Tạp chí Khoa học và Công nghệ Trường Đại học Công nghệ Đông Á

Phân tích sự đánh đổi giữa hiệu năng và năng lượng trong mạng truyền thông tán xạ ngược môi trường chống nhiễu sử dụng học tăng cường đa tác tử

Nguyễn Thái Dương[1], Nguyễn Ngọc Tân[1*]

[1] Khoa Công nghệ thông tin, Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội

*Email: tan.nguyen@vnu.edu.vn

Tóm tắt

Sự bùng nổ của các thiết bị Internet vạn vật (IoT) trong kỷ nguyên 6G đặt ra thách thức kép về tiết kiệm năng lượng và khả năng chống chịu trước các cuộc tấn công gây nhiễu chủ động. Kỹ thuật Truyền thông Tán xạ ngược Môi trường (Ambient Backscatter Communication - AmB) nổi lên như một giải pháp hứa hẹn nhờ khả năng tận dụng sóng vô tuyến có sẵn để truyền tin với công suất cực thấp. Tuy nhiên, việc tối ưu hóa đồng thời độ tin cậy và hiệu quả năng lượng trong môi trường nhiễu là bài toán phức tạp. Bài báo này nghiên cứu bài toán cân bằng tài nguyên động, trong đó tín hiệu gây nhiễu được thu thập để cung cấp năng lượng vận hành hệ thống. Chúng tôi đề xuất khung giải pháp IA-MADDPG, kết hợp Học tăng cường sâu đa tác tử với cơ chế Học bắt chước nhằm khắc phục vấn đề hội tụ chậm trong không gian hành động liên tục. Kết quả mô phỏng cho thấy, thay vì tối đa hóa thông lượng một cách mù quáng, phương pháp đề xuất chấp nhận mức tiêu thụ năng lượng cao hơn để duy trì sự ổn định của hiệu suất thông lượng trên năng lượng trong các điều kiện khắc nghiệt.

Từ khóa: *Truyền thông tán xạ ngược môi trường, Học tăng cường sâu đa tác tử, Chống nhiễu, Hiệu quả năng lượng, mạng truyền trực tiếp.*

Abstract

The deployment of Internet of Things (IoT) devices in 6G networks introduces dual challenges regarding energy efficiency and resilience against active jamming attacks. Ambient Backscatter Communication (AmB) provides a mechanism to address these constraints by modulating existing radio frequency (RF) waves for ultra-low-power transmission. However, the joint optimization of transmission reliability and energy efficiency in non-stationary jamming environments constitutes a complex resource allocation problem. This paper investigates a dynamic resource balancing model wherein adversarial jamming signals are harvested to supply operational energy for the AmB system. We formulate the joint anti-jamming and energy allocation problem as a partially observable Markov decision process (POMDP). To solve this, we propose an Imitation-Augmented Multi-Agent Deep Deterministic Policy Gradient (IA-MADDPG) framework. This architecture integrates multi-agent deep reinforcement learning (MARL) with an imitation learning mechanism to mitigate the slow convergence and high exploration overhead inherent to continuous action spaces. Simulation results indicate that the IA-MADDPG framework strictly avoids unconstrained throughput maximization; instead, it strategically allocates higher energy consumption to guarantee the stability of the throughput-to-energy efficiency metric under severe jamming conditions.

Keyword: *Ambient Backscatter Communication, Multi-agent deep reinforcement learning, Anti-Jamming, Energy efficiency, Direct transmission network.*

1. MỞ ĐẦU

Trong lộ trình phát triển của mạng di động thế hệ thứ 6 (6G), việc kết nối hàng tỷ thiết bị IoT công suất thấp đòi hỏi các giải pháp truyền thông mang tính đột phá về hiệu quả năng lượng. Mặc dù các công nghệ hiện tại đã đạt được những bước tiến lớn về tốc độ, rào cản về tuổi thọ pin và sự khan hiếm phổ tần vẫn là những nút thắt cổ chai [1]. Đặc biệt, trong các kịch bản quân sự hoặc an ninh cao, các thiết bị IoT thường xuyên phải đối mặt với các cuộc tấn công gây nhiễu nhằm làm gián đoạn kết nối.

Các phương pháp chống nhiễu truyền thống như Trải phổ nhảy tần (FHSS) hay Trải phổ chuỗi trực tiếp (DSSS) thường yêu cầu các thiết bị phải tiêu tốn một lượng năng lượng lớn để phát tín hiệu mạnh hơn nhiễu hoặc liên tục thay đổi tần số công tác [2]. Điều này đi ngược lại với triết lý thiết kế sạch và bền vững của mạng IoT mật độ cao. Gần đây, kỹ thuật Truyền thông tán xạ ngược môi trường đã được đề xuất như một hướng đi mới, cho phép các thiết bị giao tiếp bằng cách điều biến và phản xạ lại các sóng vô tuyến có sẵn trong môi trường (như sóng TV, Wi-Fi, 5G) thay vì tự tạo ra sóng mang [3].

