

NÂNG CAO CHẤT LƯỢNG TRẢI NGHIỆM DỊCH VỤ CỦA NGƯỜI DÙNG CHO MẠNG THÔNG TIN THẾ HỆ MỚI

ON THE ENHANCEMENT OF USER QUALITY OF EXPERIENCE FOR NEXT GENERATION WIRELESS NETWORKS

Lê Thị Phương Mai*, Thái Văn Tiến, Đặng Thiên Bình

Trường Đại học Bách khoa - Đại học Đà Nẵng, Việt Nam¹

*Tác giả liên hệ / Corresponding author: lpm@dut.udn.vn

(Nhận bài / Received: 21/3/2024; Sửa bài / Revised: 21/4/2024; Chấp nhận đăng / Accepted: 22/4/2024)

Tóm tắt - Bài báo này nghiên cứu việc ứng dụng công nghệ điện toán biên di động MEC tại các trạm phát đa anten MIMO cho mạng vô tuyến thế hệ mới (sau 5G/6G). Trong mô hình hệ thống này, nhóm tác giả nghiên cứu vấn đề tối ưu chất lượng trải nghiệm của người dùng thông qua việc giảm độ trễ thời gian offloading dữ liệu. Cụ thể, nhóm tác giả thiết lập bài toán tối ưu đồng thời công suất phân bổ cho các người dùng và tỉ lệ offloading dữ liệu/tính toán cục bộ cho một mạng vô tuyến. Để giải quyết bài toán đưa ra, nhóm tác giả đề xuất một thuật toán dựa trên mô hình học tăng cường sâu (DRL), với các ràng buộc về công suất phân bổ cho người dùng và tỉ lệ offloading dữ liệu. Cuối cùng, trình bày các kết quả mô phỏng tính toán số để kiểm chứng tính hiệu quả của thuật toán được đề ra.

Từ khóa - DDPG; MIMO; điện toán biên di động (MEC); phân bổ công suất; học tăng cường sâu (DRL); IoT; kết nối không lồ.

1. Giới thiệu chung

Kể từ cuối thập kỷ trước, mạng không dây thế hệ thứ 5 (5G) đã được triển khai trên toàn cầu và ở một số quốc gia và đã đạt tới quy mô rất lớn. Tuy nhiên, sự xuất hiện của các ứng dụng và dịch vụ mới cùng với sự tăng trưởng chưa từng có cả về lưu lượng không dây và số lượng thuê bao di động đã thúc đẩy mạnh mẽ việc nghiên cứu và phát triển của mạng không dây thế hệ thứ sáu (6G) [1].

Điện toán biên di động (Mobile Edge Computing - MEC) là một trong những giải pháp triển vọng cho các ứng dụng đòi hỏi tính toán phức tạp và độ trễ quan trọng như thực tế ảo (Virtual Reality - VR), thực tế tăng cường (Augmented Reality - AR), xe tự hành, phẫu thuật từ xa, phương tiện bay không người lái (Unmanned Autonomous Vehicles - UAVs) và Internet vạn vật (Internet-of-Things - IoT). Vì những ứng dụng này yêu cầu độ trễ nghiêm ngặt và tính toán phức tạp, khả năng tính toán của các thiết bị có thể không đáp ứng kịp với nhu cầu thực tế. So với điện toán đám mây, trong các hệ thống MEC, người dùng offload các tác vụ tính toán tải phức tạp cho các máy chủ MEC với khả năng xử lý tính toán số mạnh mẽ và ở vị trí gần người dùng để thực thi [2]. Do đó, MEC có thể giảm đáng kể độ trễ tính toán và giảm tải lưu lượng trên mạng backhaul [3]. Trong hệ thống MEC, việc đạt được hoạt động tính toán trên cạnh biên của mạng cũng phụ thuộc vào tốc độ truyền dữ liệu. Một trong những công nghệ chính cho thế hệ mạng tiên tiến hiện nay như Massive MIMO đang được áp dụng ngày càng nhiều trong các hệ thống vô tuyến [4], [5]. Nhờ

Abstract – This paper investigates the application of Mobile Edge Computing (MEC) technology at base stations (BS) equipped with MIMO (Multiple Input Multiple Output) antennas for next-generation wireless networks (i.e. beyond 5G/6G). In this model, the authors aim to improve the user experience by reducing data offloading latency. Specifically, the authors formulate a joint optimization problem of power allocation and data offloading/local computing ratio for the considered wireless network. To address this problem, the authors propose an algorithm based on Deep Reinforcement Learning (DRL), involving the constraints of user power allocation and data offloading ratio. Finally, we provide numerical results via simulation to verify the effectiveness of the proposed algorithm.

