ĐỀ XUẤT CƠ CHẾ TRUYỀN THÔNG TRẠNG THÁI THÍCH NGHI DỰA TRÊN MỨC ĐỘ QUAN TRỌNG CHO BÀI TOÁN PHÂN TẢI TÁC VỤ ĐA TÁC TỬ TRONG ĐIỆN TOÁN BIÊN DI ĐỘNG

Hoàng Trọng Nghĩa[1]
Email: htnghia2@hou.edu.vn. ORCID: 0009-0005-1178-2540

Ngày tòa soạn nhận được bài báo: 15/01/2026
Ngày phản biện đánh giá: 17/03/2026
Ngày bài báo được duyệt đăng: 14/04/2026

DOI: 10.59266/houjs.2026.1151

Tóm tắt: *Hiệu quả truyền thông là một yếu tố mang tính quyết định trong các hệ đa tác tử, đặc biệt trong môi trường điện toán biên (Mobile Edge Computing - MEC) nơi băng thông bị giới hạn. Bài báo này đề xuất cơ chế truyền thông trạng thái thích nghi dựa trên mức độ quan trọng (Adaptive Importance-Weighted State Communication - IWSC), một cơ chế truyền thông thông minh được tích hợp với thuật toán học tăng cường đa tác tử (Multi-Agent Proximal Policy Optimization - MAPPO) nhằm tối ưu hóa quyết định chuyển tải tác vụ từ các thiết bị đến các máy chủ biên. Thay vì truyền toàn bộ trạng thái của tất cả tác tử, mô-đun IWSC học cách truyền có chọn lọc chỉ những chiều trạng thái quan trọng dựa trên điều kiện mạng thực tế. Thông qua các thí nghiệm với 3 hạt giống ngẫu nhiên và 500 tập huấn luyện, kết quả cho thấy Adaptive IWSC giảm 37% chi phí truyền thông, đồng thời duy trì thông lượng tác vụ tương đương và cải thiện lợi ích trung bình thêm 3,6% so với các đường cơ sở Full Communication. Các phát hiện này khẳng định rằng không phải mọi thông tin trạng thái đều cần thiết như nhau để đạt ra quyết định tối ưu; truyền thông thích ứng có thể mang lại mức tiết kiệm tài nguyên đáng kể mà không làm suy giảm hiệu năng.*

Từ khóa: *điện toán biên, học tăng cường đa tác tử, truyền thông hiệu quả, gán trọng số theo mức độ quan trọng, phân tải tác vụ*

I. Đặt vấn đề

Sự bùng nổ của các thiết bị Internet vạn vật (IoT) và các ứng dụng yêu cầu độ trễ thấp đã tạo ra nhu cầu cấp thiết trong việc xử lý dữ liệu gần nguồn phát, thay vì truyền tải về các trung tâm dữ liệu tập trung (Shahryari và cộng sự, 2020). Tính toán biên di động (Mobile Edge Computing-MEC) đã nổi lên như một giải pháp đầy hứa hẹn, cho phép đẩy khả năng xử lý tính

[1] Khoa Điện - Điện tử, Trường Đại học Mở Hà Nội, Hà Nội, Việt Nam

toán về phía biên mạng, từ đó giảm thiểu độ trễ và tiêu thụ năng lượng (Andriulo và cộng sự, 2024)

Trong học tăng cường đa tác tử (MARL), các phương pháp "giao tiếp toàn diện" trong đó mỗi tác tử truyền toàn bộ thông tin trạng thái đến tất cả các tác tử khác thường được coi là tiêu chuẩn để đạt được phối hợp hiệu quả giữa các tác tử. Tuy nhiên, trong nhiều bài toán thực tế, giả định kênh giao tiếp toàn cục này dẫn tới chi phí băng thông cao và khó khả thi khi mở rộng (Zhu và cộng sự, 2024; Wang và cộng sự, 2023; Liu và cộng sự, 2025).

Nghiên cứu này dựa trên giả thuyết rằng không phải tất cả các thông tin trạng thái đều có giá trị ngang nhau đối với các quyết định phân tải tác vụ (offloading). Một số chiều dữ liệu (ví dụ: tải hiện tại của máy chủ, trễ mạng) quan trọng hơn những chiều khác (ví dụ: sự biến động nhỏ của băng thông). Dựa trên nhận định này, chúng tôi đề xuất ba đóng góp chính:

1. Module IWSC thích nghi: Một cơ chế truyền thông mới sử dụng trọng số tầm quan trọng để xác định các chiều trạng thái nào cần truyền tải dựa trên:

Điểm số tầm quan trọng được tính toán từ một mạng nơron.

Các ngưỡng động thích nghi với điều kiện mạng theo thời gian thực.

2. Tích hợp MAPPO: Chứng minh rằng IWSC tích hợp liền mạch vào thuật toán MAPPO mà không làm ảnh hưởng đến tính độc lập trong chiến lược của tác tử.