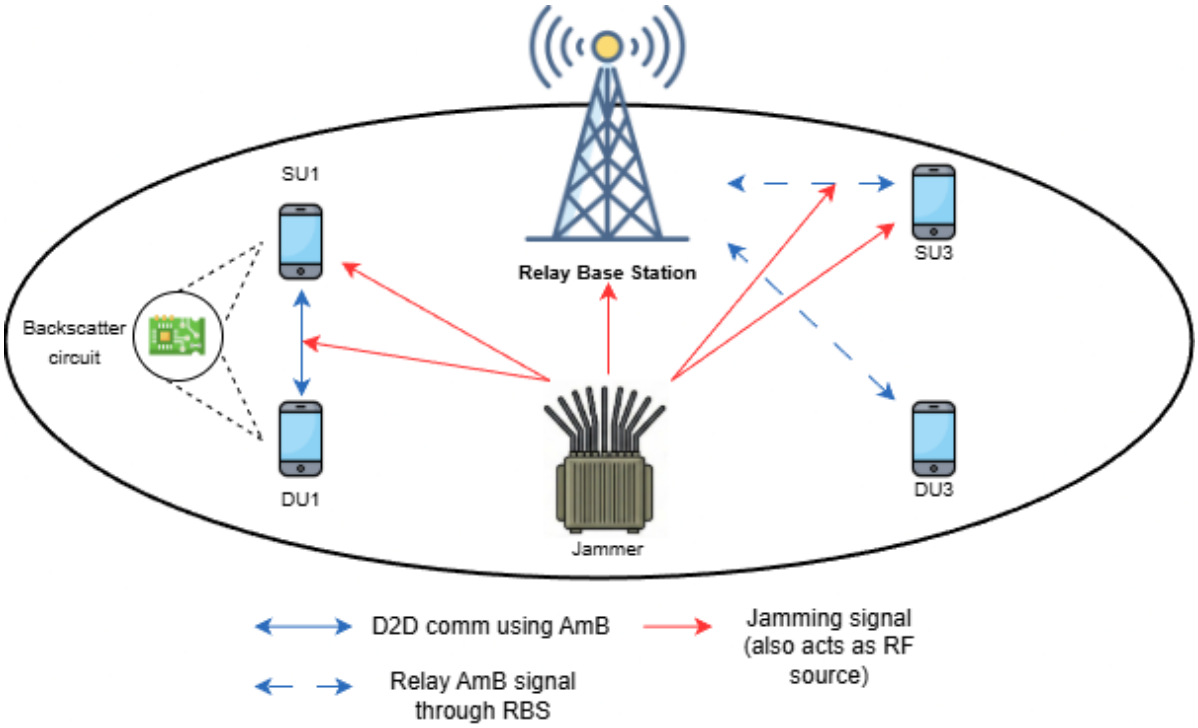
Một hướng tiếp cận đầy tiềm năng nhưng chưa được khai thác triệt để là việc coi tín hiệu gây nhiễu không phải là nhiễu, mà là một nguồn năng lượng sóng mang công suất lớn [4]. Trong kịch bản này, thiết bị Truyền thông tán xạ ngược môi trường có thể tận dụng chính năng lượng từ kẻ tấn công để phản xạ tín hiệu của mình, biến nguy cơ thành cơ hội. Tuy nhiên, việc hiện thực hóa ý tưởng này gặp phải thách thức lớn trong việc quản lý tài nguyên: Điều chỉnh hệ số phản xạ (α) để cân bằng giữa cường độ tín hiệu phát đi và năng lượng tiêu hao cho mạch điện và Lựa chọn chế độ truyền trực tiếp hoặc truyền qua trung gian để tối ưu hóa năng lượng toàn trình [5].

Các phương pháp tối ưu hóa truyền thống thường dựa trên các mô hình tĩnh hoặc tham lam, vốn không thể nắm bắt được tính động của môi trường nhiễu và sự tương tác phức tạp giữa các người dùng [6]. Học tăng cường sâu (Deep Reinforcement Learning - DRL), đặc biệt là mô hình đa tác tử (Multi-Agent DRL), cung cấp một công cụ mạnh mẽ để giải quyết các bài toán ra quyết định trong môi trường phi tập trung. Tuy nhiên, các thuật toán nền tảng như MADDPG thường gặp vấn đề "khởi đầu lạnh", dẫn đến tốn kém thời gian và năng lượng cho quá trình thăm dò [7] [8].

Bài báo này đề xuất và phân tích một khung giải pháp dựa trên thuật toán IA-MADDPG (Imitation-Augmented MADDPG). Đóng góp chính của bài báo tập trung vào việc **phân tích sâu sắc khía cạnh hiệu quả năng lượng**, làm rõ cơ chế mà tại đó các tác tử học máy chấp nhận hy sinh một phần năng lượng tiêu thụ để đổi lấy sự đảm bảo về chất lượng dịch vụ (QoS) và thông lượng trong điều kiện nhiễu khắc nghiệt.

