, A. Gkamas, and S. A. K. Salgado, "Energy Efficient Mechanism over LoRa for Search and Rescue operations," *2021 Int. Symp. Networks, Comput. Commun. ISNCC 2021*, 2021, doi: 10.1109/ISNCC52172.2021.9615822.

[16] L. Mészáros, A. Varga, and M. Kirsche, "INET Framework 4," in *Recent Advances in Network Simulation. EAI/Springer Innovations in Communication and Computing*, Springer, Cham, 2019, pp. 55-106.

[17] Omnet, "MassMobility Model." https://doc.omnetpp.org/inet/api-current/neddoc/inet.mobility.single.MassMobility.html.

VAQR: A REINFORCEMENT LEARNING APPROACH FOR ROUTING IN FANET

Mai Cuong Tho, Nguyen Thi Huong Ly, Le Huu Binh, Vo Thanh Tu

ABSTRACT: *This paper presents research results on the application of reinforcement learning for routing protocols in FANET. A position-based routing algorithm (VAQR) is proposed with the idea of constructing a reward function using two metric of relative velocity and angle. The routing table in each node is updated using the Q-learning algorithm with the proposed reward function. Simulation results on OMNet++ show that the proposed algorithm achieves high performance in terms of successful packet transfer rate and throughput compared to GPSR protocol, with the average transmission delay is higher than that of the GPSR protocol within an acceptable range.*