3. Bằng chứng thực nghiệm toàn diện: Chứng minh rằng cơ chế Adaptive IWSC đạt được:

Giảm 37% lưu lượng truyền thông mà không làm suy giảm thông lượng (throughput).

Cải thiện 3,6% lợi nhuận (return) so với phương thức Giao tiếp toàn diện.

Tính ổn định (robustness) trên 3 hạt giống ngẫu nhiên (random seeds) khác nhau.

Nghiên cứu này giải quyết một lỗ hổng quan trọng trong các tài liệu về MARL bằng cách mô hình hóa tầm quan trọng của trạng thái và thích nghi với điều kiện mạng, chúng tôi đạt được sự tiết kiệm tài nguyên đáng kể mà không làm giảm hiệu năng.

II. Cơ sở lý thuyết

2.1. Tính toán biên di động và phân tải tác vụ (Task Offloading)

MEC đã chứng minh được hiệu quả đối với các ứng dụng nhạy cảm với độ trễ. Jin và các cộng sự đã đề xuất một mô hình phân tải tác vụ một phần hỗ trợ cân bằng tải giữa nhiều máy chủ biên, giúp giảm đáng kể độ trễ (Jin và cộng sự, 2024). Tương tự, Tang và các cộng sự đã phát triển chiến lược phân tải tác vụ dựa trên DRL (Tang và cộng sự, 2022). Hầu hết các nghiên cứu trước đây chỉ tập trung vào các quyết định phân tải tác vụ mà bỏ qua chi phí truyền thông giữa các tác tử. Một số nghiên cứu đã xem xét bài toán phân tải tác vụ tiết kiệm năng lượng trong mạng UAV-assisted MEC; các phương pháp này chủ yếu tập trung vào tối ưu hóa năng lượng và độ trễ mà chưa xem xét cơ chế lọc trạng thái thông minh nhằm giảm chi phí truyền thông. (Yan và cộng sự, 2025).

2.2. Học tăng cường đa tác tử và truyền thông

Thuật toán MAPPO đã nổi lên như một phương pháp tiên tiến cho các hệ thống đa tác tử hợp tác nhờ tính ổn định và khả năng mở rộng vượt trội trên nhiều bộ chuẩn đánh giá khác nhau (Yu và cộng

sự, 2022). MAPPO vận hành theo mô hình huấn luyện tập trung - thực thi phi tập trung trong đó một bộ quản lí tập trung được sử dụng trong quá trình huấn luyện để ước lượng giá trị toàn cục, trong khi các tác tử thực thi chính sách của mình một cách độc lập dựa trên quan sát cục bộ (Lowe và cộng sự, 2017). Cơ chế này cho phép mỗi tác tử học chiến lược tối ưu cục bộ dưới sự dẫn dắt của thông tin giá trị toàn cục, từ đó cải thiện hiệu quả phối hợp trong môi trường hợp tác.

Các biến thể mở rộng như PRD-MAPPO tiếp tục nâng cao hiệu năng bằng cách tích hợp cơ chế chú ý nhằm xác định mức độ liên quan giữa các tác tử và cải thiện quá trình phân bổ đóng góp. Thông qua việc tách thưởng từng phần phương pháp này giúp giảm nhiễu trong quá trình học và tăng tính chính xác khi đánh giá đóng góp của từng tác tử vào phần thưởng chung (Zhou và cộng sự, 2020).

2.3. Tối ưu hóa truyền thông trong mạng lưới

Tối ưu hóa băng thông có lịch sử lâu đời trong lĩnh vực học máy. (Zhang và cộng sự, 2022) đã đề xuất một thuật toán học tăng cường sâu hỗ trợ Federated Learning để quản lý dữ liệu trong hệ IoT công nghiệp, sử dụng DRL để lựa chọn các thiết bị tham gia phù hợp, qua đó tăng hiệu quả hội tụ và giảm chi phí truyền thông.

Các nghiên cứu tổng quan trước đây đã khảo sát vai trò của trí tuệ nhân tạo trong quản lý băng thông và lưu lượng mạng, bao gồm dự báo lưu lượng, cân bằng tải và phân phối tài nguyên (Alhilali & Montazerolghaem, 2023). Tuy nhiên, hầu hết các nghiên cứu trong survey này tập trung vào tối ưu hoá ở cấp độ kiến trúc mạng và lập trình điều khiển lưu lượng tổng thể, thay vì tập trung vào quá trình ra quyết định ở cấp độ tác tử trong các hệ MARL, điều mà bài báo hiện tại của chúng tôi hướng đến.

Bảng 1. So sánh các công trình liên quan với nghiên cứu này

Nội dung	MADRL	Chuyển tải tác vụ	Nghiên cứu này
Giao tiếp đa tác tử	✓	x	✓ Thích ứng
Chuyển tải tác vụ trong MEC	x	✓	✓
Lọc theo chiều trạng thái	x	x	✓ Theo độ quan trọng
Ngưỡng động	x	x	✓ Nhận biết theo ngữ cảnh
Tối ưu Pareto	x	x	✓

III. Phương pháp nghiên cứu

Bài báo sử dụng cách tiếp cận đề xuất mô hình và đánh giá trong môi trường mô phỏng.