2. MÔ HÌNH HỆ THỐNG VÀ BÀI TOÁN TỐI ƯU

2.1. Mô hình Kịch bản Mạng



Hình 1. Mô hình Kịch bản Mạng

Chúng tôi xem xét một mạng truyền thông không dây trong vùng diện tích 120×120 mét, bao gồm N cặp người dùng nguồn - đích (Source User - SU, Destination User - DU) và một trạm chuyển tiếp (Relay Base Station - rBS) đặt tại trung tâm. Hệ thống chịu tác động bởi một tác nhân gây nhiễu công suất lớn di chuyển ngẫu nhiên và phát nhiễu liên tục.

Khác với các hệ thống vô tuyến chủ động, mỗi SU_i trong mô hình này được trang bị thẻ Backscatter thụ động. Thay vì sử dụng bộ dao động nội để tạo sóng mang, SU_i thu nhận tín hiệu từ tác nhân gây nhiễu và phản xạ lại nó tới DU_i hoặc rBS.

Quá trình truyền tin có thể diễn ra theo hai chế độ (m_i):

- **Chế độ truyền trực tiếp:** Truyền trực tiếp từ SU_i đến DU_i thông qua phản xạ tín hiệu tác nhân gây nhiễu. Chế độ này tiết kiệm năng lượng truyền dẫn nhưng dễ bị ảnh hưởng nếu khoảng cách xa hoặc nhiễu tại máy thu quá lớn.
- **Chế độ truyền qua trung gian:** Tín hiệu được phản xạ từ SU_i đến rBS, sau đó rBS giải mã và chuyển tiếp đến DU_i. Chế độ này tăng độ tin cậy nhờ lợi ích phân tập không gian nhưng có thể gây độ trễ và tiêu tốn tài nguyên hệ thống.

2.2. Mô hình Tín hiệu và Tiêu thụ Năng lượng

Đây là phần trọng tâm để phân tích hiệu quả năng lượng. Quá trình backscatter được thực hiện bằng cách thay đổi trở kháng anten để hấp thụ hoặc phản xạ tín hiệu tới. Gọi $\alpha_i \in [0,1]$ là hệ số phản xạ của SU_i.

Tín hiệu thu được tại đích đến D_i trong chế độ truyền trực tiếp được biểu diễn:

$$y_{D_i}(t) = \sqrt{P_J} g_{J,S_i} g_{S_i,D_i} \alpha_i x_{S_i}(t) + n_{D_i}(t), \quad (1)$$

Trong đó, P_J là công suất phát của tác nhân gây nhiễu, $g_{A,B}$ là độ lợi kênh giữa nút A và B, và n_{D_i} là tạp âm nhiệt.

Mặc dù Truyền thông tán xạ ngược môi trường không tiêu tốn năng lượng để tạo sóng mang, nhưng nó vẫn tiêu thụ năng lượng cho việc chuyển mạch trở kháng và hoạt động của vi mạch điều khiển. Tổng năng lượng tiêu thụ E_i cho một lần truyền tin của người dùng i được mô hình hóa như sau:

$$E_i = E_{\text{circuit}} + \eta |\alpha_i|^2, \quad (2)$$

Trong đó:

- $E_{\text{circuit}} = c_{\text{switch}} + c_{\text{sense}}$ là năng lượng cố định cho việc chuyển mạch và cảm biến môi trường.
- $\eta |\alpha_i|^2$ là thành phần năng lượng động, phụ thuộc bình phương vào hệ số phản xạ α_i.

Từ phương trình trên, việc tăng α_i sẽ làm tăng cường độ tín hiệu phản xạ (tăng SINR), nhưng đồng thời làm tăng chi phí năng lượng theo hàm bậc hai. Đây chính là trọng tâm của sự đánh đổi (trade-off) mà bài báo này phân tích.

2.3. Thiết lập Bài toán Tối ưu Đa mục tiêu

Mục tiêu của hệ thống là tối đa hóa thông lượng đồng thời kiểm soát mức tiêu thụ năng lượng. Tỷ số tín hiệu trên nhiễu cộng tạp âm (SINR) tại máy thu phụ thuộc vào chế độ truyền dẫn:

- SINR chế độ truyền trực tiếp ($SINR_i^{D2D}$) tỉ lệ thuận với $\alpha_i^2 P_J$.
- SINR chế độ truyền qua trung gian ($SINR_i^{Relay}$) là giá trị nhỏ nhất của SINR hai chặng (SU-rBS và rBS-DU).

Bài toán được mô hình hóa dưới dạng Quá trình Quyết định Markov Quan sát Một phần (POMDP), trong đó mỗi tác tử SU_i quan sát trạng thái kênh cục bộ và đưa ra hành động $a_i = [\alpha_i, m_i]$.