Key words - DDPG; MIMO; Mobile Edge Computing (MEC); Power Allocation; Deep Reinforcement Learning (DRL); IoT; massive connectivity.

vào việc sử dụng một số lượng lớn anten tại trạm phát (Base Station - BS), Massive MIMO có thể đồng thời phục vụ một số lượng người dùng cao hơn và cải thiện đáng kể hiệu suất phổ và năng lượng của hệ thống. Vì vậy, việc kết hợp Massive MIMO với MEC có thể mang lại tốc độ truyền dữ liệu cao hơn, đồng thời hỗ trợ một số lượng người dùng lớn hơn trong việc giảm tải (offloading) và giảm độ trễ truyền dữ liệu không dây, đặc biệt là ở đường lên (uplink - UL) [6].

1.1. Tổng quan các công trình liên quan

Một số nghiên cứu gần đây liên quan đến hệ thống kết hợp Massive MIMO và MEC có thể kể đến như [7]–[13]. Cụ thể, các tác giả trong [7] đã nghiên cứu việc tối ưu hóa tiêu thụ năng lượng và giảm thiểu độ trễ tối đa cho một hệ thống MEC được hỗ trợ bởi Massive MIMO với ràng buộc về công suất phát. Trong khi đó, công trình [8] nghiên cứu việc tối ưu đồng thời phân bổ tài nguyên năng lượng và kỹ thuật beamforming cho hệ thống điện toán biên di động đa truy cập bao gồm các điểm truy cập (Access Points - AP) đa anten và người dùng đơn anten với mục tiêu tối ưu hiệu suất năng lượng (Energy Efficiency - EE) của hệ thống MEC. Tương tự, trong [9], các tác giả đã thiết lập bài toán tối ưu năng lượng nhằm giảm thiểu tổng năng lượng tiêu thụ của hệ thống MEC đa người dùng trong khi vẫn đảm bảo điều kiện độ trễ tối thiểu trong quá trình truyền tải lên và tải xuống.

Về vấn đề liên quan đến kênh truyền có thông tin trạng thái kênh (Channel State Information - CSI) hoàn hảo và

¹ The University of Danang – University of Science and Technology, Vietnam (Le Thi Phuong Mai, Thai Van Tien, Dang Thien Binh)

không hoàn hảo, nghiên cứu [10] đã thiết lập bài toán tối ưu cho một hệ thống MIMO đơn cell nhằm giảm thiểu tiêu thụ năng lượng có tính toán trọng số tối đa đối với các nguồn tài nguyên tính toán cố sẵn và độ trễ dưới giới hạn cho phép.

Các nghiên cứu tiếp theo xem xét việc giảm thiểu độ trễ cho các hệ thống MEC dựa trên công nghệ Massive MIMO.

Cụ thể, các tác giả trong [11] đã nghiên cứu vấn đề tối ưu đồng thời phân bổ tài nguyên truyền thông và tính toán kết hợp với các ràng buộc về tài nguyên tính toán tại máy chủ MEC, công suất truyền tải dữ liệu và tín hiệu điều khiển cho một mạng MEC-Massive MIMO đơn cell. Mục tiêu chính của bài toán tối ưu trong công trình này là nhằm giảm thiểu độ trễ offloading lớn nhất khi phục vụ nhiều người dùng. Mặt khác, vấn đề giải quyết bài toán giảm thiểu độ trễ tính toán offloading lớn nhất cho tất cả người dùng đã được trình bày trong công trình [12] cho mạng MEC đơn cell đa người dùng với Massive MIMO. Trong khi đó, nghiên cứu [13] đã đề xuất một sơ đồ offloading nhi phân cho hệ thống gồm người dùng và trạm phát đa anten, với mục tiêu giảm thiểu tổng chi phí của tổng trọng số tiêu thụ năng lượng và độ trễ thời gian bằng cách kết hợp thiết kế bộ tiền mã hóa và phân bổ tài nguyên tính toán.