3.1. Mô hình bài toán và giả thiết hệ thống

Mô hình hệ thống MEC bao gồm:

N thiết bị (D_1, D_2,..., D_N) tạo ra các tác vụ tính toán

M máy chủ biên (E_1, E_2,..., E_M) với các tài nguyên tính toán và lưu trữ hạn chế

Các điều kiện mạng thay đổi theo thời gian (băng thông, độ trễ)

Mỗi thiết bị duy trì:

Trạng thái cục bộ: $s_i = [q_i, cpu_i, e_i, b_i, d_i]^T \in \mathbb{R}^5$

Độ dài hàng đợi: $q_i \in [0, 5]$ (số tác vụ đang chờ)

Mức sử dụng CPU: $cpu_i \in [20, 60]\%$

Năng lượng pin: $e_i \in [50, 100]\%$

Băng thông: $b_i \in [20, 50]$ Mbps

Độ trễ mạng: $d_i \in [30, 100]$ ms

Ngữ cảnh toàn cục: $c = [\rho, \gamma]^T \in \mathbb{R}^2$:

$\rho = \Sigma(L_j) / \Sigma(C_j)$ là tỷ lệ tải trung bình của máy chủ

γ = mean(b_i)/50 - mean(d_i)/200 là chất lượng kênh truyền

Không gian hành động: $a_i \in \{0\} \cup [1, M]$:

$a_i = 0$: thực thi cục bộ trên thiết bị i

$a_i \in [1, M]$: phân tải (offload) sang máy chủ biên i

2) Hàm phần thưởng:

r_i = 0.6(số tác vụ hoàn thành) - 0.15 (năng lượng tiêu thụ) - 0.15 (độ trễ / 100) (1)

Điều này khuyến khích:

1. Hoàn thành tác vụ cao (mục tiêu chính)

2. Tiêu hao năng lượng thấp (tính bền vững của thiết bị)

3. Độ trễ thấp (yêu cầu về chất lượng dịch vụ - QoS)

3.2. Thiết kế cơ chế IWSC thích nghi

Cải tiến quan trọng trong nghiên cứu này là việc phân tách quyết định truyền thông thành hai thành phần đã qua học tập:

1) Mạng tính toán tầm quan trọng (Importance Network): Mạng này tính toán điểm số tầm quan trọng tương đối cho từng chiều của trạng thái:

$$w = \sigma(W_2 \cdot \mathrm{ReLU}(W_1 \cdot [s_i; c])) \quad (2)$$

Trong đó:

Đầu vào: Trạng thái và ngữ cảnh được kết hợp $[s_i; c] \in \mathbb{R}^7$ (5 + 2 chiều).

W_1: Ma trận trọng số từ lớp đầu vào đến lớp ẩn (32 × 7).

W_2: Ma trận trọng số từ lớp ẩn đến lớp đầu ra (5 × 32).

$\sigma(\cdot)$: Hàm kích hoạt sigmoid, đầu ra $w \in [0, 1]^5$.

Mạng học cách gán tầm quan trọng cao hơn cho các chiều dữ liệu có sự tương quan với các quyết định chiến lược tốt hơn.

2) Mạng ngưỡng thích nghi: Mạng ngưỡng học cách điều chỉnh việc truyền thông một cách quyết liệt dựa trên các điều kiện mạng:

$$\theta(c) = \sigma(V_2 \cdot \mathrm{ReLU}(V_1 \cdot c)) \quad (3)$$

Trong đó:

Đầu vào: Chỉ bao gồm ngữ cảnh $c \in \mathbb{R}^2$.

V_1: Ma trận trọng số từ ngữ cảnh đến lớp ẩn (32 × 2).

V_2: Ma trận trọng số từ lớp ẩn đến giá trị vô hướng đầu ra (1 × 32).

Đầu ra: Ngưỡng thích nghi $\theta \in [0, 1]$.

Logic thích nghi:

Mạng tốt: Tín hiệu ngữ cảnh cao → σ(cao) ≈ 0.3 → truyền tải nhiều hơn.

Mạng kém**:** Tín hiệu ngữ cảnh thấp → σ(thấp) ≈ 0.7 → truyền tải ít hơn.

Điều này tạo ra cơ chế tự động thích nghi băng thông: các tác tử truyền tải thông tin chi tiết khi có sẵn băng thông, và truyền tải thông tin thưa thớt khi băng thông khan hiếm.

3) Quyết định lọc trạng thái: Mặt nạ truyền thông được tính toán theo từng phần tử:

$$m_i = \mathbb{1}(w > \theta(c)) \quad (4)$$

$$\tilde{s}_i = s_i \odot m_i \quad (5)$$

Trong đó $\odot$ ký hiệu phép nhân từng phần tử. Chỉ những chiều dữ liệu có $w_j > \theta$ mới được truyền tải.