Hàm phần thưởng (Reward Function) r_i cho mỗi tác tử được thiết kế để cân bằng giữa thông lượng và năng lượng tiêu thụ:

$$r_i = w_1 \log_2(1 + \text{SINR}_i) + w_2 \tanh(\text{SINR}_i - \Gamma) - w_3 \alpha_i^2, \quad (3)$$

Trong đó, thành phần thứ nhất khuyến khích tăng thông lượng, thành phần thứ hai đảm bảo SINR vượt ngưỡng an toàn Γ, và thành phần thứ ba ($-w_3 \alpha_i^2$) phạt nặng cho việc tiêu tốn năng lượng quá mức. Việc điều chỉnh các trọng số w sẽ quyết định chiến lược của mạng: thiên về hiệu năng (Performance-centric) hay thiên về tiết kiệm năng lượng (Energy-centric).

3. THUẬT TOÁN ĐỀ XUẤT: IA-MADDPG

Để giải quyết bài toán tối ưu đa mục tiêu trong môi trường nhiễu động như đã mô tả, chúng tôi đề xuất khung giải pháp IA-MADDPG. Khung giải pháp này được xây dựng dựa trên nền tảng của thuật toán MADDPG [7], nhưng tích hợp các cơ chế tiên tiến để khắc phục nhược điểm hội tụ chậm và kém ổn định ("khởi đầu lạnh") của các phương pháp Học tăng cường (RL) truyền thống.

3.1. Học Bắt chước (Imitation Learning) để Tăng tốc Hội tụ

Trong giai đoạn đầu của quá trình huấn luyện, việc để các tác tử khám phá ngẫu nhiên không gian hành động liên tục là cực kỳ lãng phí năng lượng và thời gian. Để giải quyết vấn đề này, chúng tôi giới thiệu một chính sách chuyên gia phân tích $\mu_E(o_i)$. Chuyên gia này sử dụng chiến lược tham lam (Greedy) dựa trên thông tin kênh truyền tức thời để tối ưu hóa SINR [8] .

Cơ chế thu thập dữ liệu từ chiến lược tham lam: Cụ thể, tại mỗi bước thời gian của giai đoạn khởi động, chính sách chuyên gia quan sát trạng thái kênh cục bộ o_i (bao gồm $h_{J,SU}$, $h_{SU,DU}$, và mức nhiễu) để tính toán cấu hình hành động $a_i = \{\alpha_i, m_i\}$ nhằm cực đại hóa giá trị hàm phần thưởng r_i tại phương trình (3). Quá trình ra quyết định này mang tính thiển cận, chỉ tập trung vào tối ưu hóa lợi ích tức thời mà không cập nhật hàm giá trị dài hạn. Các bộ dữ liệu chuyển đổi trạng thái (s, a, r, s') sinh ra từ chiến lược này được thu thập và đẩy liên tục vào bộ đệm kinh nghiệm D cho đến khi đạt ngưỡng dung lượng.

Quá trình học được tăng tốc thông qua hai cơ chế:

- **Khởi động chuyên gia (Expert Warm-up):** Trước khi bắt đầu quá trình huấn luyện RL, bộ đệm kinh nghiệm $\mathcal{D}$ được lấp đầy một phần bởi các mẫu chuyển đổi trạng thái (s, a, r, s') sinh ra từ chiến lược tham lam (Greedy). Điều này đảm bảo rằng ngay từ các bước cập nhật gradient đầu tiên, mạng nơ-ron đã được học từ các dữ liệu chất lượng cao thay vì dữ liệu ngẫu nhiên .
- **Sao chép hành vi (Behavior Cloning - BC):** Hàm mất mát của mạng Actor được sửa đổi để bao gồm thành phần sai số bắt chước. Điều này khuyến khích tác tử mô phỏng hành vi của chuyên gia trong giai đoạn đầu, đồng thời vẫn giữ khả năng khám phá để tìm ra chiến lược tốt hơn về sau. Hàm mất mát tổng hợp được định nghĩa:

$$L(\theta_i) = L_{\text{PG}} + \lambda_{\text{IL}} E\left[\left|\left|\mu_i(o_i) - \mu_E(o_i)\right|\right|^2\right], \quad (4)$$

Trong đó, $\mathcal{L}_{\mathcal{PG}}$ là mất mát theo Policy Gradient tiêu chuẩn, μ_E là chính sách chuyên gia, và λ_{IL} là hệ số trọng số giảm dần theo thời gian.