1.2. Các đóng góp chính và cấu trúc bài báo

Trong bài báo này, nhóm tác giả tập trung vào việc tối ưu hóa chất lượng trải nghiệm (QoE) tại các điểm biên của một mạng vô tuyến kết hợp công nghệ MEC và MIMO thông qua việc giảm thiểu thời gian tính toán trong khi đáp ứng được các yêu cầu phân bổ công suất. Các đóng góp chính của bài báo có thể được tóm tắt như sau:

- Bài báo đề xuất một mô hình toán học kết hợp tính toán biên và chia sẻ tài nguyên trong mạng MEC-MIMO. Tiếp theo, nhóm tác giả thiết lập bài toán tối ưu hóa nhằm giảm thiểu thời gian tính toán tối đa, trong đó vấn đề này thuộc dạng bài toán MINC (Mixed-Integer Non-Convex) không liên tục.

- Để giải quyết bài toán tối ưu này, đề xuất một thuật toán dựa trên học tăng cường sâu (Deep Reinforcement Learning), cụ thể là dựa trên phương pháp Deep Deterministic Policy Gradient (DDPG), trong đó tính đến các ràng buộc về tỷ lệ offloading và phân bổ công suất.

Chú thích: Trong bài báo này, $diag(\mathbf{x})$ ký hiệu cho ma trận đường chéo. Vector cột và ma trận được ký hiệu bằng chữ cái viết đậm chữ thường và chữ cái viết đậm in hoa. Các ký hiệu $E[\cdot]$, $|\cdot|$, và $\|\cdot\|$ ký hiệu lần lượt cho các toán tử kỳ vọng, giá trị tuyệt đối và chuẩn Euclid. Phép toán tử Hermitian của vector \mathbf{x} và ma trận \mathbf{A} được ký hiệu bởi \mathbf{x}^H và \mathbf{A}^H , và phương sai được ký hiệu là $\text{Var}(\cdot)$. Ngoài ra, $x \sim \mathcal{CN}(0,1)$ quy ước cho biểu diễn của biến x tuân theo phân phối Gaussian phức đối xứng với trung bình bằng không và phương sai bằng một.

2. Mô hình hệ thống

Trong bài báo này, nhóm tác giả xem xét kịch bản một mạng MEC-MIMO với mật độ người dùng cao. Trạm cơ sở (BS) được trang bị M ăng-ten để phục vụ K người dùng (User Equipments - UE) đơn ăng-ten. Nhằm tăng khả năng tính toán cho người dùng, nhóm tác giả đề xuất đặt một máy chủ MEC tại BS để nhận dữ liệu giảm tải (offloading) từ các UEs.

Trong mô hình này, quá trình truyền dữ liệu đường lên UL được xem như một quá trình offloading. Không mất

tính tổng quát, giả sử rằng đáp ứng kênh từ UE k , với $k \in \mathcal{K} \triangleq \{1, 2, \dots, K\}$ đến BS được ký hiệu là $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$. Các kênh được đặc trưng bởi fading quy mô nhỏ và fading quy mô lớn, tức là $\mathbf{h}_k = \beta_k \bar{\mathbf{h}}_k$. Ở đây, fading quy mô lớn β_k biểu thị cho mất mát đường truyền và hiện tượng shadowing gây ra, trong khi đó fading quy mô nhỏ ký hiệu bởi $\bar{\mathbf{h}}_k$, được mô tả là một vector của biến ngẫu nhiên, với mỗi phần tử của vector tuân theo phân phối phức Gaussian $\mathcal{CN}(0,1)$.