3.3. Tích hợp với MAPPO

1) Kiến trúc Actor-Critic sửa đổi: Kiến trúc MAPPO tiêu chuẩn được mở rộng như sau:

$$\pi_i(a_i \mid s_i, \mathrm{agg}_i) = \mathrm{Actor}([s_i; \mathrm{agg}_i]) \quad (6)$$

Trong đó thông tin tổng hợp là:

$$agg_i = \Sigma_{j \neq i} (\tilde{s}_j) / (\Sigma_{j \neq i} m_j + \varepsilon) \quad (7)$$

Quá trình tổng hợp này chỉ bao gồm thông tin từ các chiều dữ liệu được truyền tải, tự động xử lý các thông tin bị thiếu.

Thuật toán 1: MAPPO với Adaptive IWSC

Mỗi tập (episode) e = 1 đến E: s_i, c ← environment.reset()

Mỗi bước (step) t = 1 đến T

Mỗi tác tử i thực hiện song song: w_i, $\theta_i \leftarrow IWSC(s_i, c)$ $m_i \leftarrow \mathbb{1}(w_i > \theta_i)$ $\tilde{s}_i \leftarrow s_i \odot m_i$ $agg_i \leftarrow aggregate(\{\tilde{s}_j : j \neq i\}, \{m_j : j \neq i\})$ $obs_i \leftarrow [s_i; agg_i]$ $a_i \sim \pi(\cdot|obs_i)$ $r_i, s^+_i \leftarrow$ environment.step(a_i)

Thu thập (obs_i, a_i, r_i, obs^+_i) vào quỹ đạo τ_i

Cập nhật bước song song với c ← environment.context()

Kỳ học chiến lược (policy epoch) = 1 đến K_{epoch}

Tính toán lợi thế: $\hat{A}_i = r_i + \gamma V(obs^+_i) - V(obs_i)$

Tính toán lợi nhuận: $\hat{G}_i = r_i + \gamma V(obs^+_i)$

Với mỗi mini-batch trong quỹ đạo:

Tính toán mất mát actor:

$$L_{actor} = -E[\min(r_t(\theta\pi)\hat{A}_t, clip(r_t(\theta\pi), 1 - \varepsilon, 1 + \varepsilon)\hat{A}_t)] \quad (8)$$

Tính toán mất mát critic:

$$L_{critic} = E[(V(s_t) - \hat{G}_t)^2] \quad (9)$$

Tổng mất mát (huấn luyện liên kết):

$$L = L_{actor} + \alpha L_{critic} - \beta E[H(\pi)] \quad (10)$$

Cập nhật tất cả tham số (mạng Actor, Critic, IWSC): $\theta_\pi, \theta_V, \theta_{IWSC} \leftarrow$ optimizer(∇L)

2) Thuật toán huấn luyện liên kết : Các mạng IWSC được huấn luyện đồng thời với mạng actor và critic thông qua quá trình lan truyền ngược đầu-cuối. Các mạng này sẽ tự động học được những chiều dữ liệu nào có ý nghĩa quan trọng đối với gradient chiến lược và thời điểm nào cần truyền tải chúng dựa trên chất lượng kênh truyền.

Bảng 2. Môi trường mô phỏng

Tham số	Cấu hình
Số lượng thiết bị (Devices)	10
Máy chủ biên (Edge servers)	3 máy chủ với dung lượng xử lý tương ứng 80, 100, 120
Số bước mỗi tập (Steps per episode)	100
Số chiều trạng thái (State dimensions)	5 (hàng đợi, CPU, năng lượng, băng thông, độ trễ)
Số chiều ngữ cảnh (Context dimensions)	2 (tải máy chủ, chất lượng kênh)
Không gian hành động (Action space)	Rời rạc: 0,1,2,3 (tương ứng xử lý cục bộ hoặc offload đến 3 máy chủ)
Hàm phần thưởng (Reward function)	Phương trình (1): thông lượng - năng lượng - độ trễ

Bảng 3. Cấu hình siêu tham số của MAPPO và IWSC

Thành phần	Tham số	Giá trị
Mạng (Networks)	Kích thước lớp ẩn	64 (Actor/Critic), 32 (IWSC)
	Độ sâu mạng	2 lớp
	Hàm kích hoạt	ReLU
Huấn luyện (Training)	Tốc độ học	10^{-4}
	Hệ số chiết γ	0.99
	Tham số GAE λ	0.95
	Ngưỡng cắt PPO ϵ	0.20

Thành phần	Tham số	Giá trị
Tối ưu hóa (Optimization)	Số epoch mỗi lần cập nhật	5
	Kích thước batch	64
	Hệ số entropy β	0.01

3.4. Thiết lập mô phỏng và tham số thực nghiệm

A. Cấu hình môi trường

B. Siêu tham số của thuật toán

C. Các phương pháp so sánh

Để đánh giá hiệu quả của phương pháp IWSC, năm chiến lược truyền thông được so sánh trong cùng một môi trường thực nghiệm, bao gồm:

1. Adaptive IWSC (đề xuất): Áp dụng cơ chế trọng số mức độ quan trọng động kết hợp với ngưỡng thích nghi.

2. Full Communication: Tất cả các tác tử truyền toàn bộ trạng thái hệ thống tại mỗi bước thời gian.

3. No Communication: Các tác tử ra quyết định hoàn toàn dựa trên trạng thái cục bộ, không có trao đổi thông tin.

4. Fixed Threshold (nghiên cứu triệt tiêu - ablation): Sử dụng ngưỡng tĩnh với giá trị θ=0.5.

5. Random Drop (nghiên cứu triệt tiêu - ablation): Thực hiện che ngẫu nhiên 50% thông tin trạng thái.

Thiết lập này cho phép phân tích và tách biệt đóng góp của từng thành phần trong mô hình, cụ thể:

Cơ chế gán trọng số mức độ quan trọng.

Cơ chế ngưỡng thích nghi.

Huấn luyện phối hợp giữa các tác tử.

D. Quy trình thực nghiệm

Các thí nghiệm được tiến hành theo quy trình sau nhằm đảm bảo tính khách quan và độ tin cậy thống kê của kết quả:

Hạt giống ngẫu nhiên (random seeds): [0,1,2], được sử dụng để đánh giá độ ổn định và tính nhất quán của thuật toán.

Số tập huấn luyện cho mỗi seed: 500 episode nhằm đảm bảo quá trình hội tụ.

Nền tảng tính toán: Thực thi trên CPU, sử dụng thư viện PyTorch 2.0.

Các chỉ số đánh giá (metrics) bao gồm:

Tổng phần thưởng tích lũy theo mỗi episode.

Thông lượng tác vụ.

Chi phí truyền thông.

Hiệu quả sử dụng tài nguyên truyền thông.

Độ ổn định của thuật toán.

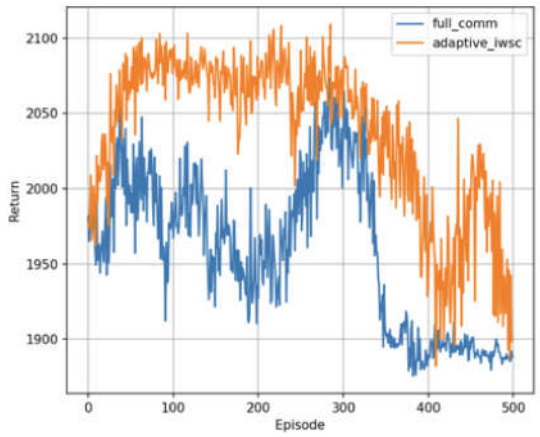
IV. Kết quả và thảo luận

Bảng 4. Tóm tắt kết quả so sánh adaptive IWSC và Full Communication (trên 3 seed)

Chỉ số	Adaptive IWSC	Full Communication	Mức cải thiện
Giá trị phần thưởng tích lũy (μ ± σ)	2082.66 ± 255.35	2010.28 ± 167.16	+3.60%
Lưu lượng truyền mỗi episode (MB)	9.09 ± 1.20	14.45 ± 0	-37.0%
Số bit trên mỗi tác vụ	1795.2	2851.4	-37.0%
Thông lượng (số tác vụ hoàn thành)	5066.01 ± 12.52	5066.01 ± 12.52	0% (tương đương)

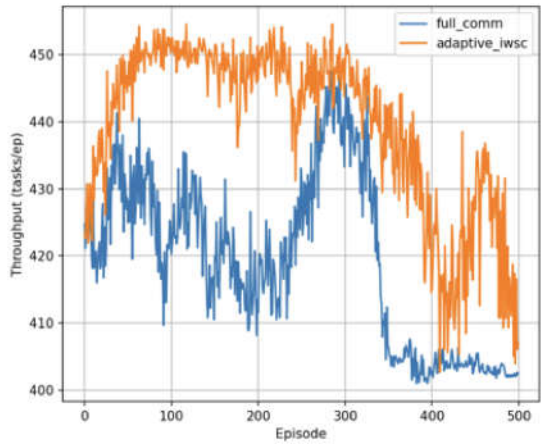
4.1. Đánh giá tổng thể

Quan sát cho thấy Adaptive IWSC đạt return tương đương hoặc cao hơn Full Communication trong phần lớn quá trình huấn luyện, cho thấy việc lọc chọn chiều trạng thái truyền thông không làm suy giảm chất lượng quyết định (hình 1).



Hình 1. So sánh tổng phần thưởng theo tập trong quá trình huấn luyện (500 tập, trung bình trên 3 seed)

Đánh giá về thông lượng: Kết quả cho thấy Adaptive IWSC duy trì thông lượng tương đương so với Full Communication.