Cơ chế suy giảm của hệ số λ_{IL}: Để đảm bảo tính tái lập của quá trình huấn luyện, cơ chế giảm dần của hệ số λ_{IL} được mô hình hóa theo hàm suy giảm mũ. Cụ thể, tại tập huấn luyện thứ e, trọng số này được cập nhật theo công thức: $\lambda_{IL}(e) = \lambda_{IL}(0) \cdot (\gamma_{IL})^e$

Trong đó, $\lambda_{IL}(0)$ là giá trị trọng số khởi tạo ban đầu và $\gamma_{IL} \in (0,1)$ là hằng số tốc độ suy giảm. Việc sử dụng hàm mũ ép buộc mô hình bám sát nhãn của chuyên gia ở các tập đầu tiên để ổn định không gian trọng số mạng nơ-ron. Sau đó, λ_{IL} suy giảm tiệm cận về 0 ở các tập sau để triệt tiêu độ chệch của chiến lược tham lam, mở rộng không gian khám phá để thuật toán MARL tự hội tụ về điểm cân bằng tối ưu hơn.

3.2. Nâng cao Tính Ổn định với PER và TD3

Để tăng cường độ ổn định trong môi trường đa tác tử biến đổi nhanh, chúng tôi tích hợp hai kỹ thuật:

- **Ưu tiên Hồi tưởng Kinh nghiệm (Prioritized Experience Replay - PER)**: Thay vì lấy mẫu ngẫu nhiên đồng đều, thuật toán ưu tiên lấy mẫu các chuyển đổi có sai số dự báo (TD-error) lớn. Điều này giúp mạng nơ-ron tập trung học các tình huống "khó" hoặc bất ngờ (ví dụ: khi tác nhân gây nhiễu thay đổi vị trí đột ngột) [9].
- **Làm mượt Chính sách Mục tiêu (Target Policy Smoothing)**: Lấy cảm hứng từ thuật toán TD3, chúng tôi thêm nhiễu được cắt (clipped noise) vào hành động mục tiêu khi cập nhật mạng Critic. Kỹ thuật này giúp giảm thiểu hiện tượng ước lượng quá cao giá trị Q (Q-value overestimation), một vấn đề thường gặp trong DDPG .

3.3. Quy trình Thuật toán Chi tiết

Quy trình huấn luyện đầy đủ của IA-MADDPG được trình bày chi tiết trong Thuật toán 1.

Thuật toán 1: Imitation-Augmented MADDPG (IA-MADDPG)

Đầu vào: Các siêu tham số mạng, Môi trường truyền thông, Chuyên gia μ^E.

Đầu ra: Chính sách tối ưu cho các tác tử SU_i.

1. **Khởi tạo:**
   - Khởi tạo mạng Actor $\mu_i(o_i|\theta_i)$ và Critic $Q_i(s,a|\phi_i)$ với trọng số ngẫu nhiên.
   - Khởi tạo các mạng mục tiêu (Target networks) μ_i' và Q_i' với trọng số tương ứng.
   - Khởi tạo Bộ đệm ưu tiên (Prioritized Replay Buffer) $\mathcal{D}$.
2. **Giai đoạn Khởi động (Warm-up Phase):**
   - For $k = 1$ to K_{warmup} do:
     - Quan sát trạng thái s; Chọn hành động từ chuyên gia $a = \mu^E(s)$.
     - Thực thi hành động, nhận phần thưởng r và trạng thái mới s'.
     - Lưu bộ chuyển đổi (s,a,r,s') vào $\mathcal{D}$.
   - End For
3. **Giai đoạn Huấn luyện (Training Phase):**
   - For episode $= 1$ to M do:
     - Khởi tạo lại môi trường, nhận trạng thái ban đầu s.
     - Khởi tạo nhiễu ngẫu nhiên $\mathcal{N}$ cho việc khám phá.
     - For $t = 1$ to T (số bước mỗi tập) do:
       - Mỗi tác tử i chọn hành động $a_i = \mu_i(o_i) + \mathcal{N}_t$.

- Thực thi hành động kết hợp $a = (a_1, \dots, a_N)$, nhận phần thưởng r và trạng thái mới s'.
- Lưu (s, a, r, s') vào $\mathcal{D}$ với độ ưu tiên cao nhất.
- Lấy mẫu ngẫu nhiên một minibatch B từ $\mathcal{D}$ dựa trên trọng số ưu tiên PER.
- Cập nhật tham số (cho mỗi tác tử i):
  - Tính mục tiêu y_j sử dụng Target Policy Smoothing (TD3).
  - Cập nhật Critic bằng cách tối thiểu hóa hàm mất mát Bellman có trọng số PER.
  - Cập nhật Actor bằng cách tối thiểu hóa hàm mất mát tổng hợp.
  - Cập nhật độ ưu tiên trong $\mathcal{D}$ dựa trên TD-error mới.
- Cập nhật mạng mục tiêu:
  - $\theta_i' \leftarrow \tau\theta_i + (1-\tau)\theta_i'$
  - $\phi_i' \leftarrow \tau\phi_i + (1-\tau)\phi_i'$
- $s \leftarrow s'$

- End For (Time step)

- End For (Episode)