Tín hiệu nhận được tại BS, ký hiệu là $\mathbf{r} \in \mathbb{C}^{M \times 1}$, có thể được biểu diễn như sau:

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{z}, \quad (1)$$

Trong đó, ma trận $\mathbf{H} \triangleq [\mathbf{h}_1 \ \mathbf{h}_2 \ \dots \ \mathbf{h}_K] \in \mathbb{C}^{M \times K}$. Vector $\mathbf{z} \in \mathbb{C}^{M \times 1}$ ký hiệu cho nhiễu trắng cộng (AWGN), trong đó mỗi phần tử của vector này tuân theo phân phối Gaussian phức $\mathcal{CN}(0, \sigma_{BS}^2)$, với σ_{BS}^2 là công suất nhiễu của mỗi phần tử. Tín hiệu truyền $\mathbf{x} \in \mathbb{C}^{K \times 1}$ được xác định như sau:

$$\mathbf{x} = \text{diag}(\mathbf{p})\tilde{\mathbf{x}}, \quad (2)$$

Trong đó, công suất phân bổ cho các UEs ký hiệu bởi $\mathbf{p} \triangleq [\sqrt{p_1} \ \sqrt{p_2} \ \dots \ \sqrt{p_K}]$, với p_k là hệ số công suất phân bổ cho UE k . Nhóm tác giả ký hiệu $\tilde{\mathbf{x}} \in \mathbb{C}^{K \times 1}$ là vector tín hiệu của K UEs, với mỗi phần tử $\tilde{x}_k, \forall k \in \mathcal{K}$ có kỳ vọng công suất của tín hiệu từ UE k đến BS xác định bởi $E[|\tilde{x}_k|^2] = 1$.

Giả sử rằng chúng ta sử dụng một máy thu tuyến tính tại trạm phát BS, tín hiệu $y \in \mathbb{C}^{K \times 1}$ hậu xử lý từ K UEs được xác định như sau:

$$\mathbf{y} = \mathbf{A}^H \mathbf{r} = \mathbf{A}^H \mathbf{H} \tilde{\mathbf{x}} + \mathbf{A}^H \mathbf{z}, \quad (3)$$

với ma trận $\mathbf{A} \in \mathbb{C}^{M \times K}$ biểu thị cho bộ tách sóng tuyến tính với tín hiệu từ UE k , tương ứng cột k , được chuẩn hóa bởi $\|\mathbf{a}_k^H \mathbf{a}_k\|^2 = 1$. Từ đây, tỉ số tín hiệu trên nhiễu SINR (signal-to-interference-plus-noise) để giải mã tín hiệu UE k được tính như sau:

$$\gamma_k = \frac{p_k |\mathbf{a}_k^H \mathbf{a}_k|^2}{\sum_{k' \neq k} p_{k'} |\mathbf{a}_k^H \mathbf{a}_{k'}|^2 + \sigma_{BS}^2} \quad (4)$$

Tốc độ truyền dữ liệu khả thi cho quá trình offload của UE k có thể biểu diễn như sau:

$$R_k(\mathbf{p}) = B \log_2(1 + \gamma_k), \quad [\text{bits/s}] \quad (5)$$

với B là băng thông hệ thống.

Quy trình thực hiện tính toán cho người dùng k có thể được chia tương ứng thành hai thành phần: offloading và tính toán cục bộ (local computing), trong đó thời gian xử lý cho các thành phần này được biểu diễn lần lượt bởi τ_k^{off} và τ_k^{loc} . Cụ thể, thời gian tính toán được xác định như sau:

$$\tau_k^{\text{MEC}} = \max\{\tau_k^{\text{off}}, \tau_k^{\text{loc}}\}, \quad (6)$$

Trong đó, thời gian cho việc offloading và tính toán cục bộ được xác định bởi

$$\tau_k^{\text{off}} = \frac{\rho_k Z_k}{R_k(\mathbf{p})} + \frac{\rho_k Z_k C_k}{f^{\text{SER}}} \quad (7)$$

$$\tau_k^{\text{loc}} = \frac{(1-\rho_k) Z_k C_k}{f_k} \quad (8)$$

với ρ_k và Z_k ký hiệu cho tỷ lệ và kích thước (dưới dạng bit) của nhiệm vụ offloading được yêu cầu bởi người dùng k . Trong khi đó, C_k đại diện cho số chu kỳ CPU cần thiết để thực thi một bit trong nhiệm vụ có Z_k bit. Ngoài ra, f^{SER} và f_k lần lượt đại diện cho tần số tính toán tại máy chủ

MEC và người dùng k .