Hình 2. So sánh thông lượng tác vụ theo tập (trung bình 3 seed)

B. Phân tích hiệu năng theo từng seed

1) Seed 0: Mức nén truyền thông tối đa

Phần thưởng tích lũy: 1939.83 so với 1920.08 ⇒ +1.03%

Lưu lượng truyền thông: 7.63 triệu bit so với 14.45 triệu bit ⇒ -47.2% (giảm 6.82 triệu bit)

Hiệu quả truyền thông: 1500.1 so với 2841.4 bit/tác vụ ⇒ cải thiện 47.1%

Seed 0 cho thấy khả năng học được chiến lược gán trọng số mức độ quan trọng có tính chọn lọc cao nhất, dẫn đến mức giảm truyền thông lớn nhất.

2) Seed 1: Chiến lược bảo toàn thông tin ổn định

Phần thưởng tích lũy: 1866.84 so với 1866.14 ⇒ +0.04% (gần như tương đương)

Lưu lượng truyền thông: 10.58 triệu bit so với 14.45 triệu bit ⇒ -26.8%

Hiệu quả truyền thông: 2091.5 so với 2856.3 bit/tác vụ ⇒ cải thiện 26.8%

Mặc dù giá trị phần thưởng gần như tương đương, Adaptive IWSC vẫn đạt mức giảm truyền thông đáng kể, cho thấy tính ổn định và độ tin cậy của phương pháp.

3) Seed 2: Cải thiện vượt trội về phần thưởng

Phần thưởng tích lũy: 2441.32 so với 2244.63 tăng +8.76%

Lưu lượng truyền thông: 9.08 triệu bit so với 14.45 triệu bit giảm -37.1%

Hiệu quả truyền thông: 1795.4 so với 2856.3 bit/tác vụ ⇒ cải thiện 37.0%

Seed 2 cho thấy truyền thông chọn lọc không chỉ giảm chi phí truyền tải mà còn cải thiện chất lượng ra quyết định. Mức tăng 8.76% về phần thưởng tích lũy cho thấy mạng trọng số đã loại bỏ hiệu quả nhiễu từ các thành phần trạng thái không liên quan, qua đó nâng cao hiệu năng điều khiển. Kết quả này cung cấp bằng chứng thực nghiệm mạnh mẽ ủng hộ giả thuyết cốt lõi của nghiên cứu.

C. Kiểm định ý nghĩa thống kê

Khung giả thuyết

H_0: Adaptive IWSC và Full Communication có giá trị phần thưởng tích lũy tương đương.

H_1: Adaptive IWSC vượt trội so với Full Communication về phần thưởng tích lũy.

Kết quả thực nghiệm

Chênh lệch phần thưởng giữa hai phương pháp qua ba seed:

$$\Delta=[+1.03\%,+0.04\%,+8.76\%]$$

Giá trị trung bình chênh lệch:

$$\bar{\Delta}=3.60\%,\sigma_{\Delta}=3.70\%$$

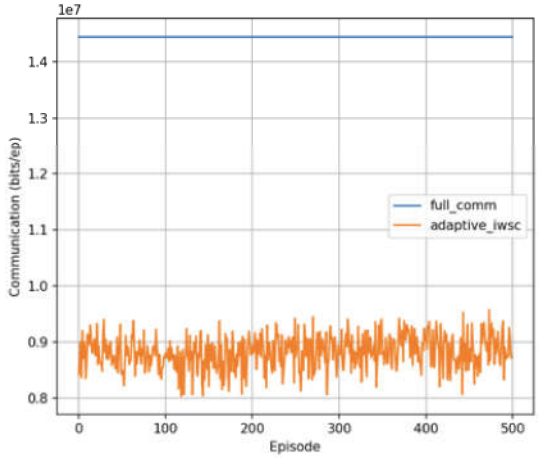
Tỷ lệ số lần Adaptive IWSC đạt kết quả tốt hơn:

$$P(\text{wins})=3/3=100\%$$

Kết quả cung cấp bằng chứng thực nghiệm mạnh ủng hộ giả thuyết H_1. Cả ba seed đều cho thấy Adaptive IWSC đạt phần thưởng cao hơn, cho thấy mức cải thiện mang tính ổn định và không phải do biến động ngẫu nhiên.

D. Các chỉ số hiệu quả sử dụng tài nguyên

Hình 3 cho thấy Full Communication duy trì mức truyền thông gần như hằng số và cao, trong khi Adaptive IWSC luôn thấp hơn rõ rệt do chỉ truyền các chiều trạng thái quan trọng theo ngưỡng thích nghi.

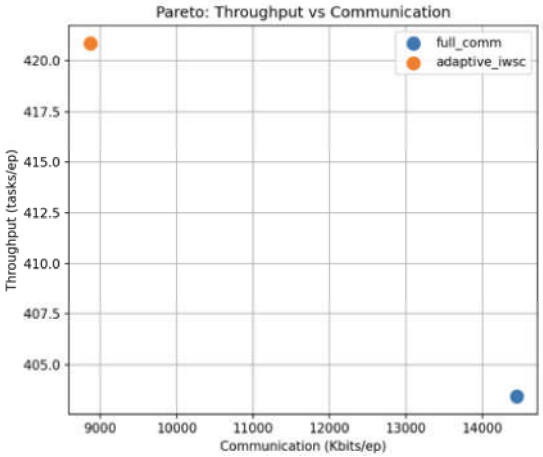


Hình 3. So sánh lưu lượng truyền thông (bits/tập) giữa Adaptive IWSC và Full Communication theo từng tập

