Handover Optimization in LTE Networks Using Contextual Bandit Reinforcement Learning and Real-World Data

M M Sadman Shafi
Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur, Bangladesh
sadmanshafi@iut-dhaka.edu

K M Istiaque
Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur, Bangladesh
istiaque20@iut-dhaka.edu

Shafayet Sadik Sowad
Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur, Bangladesh
shafayetsadik@iut-dhaka.edu

Mohammad Tawhid Kawser
Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur, Bangladesh
kawser@iut-dhaka.edu

Abstract*— Handover (HO) optimization is crucial in Long Term Evolution (LTE) networks to ensure seamless connectivity and minimize unnecessary signaling. Conventional threshold-based HO mechanisms lack adaptability to dynamic radio conditions, leading to suboptimal decisions and ping-pong effects. This paper presents a reinforcement learning (RL)–based HO decision framework trained and evaluated on real-world LTE drive-test datasets from a dense urban environment in Dhaka, Bangladesh. The HO problem is modeled as a contextual bandit and solved using tabular Q-learning with a binary action space. The reward function integrates handover margin (HOM) and time-to-trigger (TTT) constraints to promote link stability and penalize unstable HOs. Evaluation on unseen traces uses nearest-neighbour state matching, while an external TTT filter further improves temporal stability. Results show consistent suppression of unnecessary HOs and alignment with reward objectives, confirming the framework's robustness and adaptability for practical LTE deployments.

***Keywords*— Long Term Evolution (LTE), Handover Optimization, Reinforcement Learning, Q-Learning, Drive Test Data, Time-to-Trigger (TTT), Handover Margin (HOM)**

I. INTRODUCTION

In Long Term Evolution (LTE) networks, handover (HO) optimization is critical for sustaining seamless connectivity, minimizing call drops, and ensuring consistent user experience amid fluctuating radio conditions. Poorly timed or unnecessary handovers not only degrade service quality but also increase signaling overhead and drain network resources, ultimately impacting both users and operators. Traditional rule-based HO algorithms, often relying on fixed thresholds and timers, lack the flexibility to adapt dynamically to diverse scenarios such as varying user mobility, interference levels, and network load.

Reinforcement Learning (RL) introduces a promising data-driven framework that enables adaptive decision-making by continuously learning from environmental feedback. Unlike static methods, RL agents can autonomously discover effective HO policies by balancing exploration and exploitation, thus optimizing trade-offs between handover frequency and connection stability. Q-learning, a prominent model-free RL algorithm, learns the optimal action-value function without needing explicit knowledge of network dynamics, making it well-suited for complex and time-varying wireless environments.

To realize robust RL-based HO optimization, training agents with realistic datasets is indispensable. Drive test measurements capture nuanced mobility patterns, signal quality variations, and cell topology intricacies under real operational conditions, providing high-fidelity inputs for learning algorithms. Leveraging such rich datasets ensures that learned policies generalize well across different network scenarios, enabling more intelligent, context-aware handover decisions that improve overall network efficiency and user satisfaction. Moreover, real-world measurements expose the agent to non-idealities such as measurement noise, interference spikes, and irregular mobility patterns that are often absent in simulations. This exposure equips the model with better resilience to unpredictable field conditions, enhancing deployment readiness. By grounding the learning process in empirical data, the resulting HO policies can be directly translated into practical network optimization strategies with minimal adaptation overhead.

The rest of this paper is structured as follows: Section 2 reviews related works; Section 3 details the proposed methodology; Section 4 presents evaluation and results; and Section 5 concludes the study.

II. RELATED WORKS

Recent efforts in LTE network performance evaluation have extensively leveraged drive test data to characterize wireless connectivity in complex urban environments. A comprehensive urban LTE dataset was developed capturing detailed mobility and connectivity patterns within a busy market environment, enabling deeper analysis of wireless performance [1]. Another study focused on 2G handover evaluation through RF drive test log analysis, highlighting practical coverage assessment techniques [2]. Regional LTE performance studies in Nigeria revealed operator disparities in speed, latency, coverage, and packet loss, providing recommendations to improve infrastructure in suburban and rural areas [3]. Extensive long-term drive test data was used to apply machine learning models for predicting network metrics and understanding signal interactions [4]. KPI-focused drive tests in Iraq analyzed LTE performance parameters such as RSRP, RSRQ, SINR, and throughput to guide network management [5]. A critical review of cellular network monitoring applications highlighted the importance of reliable, scalable tools for effective data-driven network optimization [6].