Trong bài báo này, nhóm tác giả tập trung vào việc nâng cao chất lượng trải nghiệm QoE tại biên của mạng MEC-MIMO bằng cách giảm độ trễ tính toán τ_k^{MEC} . Điều này nhằm mục đích nâng cao trải nghiệm người dùng trong các ứng dụng thời gian thực. Sự công bằng QoE trong quản lý mạng được đặt ra dưới dạng một vấn đề tối ưu nhằm giảm thời gian thực thi cho tất cả người dùng:

$$\text{minimize } \max_{p, \rho} \max_{k \in \mathcal{K}} \{\tau_k^{\text{MEC}}\} \quad (9a)$$

$$\text{s. t. } 0 \leq p_k \leq 1 \quad (9b)$$

$$\rho_k \in \{0, 1\}, k \in \mathcal{K}, \quad (9c)$$

Điều kiện ràng buộc (9b) liên quan đến việc phân bổ công suất, cụ thể hơn là hạn chế lượng công suất truyền tải mà người dùng k có thể sử dụng không vượt quá giá trị giới hạn đã xác định trước. Ràng buộc (9c) đảm bảo rằng việc xử lý dữ liệu cho người dùng k được thực hiện thông qua tính toán cục bộ ($\rho_k = 0$) hoặc offload dữ liệu ($\rho_k = 1$).

3. Đề xuất thuật toán tối ưu cho bài toán đề ra

Để giải quyết hiệu quả bài toán (9), trong phần này nhóm tác giả giới thiệu thuật toán đề xuất cho bài toán tối ưu thiết lập ở trên, trong đó thuật toán này dựa trên phương pháp DDPG, một thuật toán phổ biến trong mảng trí tuệ nhân tạo (Artificial Intelligence - AI).

3.1. Giới thiệu về DDPG

Trong lĩnh vực học tăng cường sâu (Deep Reinforcement Learning - DRL), DDPG là một phương pháp mạnh mẽ nhờ khả năng xử lý không gian hành động liên tục. Bằng cách sử dụng một phần của MDPs (Markov Decision Processes), DDPG cho phép các tác nhân tối ưu hóa tổng thưởng trong một số bước thời gian rời rạc hạn chế. Chúng ta có thể đưa ra các giải pháp gần đúng (heuristics) thông qua việc đánh giá các trạng thái, hành động, xác suất chuyển tiếp và thưởng. Cụ thể hơn, giá trị chính được tăng dần theo cập nhật mục tiêu Bellman:

$$y = r^{(t)} + \gamma Q \left(\theta_c^{(target)} \Big|_{s^{(t+1)}, a'} \right), \quad (10)$$

với mạng mục tiêu critic được biểu diễn bởi $\theta_c^{(target)}$. Để đạt được điều này, bộ nhớ lưu trữ, được biểu diễn bằng ký hiệu D , bao gồm một tập hợp các chuyển tiếp sau (s, a, r, s') .

Mô hình DDPG cập nhật mạng mục tiêu thông qua kỹ thuật trung bình Polyak như sau:

$$\theta_a^{(target)} \leftarrow \rho \theta_a^{(train)} + (1 - \rho) \theta_a^{(target)}, \quad (11)$$

$$\theta_c^{(target)} \leftarrow \rho \theta_c^{(train)} + (1 - \rho) \theta_c^{(target)}, \quad (12)$$

Trong đó, siêu tham số (hyperparameter) được ký hiệu bởi ρ , thường được gán một giá trị trong khoảng từ 0 đến 1, thường gần với 1.

Sai số trung bình bình phương (Mean Squared Error - MSE) được sử dụng để đo lường cải thiện hiệu suất bằng cách so sánh giá trị Q tại bước thời gian t với mục tiêu cập nhật Bellman. Phương pháp tối thiểu hóa toán học MSE bao gồm kết hợp giữa phần thưởng $r^{(t)}$ và giá trị Q tại bước thời gian $t + 1$, được biểu diễn qua công thức sau:

$$\mathcal{L}(\theta_c^{(train)}) = E_{(s, a, r, s')} \mathbb{D} \sim (Q(\theta_c^{(train)}) \Big|_{s^{(t+1)}, a'} - y)^2 \quad (13)$$