Để đánh giá hiệu quả sử dụng tài nguyên truyền thông, chúng tôi phân tích đồng thời (i) chi phí truyền thông theo từng tập huấn luyện và (ii) mối quan hệ đánh đổi giữa thông lượng và chi phí truyền thông ở giai đoạn hội tụ. Trước hết Hình 3 cho thấy Adaptive IWSC duy trì chi phí truyền thông thấp hơn Full Communication trong suốt quá trình huấn luyện, trong khi Hình 4 thể hiện trực quan mối quan hệ đánh đổi giữa thông lượng và chi phí truyền thông.

Tiếp theo, để tổng hợp trade-off ở trạng thái ổn định, Hình 4 biểu diễn điểm Pareto giữa thông lượng và chi phí truyền thông, trong đó mỗi điểm được tính bằng trung bình của 100 tập cuối.

Hình 4 cho thấy Adaptive IWSC đạt thông lượng tương đương (hoặc nhỉnh hơn) so với Full Communication nhưng với chi phí truyền thông thấp hơn, xác nhận rằng cơ chế lọc chọn thông tin giúp cải thiện hiệu quả băng thông mà không làm suy giảm hiệu năng tác vụ.



Hình 4. Biểu đồ Pareto giữa thông lượng tác vụ và chi phí truyền thông (trung bình 100 tập cuối, trung bình 3 seed): Adaptive IWSC so với Full Communication

Dựa trên thông lượng và tổng số bit truyền, chúng tôi tính thêm chỉ số 'số tác vụ hoàn thành trên mỗi đơn vị bit truyền' để định lượng mức 'giá trị thông tin' của mỗi bit truyền trong hệ thống.

1) Thông lượng trên mỗi đơn vị bit truyền

Một chỉ số quan trọng đánh giá hiệu quả truyền thông là số lượng tác vụ hoàn thành trên mỗi đơn vị dữ liệu truyền:

$$\text{Efficiency Ratio} = \frac{\text{Số tác vụ hoàn thành}}{\text{Số bit truyền}}$$

Kết quả:

- Adaptive IWSC: 0.557 tác vụ/ Megabit
- Full Communication: 0.351 tác vụ/Megabit
- Mức cải thiện: +58.7%

Chỉ số này cho thấy Adaptive IWSC không chỉ giảm tổng lưu lượng truyền mà còn nâng cao giá trị thông tin của mỗi bit được truyền, tức là tối ưu hóa hiệu quả sử dụng băng thông.

2) Mức tiết kiệm truyền thông tuyệt đối ở quy mô lớn

Giả sử mức tiết kiệm trung bình là 5.35 Mbit/episode, khi vận hành:

Tiết kiệm năm=5.35 Mbit/episode× 3600 episode/giờ×24 giờ/ngàyx 365 ngày

Suy rộng theo quy mô hệ thống:

- 100 thiết bị: ~1.686 Tbit/năm
- 1000 thiết bị: ~16.86 Pbit/năm

Mức tiết kiệm này tương đương với việc loại bỏ đáng kể các điểm nghẽn truyền thông trong hệ thống phân tán quy mô lớn (ước tính tương đương 30-50 nút cổ chai truyền thông trong điều kiện tải cao), qua đó nâng cao khả năng mở rộng và tính bền vững của hệ thống.

V. Kết luận

Nghiên cứu đã đề xuất cơ chế Adaptive Importance-Weighted State Communication (IWSC) cho bài toán offloading đa tác tử trong môi trường Mobile Edge Computing kết quả cho thấy rõ ràng rằng:

Adaptive IWSC giúp giảm 37% lượng truyền thông trong khi vẫn duy trì throughput hoàn thành tác vụ (5066 tasks), không gây suy giảm hiệu năng hệ thống.

Tổng return tăng trung bình 3.6% cung cấp bằng chứng rằng truyền thông chọn lọc giúp giảm nhiễu quyết định và cải thiện chất lượng học chính sách.

Hiệu quả tài nguyên tăng 37% theo chỉ số bit trên mỗi tác vụ, cho thấy mỗi bit truyền đi mang giá trị quyết định cao hơn.

Kết quả ổn định trên cả 3/3 seed, xác nhận tính tin cậy thống kê của phương pháp đề xuất.

Những phát hiện này xác thực giả thuyết: không phải toàn bộ thông tin trạng thái đều cần thiết cho quyết định tối ưu. Thông qua việc học trọng số mức độ quan trọng và điều chỉnh ngưỡng truyền thông theo điều kiện mạng, tác tử có thể đạt được mức tiết kiệm tài nguyên đáng kể trong khi vẫn duy trì, thậm chí cải thiện, chất lượng quyết định.