Handover optimization using an improved genetic algorithm dynamically adapts hysteresis tolerance for high-speed users, improving handover trigger and success rates while ensuring communication quality [7]. Load balancing self-optimization in 5G efficiently redistributes traffic from overloaded to low-load cells, significantly reducing ping-pong handover, radio link failures, and improving spectral efficiency across mobility scenarios [8]. Metaheuristic algorithms like Grey Wolf and Mayfly optimizations improve vertical handover performance in heterogeneous wireless networks by reducing energy consumption, delay, call drops, and increasing throughput [9]. A fuzzy logic-based handover scheme dynamically adjusts handover margin and time-to-trigger based on signal quality and user speed, reducing handover frequency and ping-pong effects in ultra-dense 5G networks [10]. Scalable mobility management using proximity-based clustering and aerial access networks addresses high handover rates in dense 5G environments, resolving security and privacy challenges overlooked by previous distributed schemes [11]. In [12], Alnabhana et al. propose an RSSI, mobility-direction and BS-capacity-aware handover scheme for dense femtocells, cutting unnecessary HOs by ≈30% and lowering delay by ~10 ms. The work in [13][13] introduces a GA-based vertical-handover selection (GAfVH) for multi-RAT 5G, reducing HO frequency and enabling application-specific network choice. In [14], Shayea et al. present IDHPO-AWF, which fuses SINR, load, and speed metrics via an automatic-weight function to optimize per-UE HCPs, improving RSRP, HOP and RLF. The RHOT-FLC in [15] adjusts TTT and HOM using fuzzy logic, yielding significant mobility robustness gains. A self-optimization framework in [16] tunes Threshold, Hysteresis and TTT based on channel quality, mobility and latency, lowering HO rates and ping-pong. The study in [17] highlights HCP trade-offs across user speeds, motivating adaptive settings, while [18] introduces CSVAG, a hybrid CS/GA method that boosts throughput (~17% over CS) and reduces latency and failures.

Baud Haryo Prananto et al. [19] extend RIC-based mobility prediction by incorporating UE-motion features into vector autoregression and replacing it with a neural network, improving target-cell selection in high-mobility coverage-hole scenarios. In [20], supervised ML is applied to map HOM/TTT parameters to HOPP and HOF metrics, where nested cross-validation yields lower MAE and MSE than standard validation. The ANN–fuzzy logic handover protocol in [21] jointly addresses authentication and QoS, achieving ≈27% higher success rates with notable reductions in HOF and ping-pong events. A DRL-based satellite handover scheme in [22] leverages visible time, RSS, distance and load information, reducing total handovers by over 21% under zero-failure conditions. Multi-agent RL in [23] enables distributed satellite handover coordination that balances load while lowering average handovers and user blocking probability compared to localized criteria. Q-learning in [24] dynamically selects LTE HOM/TTT pairs to minimize handover rate and latency while maximizing throughput, outperforming fuzzy-logic benchmarks. The approach in [25] integrates subtractive clustering with Q-learning to self-generate fuzzy membership functions and rules, reducing handover, ping-pong and failure rates while improving throughput and latency.

Most existing HO optimization studies rely on simulations, while real drive-test datasets are typically limited to signal quality assessment; in contrast, our work integrates real-world LTE drive-test data with a novel tabular Q-learning framework for reinforcement learning-based HO optimization.

III. Methodology

This section details the drive test, data preprocessing and reinforcement learning framework for handover optimization, including state-action-reward design with HOM and TTT, contextual bandit modeling and Q-learning training.

A. *Drive Test*

Drive Test Setup and Data Collection:

The drive test was conducted in Dhaka, Bangladesh, targeting LTE eNodeBs identified via the OpenCellID database. A high-density urban route (~13 km) from Le Meridien Dhaka (Uttara) to BRAC University was selected to induce frequent inter-cell handovers. The route included stationary, pedestrian, and vehicular movement under mixed traffic conditions, enabling diverse mobility scenarios and repeated encounters with the same cells.

Measurements were taken on the Grameenphone LTE network over three days—October 15, November 6, and November 16—during evening peak hours (16:00–22:00). Each day comprised two round trips, introducing non-uniform mobility patterns and providing temporal diversity for learning-based analysis. Network infrastructure remained unchanged across the measurement period. The map view of the drive test area is shown in figure 1.