Bảng 1. Tham số mô phỏng

Tham số	Giá trị
Bán kính của micro-cell	100 m
Băng thông hệ thống	20 MHz
Công suất nhiễu tại UE	-174 dBm/Hz
Công suất huấn luyện uplink UEs	$P_{\text{up}} = 10$ dBm
Số lượng UE	$K = 8$
Số lượng anten ở BS	$M \in \{32, 64, 128\}$
Tần số tính toán tại máy chủ MEC	$f_{\text{SER}} = 5$ GHz
Tần số tính toán tại UE k	$f_k = 1$ GHz

3.2. Thuật toán đề xuất dựa trên DDPG

Phần này nhằm phát triển một thuật toán sử dụng DDPG để xác định việc phân bổ công suất và phân công nhiệm vụ offloading dữ liệu. Độ trễ min-max sau đó được xác định thông qua việc phân tích các biến đổi trong thiết kế phần thưởng (reward), không gian hành động (action space) và không gian trạng thái/ quan sát (state/observation space). Sau đây là mô tả chi tiết về những thành phần này:

- Không gian trạng thái/ quan sát: Tại mỗi bước thời gian t , một quan sát được tạo ra bằng cách sử dụng trạng thái môi trường hiện tại $s^{(t)}$, bao gồm độ trễ, tốc độ truyền dữ liệu và giải pháp đồng thời. Do đó, không gian trạng thái tại bước thời gian t có thể được xác định cụ thể như sau:

$$s^{(t)} \triangleq [p_1^{(t-1)}, \dots, p_K^{(t-1)}, \rho_1^{(t-1)}, \dots, \rho_K^{(t-1)}, (\tau_1^{\text{MEC}})^{(t-1)}, \dots, (\tau_K^{\text{MEC}})^{(t-1)}, R_1^{(t-1)}, \dots, R_K^{(t-1)}] \quad (14)$$

- Không gian hành động: Không gian hành động tại thời điểm t được xác định như sau:

$$a^{(t)} = [p_1^{(t)}, \dots, p_K^{(t)}, \rho_1^{(t)}, \dots, \rho_K^{(t)}]. \quad (15)$$

- Phần thưởng nhận được tại bước thời gian t sau khi thực hiện hành động $a^{(t)}$ trong trường hợp này là độ trễ tích lũy với

$$r^{(t)} = - \max_{k \in \mathcal{K}} \{ (\tau_k^{\text{MEC}})^{(t)} \}. \quad (16)$$

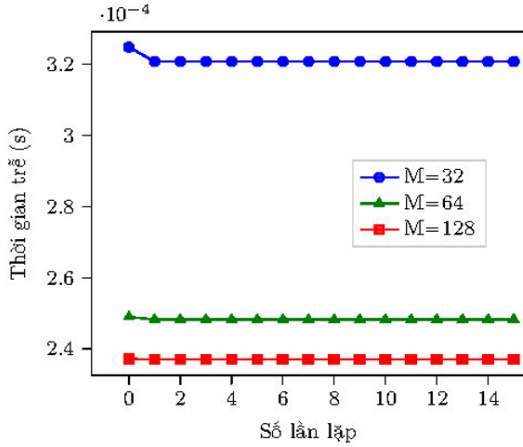
4. Kết quả số

Trong phần này, nhóm tác giả sẽ trình bày các kết quả số cho việc giải bài toán tối ưu hóa chất lượng trải nghiệm QoE đối với mạng MEC-MIMO đang xem xét. Hiệu quả của thuật toán được đánh giá thông qua mô phỏng Python với việc lấy trung bình sau khi tạo ngẫu nhiên 1000 kênh vô tuyến.

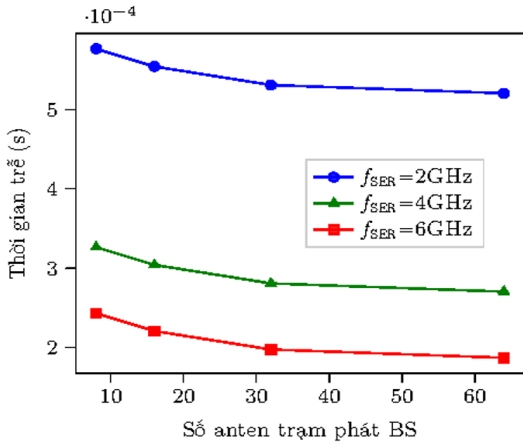
Như đã đề cập trong mục 2, nhóm tác giả xem xét một mô hình mạng với trạm BS gồm M anten truyền dữ liệu đến K UEs thiết bị kết nối đơn anten. Các UE được phân bố ngẫu nhiên trong một khu vực hình tròn với bán kính 100 mét. Công suất huấn luyện uplink được thiết lập ở mức 10 dBm cho mỗi UE. Ngoài ra, Bảng 1 liệt kê các tham số hệ thống còn lại.