4. ĐÁNH GIÁ HIỆU SUẤT VÀ PHÂN TÍCH

4.1. Thiết lập Mô phỏng và Kịch bản đánh giá

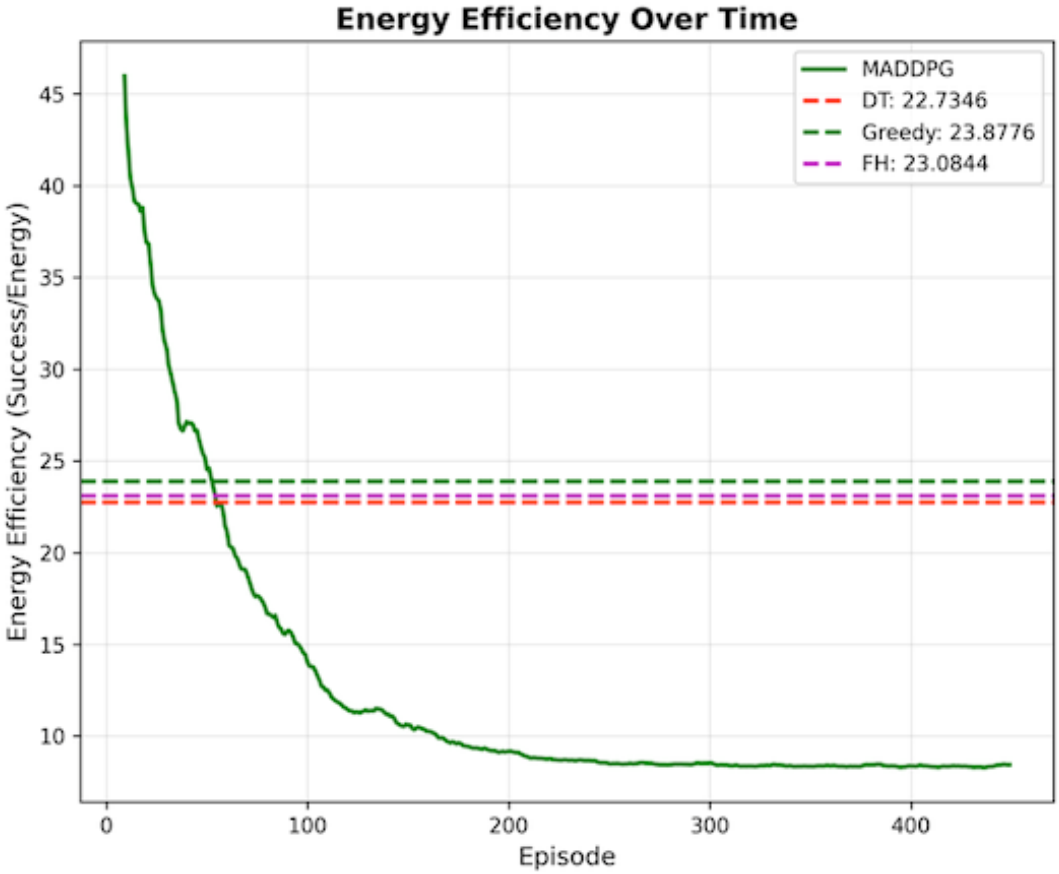
Hệ thống được triển khai trên nền tảng PyTorch với các tham số mô phỏng bám sát thực tế môi trường nhiễu động. Cụ thể, công suất tác nhân gây nhiễu được thiết lập $P_J = 1.0$ W, tạo ra áp lực nhiễu lớn lên mạng lưới. Các tác tử (Agent) sử dụng mạng nơ-ron sâu với cấu trúc Actor [256, 256, 128] và Critic [512, 256, 128, 64] để học chiến lược đối phó.

Để làm nổi bật khả năng thích nghi của thuật toán đề xuất (IA-MADDPG), chúng tôi so sánh với ba chiến lược cơ sở đại diện cho tư duy thiết kế tĩnh:

- Truyền dẫn trực tiếp (Direct Transmission - DT) & Tham lam (Greedy): Sử dụng hệ số phản xạ cố định ở mức thấp ($\alpha = 0.2$). Đây là chiến lược "an toàn" nhằm tối ưu hóa chỉ số tiết kiệm năng lượng lý thuyết.
- Nhảy tần ngẫu nhiên (Frequency Hopping - FH): Chiến lược nhảy tần ngẫu nhiên, đại diện cho phương pháp tránh nhiễu thụ động.

4.2. Phân tích Hiệu quả Năng lượng (Energy Efficiency Analysis)

Thay vì chỉ đánh giá mức tiêu thụ năng lượng một cách cơ học, phần này đi sâu phân tích hành vi thông minh của hệ thống thông qua chỉ số Hiệu quả Năng lượng (Energy Efficiency).