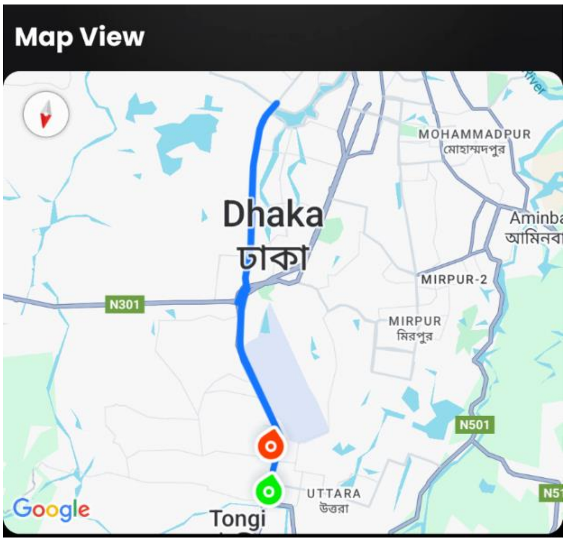

Figure 1. Map view of the Drive Test area

Equipment and Configuration:
Data collection employed XCAL-M software (Accuver) on a laptop connected to a Samsung Galaxy S10 (Exynos SoC) in Layer 3 KPIs. An Arduino Uno with Neo-7M GPS provided high-accuracy location and velocity logs. XCAL-M enabled millisecond-level sampling of RF metrics and handover signaling, synchronized with GPS coordinates and timestamps.

Captured Parameters:
From XCAL-M post-processing:

- *Handover events* — attempt counts, success counts, event types, timestamps.
- Timestamp
- Longitude, Latitude, Speed (km/h)
- Serving cell: ID, RSRP, RSRQ, CINR
- Neighbor cell: ID, RSRP, RSRQ

These parameters, aligned in time, formed the basis of the tabular dataset for reinforcement learning. The complete dataset has been made public [26].

B. *Dataset Description*

Three independent datasets were collected from the drive tests:

- Day 1 — 922 measurement points, 150 handover (HO) instances
- Day 2 — 592 measurement points, 99 HO instances
- Day 3 — 641 measurement points, 118 HO instances

Each dataset underwent a structured preprocessing pipeline: missing *RSRP* and *RSRQ* values were manually interpolated using serving cell IDs and timestamp continuity; segments affected by hardware anomalies were removed; and all non-LTE measurements were filtered out. The processed datasets were then used for reinforcement learning.

C. *Reinforcement Learning Implementation*

The reinforcement learning (RL) agent is designed with a defined state space, binary action set, and a reward function that incorporates handover margin (HOM) and time-to-trigger (TTT) parameters to optimize handover decisions based on signal quality and stability.

State
Each state is a single measurement snapshot represented by a 5-dimensional feature vector (all features standardized using the *StandardScaler* fitted on the training set):

$$s=[\textit{Serving RSRP, Serving RSRQ, Neighbour RSRP, Neighbour RSRQ, Serving CINR}]_{\text{normalized}}$$

Action
A binary action space is used:

- a=0 : No handover (stay on serving cell)
- a=1 : Trigger handover to neighbour cell

Reward Function:
The reward is a conditionally structured scalar signal that (i) rewards correct retention of a good serving link, (ii) rewards handovers that improve both signal strength and quality, and (iii) heavily penalizes unnecessary or unstable handovers. The design explicitly integrates two conventional mobility-control primitives: HOM (handover margin) and TTT (time-to-trigger).

Algorithm 1. Reward Computation for Handover Decision

Input:

- *State vector s, index i, dataset df*
- *Constants: HOM_MARGIN, TTT_SECONDS*
- *Action a ∈ {0 = No Handover, 1 = Handover}*

Output:

- *Reward R*

Steps:

1. ***Extract physical values from state:***
   - *s_RSRP ← df.Serving_Cell_RSRP[i]*
   - *n_RSRP ← df.NeighCell_RSRP[i]*
   - *s_RSRQ ← df.Serving_Cell_RSRQ(s[i])*
   - *n_RSRQ ← df.NeighCell_RSRQ(s[i])*
   - *CINR ← CINR(s[i])*
2. ***Check HOM condition:***
   - *ho_cond ← (n_RSRP ≥ s_RSRP + HOM_MARGIN)*
3. ***Check TTT persistence:***
   - *start_time ← df.timestamp_sec[i]*
   - *ttt_met ← True*
   - *j ← i*
   - *While (j+1 < len(df)):*
     - *next_time ← df.timestamp_sec[j+1]*
     - *If (next_time − start_time > TTT_SECONDS): break*
     - *If (df.NeighCell_RSRP_dBm[j+1] < df.Serving_Cell_RSRP_dBm[j+1] + HOM_MARGIN):*
     - *ttt_met ← False; break*
     - *j ← j + 1*
4. ***Compute reward:***
   - *If (a = 1)* **Handover**:
     - *If (ho_cond AND ttt_met):*
     - *If (n_RSRP > s_RSRP AND n_RSRQ > s_RSRQ AND CINR > 20): R ← 3*
     - *Else if (n_RSRP > s_RSRP AND n_RSRQ > s_RSRQ): R ← 1*
     - *Else: R ← −8*
     - *Else: R ← −20*
   - *Else (**No Handover**):*
     - *If (ho_cond AND ttt_met): R ← −5*
     - *Else if (CINR > 25 AND s_RSRP > −85): R ← 5*
     - *Else if (CINR ≤ 5 OR s_RSRP < −110): R ← −4*
     - *Else: R ← 3*
5. ***Return** R*

The reward function is designed to strongly discourage unstable or poor-quality handovers through large negative rewards (e.g., −20, −8), effectively reducing ping-pong effects and unnecessary signaling. Positive rewards (e.g., +5, +3) promote maintaining strong serving links and avoiding needless handovers. Intermediate rewards (+1, −5) capture marginal improvements or missed handover opportunities, guiding the agent towards balanced decisions that optimize network performance and stability.
In the implemented agent, HOM and TTT are parameters that gate and stabilize handover decisions.

Handover Margin (HOM):

The handover margin is defined as a fixed signal strength threshold, expressed in decibels, by which the received signal power of a neighbouring cell must exceed that of the serving cell before a handover is considered. This threshold mitigates the risk of unnecessary handovers caused by short-term measurement fluctuations or minor variations in signal quality, thereby enhancing decision stability.

Time-to-Trigger (TTT):

The time-to-trigger parameter specifies the minimum duration for which the HOM condition must remain satisfied before a handover is executed. During this interval, signal measurements are continuously monitored to ensure that the neighbouring cell maintains its advantage over the serving cell. If the advantage is lost at any point within this duration, the handover is aborted. This persistence mechanism prevents premature or oscillatory handovers and ensures that only sustained improvements in link quality lead to a cell change.

Each dataset entry is a timestamped but independent snapshot of radio conditions, with no influence on future states. Thus, the problem is modeled as a contextual bandit, aiming to learn a mapping $\pi: context \mapsto action$ that maximizes instantaneous reward. This supports per-snapshot rewards without state-transition modeling, using tabular Q-learning to estimate action utilities. HOM and TTT effects are embedded in the reward to ensure compliance with quality thresholds and temporal stability.

D. *Training Phase*

The training process employed a **Q-learning**-based reinforcement learning framework, wherein the agent incrementally updated its **Q-table** to represent the expected cumulative rewards for each state–action pair.

At each training episode, a random measurement point was sampled from the dataset to simulate the agent's perception of the environment. The current state was represented by normalized feature values, and an action was selected following an **ε-greedy** exploration–exploitation strategy. This ensured that the agent initially explored a broad range of state–action combinations before gradually converging toward exploitation of the learned policy as ε decayed.

The **Q-table** was initialized with zero values for both available actions (handover, no handover) for each state. The update of Q-values followed the canonical Q-learning rule:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\,[\,r + \gamma\ \text{max\_}a'\, Q(s',a') - Q(s,a)\,]$$

where:

- Q(s,a) is the current Q-value for state s and action a
- α is the learning rate
- r is the immediate reward
- γ is the discount factor for future rewards
- max_a′ Q(s′,a′) is the estimated value of the best action in the next state

This iterative process enabled the Q-table to gradually approximate the optimal action-value function, allowing the agent to make informed handover decisions based on accumulated experience. Table 1 presents the Q-learning parameter settings.