Trước tiên, xem xét tốc độ hội tụ của thuật toán dựa trên DDPG qua các tập huấn luyện. Hình 1 biểu diễn thời gian trễ (tính bằng giây) theo số lần lặp, trong khi duy trì một số lượng cố định các thiết bị kết nối UE với $K = 8$. Tốc độ hội tụ được nghiên cứu trong ba trường hợp có số anten trạm BS khác nhau: $M \in \{32, 64, 128\}$. Từ Hình 1, có thể thấy, thuật toán DDPG đạt được sự hội tụ nhanh chóng chỉ sau một số lần lặp nhỏ. Hơn nữa, từ Hình 1 ta thấy, nếu sử dụng nhiều anten BS lớn hơn dẫn đến thời gian trễ ngắn hơn. Khác với khoảng thời gian trễ lý thuyết của công nghệ

5G trong khoảng từ 6-9 ms [14], thuật toán đề xuất cho ra thời gian trễ chỉ khoảng 0.32 ms cho $M = 32$; 0,25 ms cho $M = 64$, và 0,24 ms cho $M = 128$.



Hình 1. Tốc độ hội tụ của thuật toán DDPG theo thời gian trễ (s) với số lượng anten BS khác nhau



Hình 2. Thời gian trễ tối thiểu khi thay đổi số lượng anten BS

Tiếp theo, Hình 2 khảo sát hiệu suất hệ thống khi thay đổi số anten BS (M) với số lượng thiết bị kết nối UE cố định ($K = 4$). Trong hình này nhóm tác giả cũng xem xét các trường hợp tần số tính toán khác nhau tại máy chủ MEC gồm $f_{SER} \in \{2,4,6\}$ GHz. Như kỳ vọng, thời gian trễ giảm khi số anten BS tăng trong cả ba kịch bản về f_{SER} , điều này khẳng định tính đúng đắn của các kết quả đã khảo sát trong Hình 1. Hơn nữa, Hình 2 cho thấy, hệ thống có thể đạt được chất lượng trải nghiệm QoE cao hơn về mặt thời gian trễ với tần số tính toán cao hơn. Ví dụ, với $M = 32$, thời gian trễ khoảng 0,2 ms cho $f_{SER} = 6$ GHz, trong khi thời gian trễ này lần lượt là khoảng 0,5 ms cho $f_{SER} = 2$ GHz và 0,3 ms cho $f_{SER} = 4$ GHz. Do đó, các kết quả số liệu nhấn mạnh tính hiệu quả của thuật toán dựa trên phương pháp DDPG trong việc giảm thiểu thời gian trễ tính toán. Cuối cùng, có thể nhận thấy có tiềm năng lớn trong việc cải thiện hiệu suất bằng cách tăng số lượng anten BS lớn hơn và nâng tần số tính toán cao hơn tại máy chủ MEC.

5. Kết luận

Trong bài báo này, nhóm tác giả giới thiệu một mô hình tính toán toán học mới cho mạng MEC-MIMO, trong đó tính đến thời gian offloading và thời gian thực thi cục bộ tại máy chủ.

Nghiên cứu này nhằm đến mục tiêu tăng cường chất lượng trải nghiệm QoE cho các thiết bị kết nối đơn anten nằm ở biên của hệ thống thông qua việc giảm độ trễ tính toán. Để giải quyết vấn đề tối ưu hóa này, phát triển một thuật toán dựa trên DDPG xem xét các ràng buộc về tỷ lệ offloading và phân bổ công suất. Thông qua mô phỏng, đánh giá hiệu suất của thuật toán đề xuất và chứng minh tính hiệu quả của thuật toán này trong việc cải thiện QoE của UE. Kết quả tính toán số cho thấy, bằng cách tận dụng một số lượng anten trạm BS cao hơn và tần số tính toán cao hơn tại máy chủ MEC thì có thể cải thiện mạnh mẽ hiệu năng hệ thống.