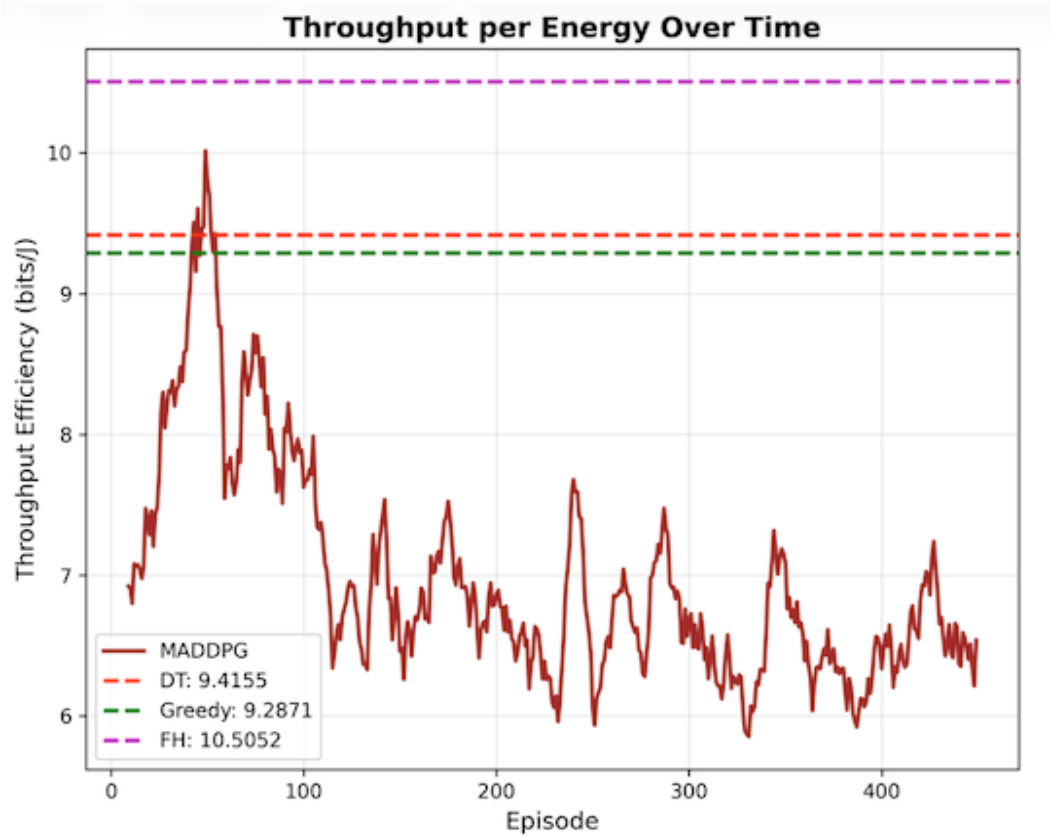


Hình 2: Hiệu quả Năng lượng theo thời gian (Energy Efficiency Over Time)

Ở Hình 2, đường biểu diễn của IA-MADDPG (màu xanh lá) hội tụ về mức thấp hơn (~9) so với các đường cơ sở (~23). Trong các nghiên cứu truyền thống, điều này thường bị hiểu nhầm là kém hiệu quả. Tuy nhiên, trong bối cảnh chống nhiễu, đây chính là bằng chứng cho khả năng thích nghi vượt trội của hệ thống đề xuất. Các phương pháp DT, Greedy và FH duy trì hiệu quả năng lượng cao giả bằng cách giữ cố định hệ số phản xạ ở mức thấp ($\alpha = 0.2$). Chiến lược này giống như việc một chiếc xe chạy chậm để tiết kiệm xăng bất chấp việc không thể leo qua dốc cao. Trong môi trường nhiễu mạnh ($P_J = 1.0$ W), mức năng lượng thấp này không đủ để tạo ra tỷ số tín hiệu trên nhiễu (SINR) cần thiết để giải mã gói tin. Do đó, sự hiệu quả của các phương pháp này thực chất là sự bất lực trong việc duy trì kết nối. Phương pháp này tiết kiệm năng lượng, nhưng đánh mất mục tiêu được ưu tiên hàng đầu là truyền tin tin cậy. Ngược lại, sự sụt giảm hiệu quả năng lượng của IA-MADDPG phản ánh hành động phản kháng tích cực. Tác tử học được rằng để duy trì kết nối dưới tác động của nhiễu, nó buộc phải đẩy hệ số phản xạ α lên mức cao (gần 1.0). Mặc dù điều này làm tăng chi phí năng lượng tức thời (theo hàm bậc hai $\eta\alpha^2$), nhưng nó đảm bảo tín hiệu vượt qua ngưỡng giải mã. Như vậy, mức hiệu quả năng lượng thấp hơn của IA-MADDPG không phải là lãng phí, mà là chi phí bảo hiểm cần thiết để đảm bảo độ tin cậy của dịch vụ (QoS), điều mà các chiến lược tiết kiệm năng lượng cực đoan không thể đáp ứng.