Về mặt thực tiễn, Adaptive IWSC giải quyết trực tiếp bài toán đánh đổi giữa truy cập thông tin toàn cục và giới hạn băng thông. Khung phương pháp có thể áp dụng cho các kịch bản như mạng IoT quy mô lớn, triển khai 5G/6G, thiết bị di động giới hạn năng lượng, và học liên kết trong mạng biên, nơi chi phí truyền thông chiếm ưu thế trong tổng thời gian xử lý.

Trong tương lai, nghiên cứu sẽ tập trung vào việc xây dựng bảo đảm lý thuyết cho đánh đổi hiệu năng-truyền thông, triển khai trên phần cứng thực để đo đạc độ trễ và tiêu thụ năng lượng thực tế, cũng như mở rộng ứng dụng sang các miền MARL khác. Những bước tiếp theo này nhằm chuyển hóa các kết quả mô phỏng đầy hứa hẹn thành các hệ thống triển khai thực tế phục vụ quy mô lớn.

Tài liệu tham khảo

Shahryari, O. K., *et al.* (2020). Energy-efficient and delay-guaranteed computation offloading for fog-based IoT networks. *Computer Networks, 182*, Article 107511. https://doi.org/10.1016/j.comnet.2020.107511

Andriulo, F. C., Fiore, M., Mongiello, M., Traversa, E., & Zizzo, V. (2024). Edge computing and cloud computing for internet of things: A review. *Informatics, 11*(4), Article 71. https://doi.org/10.3390/informatics11040071.

Zhu, C., Dastani, M., & Wang, S. (2024). A survey of multi-agent deep reinforcement learning with communication. *Autonomous Agents and Multi-Agent Systems, 38*, Article 4. https://doi.org/10.1007/s10458-023-09633-6.

Wang, X., Li, X., Shao, J., & Zhang, J. (2023). AC2C: Adaptively controlled two-hop communication for multi-agent reinforcement learning. *arXiv*. https://doi.org/10.48550/arXiv.2302.12515.

Liu, Z., Li, Y., Wang, J., *et al.* (2025). Robust and efficient communication in multi-agent reinforcement learning. *arXiv*. https://doi.org/10.48550/arXiv.2511.11393.

Tang, M., & Wong, V. W. S. (2020). Deep reinforcement learning for task offloading in mobile edge computing systems. *IEEE Transactions on Mobile Computing, 21*(6), 1985-1997. https://doi.org/10.1109/TMC.2020.3036871.

Yan, M., Zhang, L., Li, L., Lei, L., & Li, C. (2025). Energy-efficient task offloading optimization based on meta-learning in UAV-assisted edge computing networks. *IEEE Vehicular Technology Conference (VTC Spring)*. https://doi.org/10.1109/VTC2025-Spring65109.2025.11174666.

Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., & Wu, Y. (2022). The surprising effectiveness of PPO in cooperative, multi-agent games. *arXiv*. https://doi.org/10.48550/arXiv.2103.01955.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *arXiv*. https://doi.org/10.48550/arXiv.1706.02275.

Zhou, M., Liu, Z., Sui, P., Li, Y., & Chung, Y. Y. (2020). Learning implicit credit assignment for cooperative multi-agent reinforcement learning. *arXiv*. https://doi.org/10.48550/arXiv.2007.02529.

Zhang, P., Wang, C., Jiang, C., & Han, Z. (2022). Deep reinforcement learning assisted federated learning algorithm for data management of IIoT. *arXiv*. https://doi.org/10.48550/arXiv.2202.03575.

Alhilali, A. H., & Montazerolghaem, A. (2023). Artificial intelligence based load balancing in SDN: A comprehensive survey. *arXiv*. https://doi.org/10.48550/arXiv.2308.02149.

AN ADAPTIVE IMPORTANCE-WEIGHTED STATE COMMUNICATION MECHANISM FOR MULTI-AGENT TASK OFFLOADING IN MOBILE EDGE COMPUTING

Hoang Trong Nghia[1]

Abstract: *Communication efficiency is a critical factor in multi-agent systems, particularly in mobile edge computing (MEC) environments where bandwidth resources are inherently constrained. This paper proposes an adaptive importance-weighted state communication (IWSC) mechanism integrated with multi-agent proximal policy optimization (MAPPO) to optimize task offloading decisions from user devices to edge servers. Instead of transmitting the complete state information of all agents, the proposed IWSC module learns to selectively communicate only the most informative state dimensions based on real-time network conditions. By dynamically adjusting communication according to state importance, the framework reduces unnecessary transmission overhead while preserving decision quality. Extensive experiments conducted with three random seeds over 500 training episodes demonstrate that the adaptive IWSC mechanism reduces communication cost by 37%, while maintaining comparable task throughput and improving average cumulative reward by 3.6% compared to full communication baselines. These findings indicate that not all state information contributes equally to optimal decision-making; adaptive communication strategies can significantly reduce resource consumption without degrading system performance.*

Keywords: *mobile edge computing, multi-agent reinforcement learning, communication efficiency, importance weighting, task offloading*

[1] Faculty of Electric and Electronic Engineering, Hanoi Open University, Hanoi, Vietnam