Lời cảm ơn: Nghiên cứu này được hỗ trợ bởi Đề tài cấp Bộ Giáo Dục và Đào Tạo, mã số B2024.DNA.16.

TÀI LIỆU THAM KHẢO

- [1] G. Femenias and F. Riera-Palou, "Mobile Edge Computing Aided Cell-Free Massive MIMO Networks", *IEEE Transactions on Mobile Computing*, vol. 23, no. 2, pp. 1246–1261, 2024.
- [2] S. S. Yilmaz, B. O'zbek, and R. Mumtaz, "Delay Minimization for Massive MIMO Based Cooperative Mobile Edge Computing System With Secure Offloading", *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 149–161, 2023.
- [3] C. Zhao, Y. Cai, A. Liu, M. Zhao, and L. Hanzo, "Mobile Edge Computing Meets mmWave Communications: Joint Beamforming and Resource Allocation for System Delay Minimization", *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2382–2396, 2020.
- [4] M. T. P. Le, H. V. Nguyen, V. Nguyen-Duy-Nhat, and L. Sanguinetti, "QoE-Aware Power Allocation for Aerial-Relay Massive MIMO Networks", *IEEE Transactions on Network and Service Management*, vol. 21, no.1, pp. 477–489, 2024.
- [5] V. Nguyen-Duy-Nhat, M. T. P. Le, and H. Nguyen-Le, "On the optimization of weighted sum rate for MIMO broadcast channels", *The University of Danang - Journal of Science and Technology*, vol. 20, no. 6.2, pp. 44–51, 2022.
- [6] M. Zeng, W. Hao, O. A. Dobre, Z. Ding, and H. V. Poor, "Massive MIMO-Assisted Mobile Edge Computing: Exciting Possibilities for Computation Offloading", *IEEE Vehicular Technology Magazine*, vol. 15, no. 2, pp. 31–38, 2020.
- [7] M. Zeng, W. Hao, O. A. Dobre, and H. V. Poor, "Delay Minimization for Massive MIMO Assisted Mobile Edge Computing", *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6788–6792, 2020.
- [8] H. Lim and T. Hwang, "Energy-Efficient Beamforming and Resource Allocation for Multi-Antenna MEC Systems", *IEEE Access*, vol. 10, pp. 18 008–18 022, 2022.
- [9] D. Gao, H. Cheng, Z. Han, and S. Yang, "Resource Optimization for the Multi-user MIMO Systems Assisted Edge Cloud Computing", in *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*, 2021, pp. 948–953.
- [10] T. T. Nguyen, L. B. Le, and Q. Le-Trung, "Computation Offloading in MIMO Based Mobile Edge Computing Systems Under Perfect and Imperfect CSI Estimation", *IEEE Transactions on Services Computing*, vol. 14, no. 6, pp. 2011–2025, 2021.
- [11] T. Huang, Y. Zhang, H. Wu, W. Jiang, C. Yao, M. Xu, and J. Feng, "Joint Pilot and Data Transmission Power Control and Computing Resource Allocation for the Massive MIMO Based MEC Network", in *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, 2019, pp. 860–865.
- [12] W. Feng, J. Zheng, and W. Jiang, "Joint Pilot and Data Transmission Power Control and Computing Resource Allocation Algorithm for Massive MIMO-MEC Networks", *IEEE Access*, vol. 8, pp. 80 801–80 811, 2020.
- [13] C. Ding, J.-B. Wang, H. Zhang, M. Lin, and J. Wang, "Joint MU-MIMO Precoding and Resource Allocation for Mobile-Edge Computing", *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1639–1654, 2021.
- [14] E. J. Oughton, W. Lehr, K. Katsaros, I. Selinis, D. Bublely, and J. Kusuma, "Revisiting wireless internet connectivity: 5G vs Wi-Fi 6", *Telecommunications Policy*, vol. 45, no. 5, p. 102127, 2021.