4.3. Phân tích Hiệu suất Thông lượng trên Năng lượng (Throughput Efficiency)

Để minh chứng rõ hơn trí thông minh của hệ thống, chúng ta xem xét chỉ số Hiệu suất Thông lượng (Throughput Efficiency - bits/Joule) trong Hình 3. Đây là thước đo cho thấy mỗi đơn vị năng lượng bỏ ra mang lại bao nhiêu giá trị thông tin thực tế.



Hình 3: Hiệu suất Thông lượng trên Năng lượng theo thời gian

Sự khác biệt giữa đường IA-MADDPG (dao động mạnh) và các đường cơ sở (đi ngang) cho thấy hai triết lý vận hành hoàn toàn trái ngược.

- Sự cứng nhắc của các phương pháp tĩnh: Các đường nét đứt nằm ngang thể hiện sự cứng nhắc. Dù môi trường nhiễu thay đổi như thế nào, các phương pháp này vẫn áp dụng một cấu hình duy nhất. Điều này dẫn đến rủi ro "Outage" (mất kết nối) cực cao khi nhiễu biến động vượt quá khả năng chịu đựng của cấu hình $\alpha = 0.2$.
- Đường dao động của IA-MADDPG (màu nâu đỏ) là minh chứng cho quá trình dò tìm và thích nghi liên tục. Các đỉnh thể hiện những khoảnh khắc tác tử phát hiện nhiễu giảm và thông minh hạ thấp công suất để tiết kiệm năng lượng. Các đáy thể hiện những thời điểm tác tử phát hiện nhiễu tăng cao và quyết định hy sinh hiệu suất bits/Joule để dồn toàn lực duy trì kết nối. Hành vi này tương tự như cơ chế điều tiết sinh học: hệ thống tự động cân bằng giữa trạng thái tiết kiệm và tiêu hao tùy thuộc vào mức độ đe dọa của máy gây nhiễu.

Kết quả mô phỏng khẳng định sự ưu việt của IA-MADDPG không nằm ở con số tiết kiệm năng lượng thô, mà nằm ở khả năng quản trị rủi ro. Trong khi các phương pháp cơ sở chấp nhận rủi ro mất tin để bảo toàn năng lượng, IA-MADDPG sử dụng năng lượng như một chiến lược để đảm bảo tính liên tục của dịch vụ. Đây là đặc tính tiên quyết cho các mạng IoT quân sự hoặc các ứng dụng y tế, an ninh quan trọng, nơi sự gián đoạn thông tin là không thể chấp nhận được.

5. KẾT LUẬN

Bài báo đã trình bày một phân tích toàn diện về sự đánh đổi giữa hiệu năng và năng lượng trong mạng Truyền thông tán xạ ngược môi trường dưới tác động của nhiễu chủ động. Thông qua việc áp dụng thuật toán IA-MADDPG, chúng tôi đã chứng minh rằng việc sử dụng các chiến lược cố định tuy tiết kiệm năng lượng nhưng không đảm bảo độ tin cậy trong môi trường thù địch. Ngược lại, việc cho phép các tác tử học máy tự chủ điều chỉnh hệ số phản xạ và chế độ truyền dẫn - mặc dù tiêu tốn năng lượng hơn - là một sự đầu tư chiến lược cần thiết. Giải pháp đề xuất cho phép duy trì kết nối ổn định bằng cách biến đổi linh hoạt cấu hình hệ thống, phù hợp với các ứng dụng yêu cầu chất lượng dịch vụ (QoS) cao trong kỷ nguyên 6G. Các hướng phát triển tiếp theo sẽ tập trung vào việc tích hợp các ràng buộc năng lượng cứng vào không gian hành động để tìm ra điểm cân bằng Pareto tối ưu hơn nữa.

TÀI LIỆU THAM KHẢO

[1] H. Tataria, M. Shafi, A. Molisch, M. Dohler, H. Sjoland, and F. Tufvesson, "6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities," Proceedings of the IEEE, pp. 1-34, 2021.

[2] H. Pirayesh and H. Zeng, "Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 767-809, 2022.

[3] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," ACM SIGCOMM Computer Communication Review, vol. 43, no. 4, pp. 39-50, 2013.

[4] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, "Ambient Backscatter Communications: A Contemporary Survey," IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2889-2922, 2018.

[5] N. Van Huynh, D. N. Nguyen, D. Thai Hoang, E. Dutkiewicz, and M. Mueck, "Ambient Backscatter: A Novel Method to Defend Jamming Attacks for Wireless Networks," IEEE Wireless Communications Letters, vol. 9, no. 2, pp. 175-178, 2020.

[6] L. Jia, N. Qi, Z. Su, F. Chu, S. Fang, K. Wong, and C. Chae, "Game theory and reinforcement learning for anti-jamming defense in wireless communications," IEEE Communications Surveys & Tutorials, 2024.

[7] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," NIPS, 2017.

[8] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, "Imitation learning: A survey of learning methods," ACM Computing Surveys, vol. 50, no. 2, pp. 1-35, 2017.

[9] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," arXiv preprint arXiv:1511.05952, 2015.