TABLE I: Q-learning Parameter Settings

Parameter	Value
Learning rate	0.1
Discount factor	0.9
Initial exploration rate	1
Minimum exploration rate	0.01
Exploration decay	0.995
Training episodes	10,000

IV. Evaluation and Results

A. *Evaluation Phase*

To assess true generalization in a dynamic radio environment, the trained agent has to be evaluated on an unseen dataset. Testing on the same training traces risks policy memorization and gives a misleading estimate of performance in operational conditions; instead, the agent must apply its learned mapping from instantaneous radio contexts to actions when confronted with new, time-evolving measurement sequences. Accordingly, evaluation proceeds sequentially over the test trace so that results reflect the agent's behaviour in a temporally ordered, realistic scenario.

Applying a tabular action-value function to new contexts requires matching each test-state to an entry in the learned Q-table. Exact matches are unlikely because the Q-table keys are continuous, normalized feature vectors observed during training. We therefore employ a nearest-neighbour lookup in the normalized feature space: for each test vector the Euclidean distance to every stored training-state is computed and the Q-values of the nearest training-state are used to select the action. This procedure projects dynamic test contexts onto the closest learned contexts and permits direct reuse of the tabulated policy.

Two training-testing configurations were evaluated. In **Case 1**, the agent was trained on the Day 1 dataset and tested on the Day 2 dataset, while in **Case 2**, training was performed on the Day 3 dataset with testing again on the Day 2 dataset. Figures 2 and 3 present the PCA-based projections of the Q-table states from the training phase alongside the corresponding states from the test data for each case.

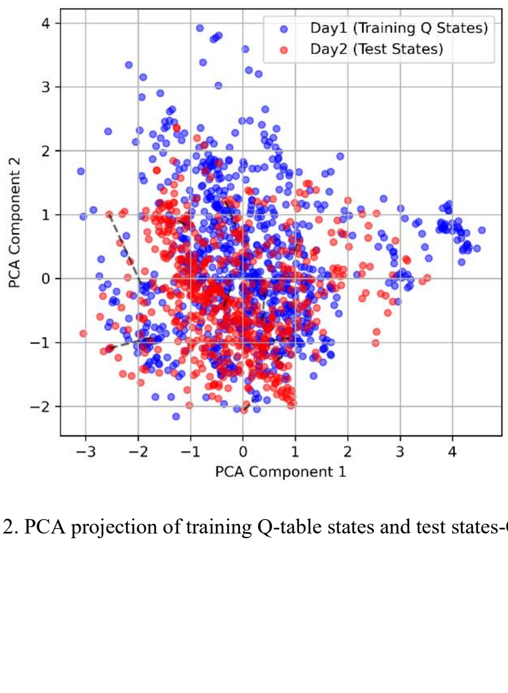

Figure 2. PCA projection of training Q-table states and test states-Case 1.

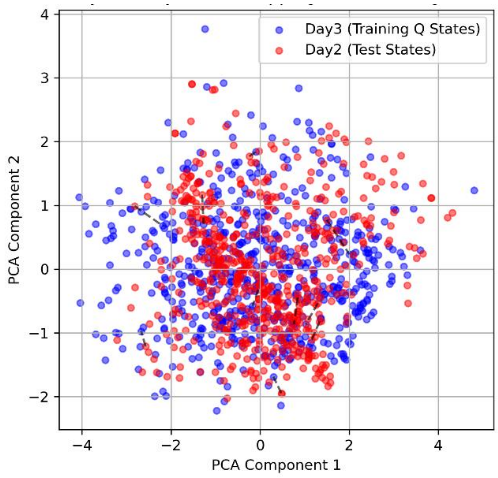

Figure 3. PCA projection of training Q-table states and test states-Case 2.

A limitation of this approach is the loss of temporal context, as nearest-neighbour matching operates on individual snapshots without considering sequential dependencies. To address this, an external Time-to-Trigger (TTT) filter is incorporated during evaluation. This post-decision constraint ensures that handover actions are only accepted if the neighbour cell advantage persists for the required TTT duration, thereby preserving the temporal stability expected in operational handover control.

B. *Result Analysis*

This subsection presents the performance evaluation of the trained RL agent under two configurations, employing a HOM of 3 dB and TTT values of 0.7 s and 1.0 s for the both of the cases. As shown in Table 2, the RL agent successfully reduces unnecessary handovers in both cases, with the optimized handover count closely aligned to the reward function's objectives.

TABLE II: RL agent's handover reduction performance across TTT settings

Case	TTT (second)	Original HO	Agent Action	TTT Blocked	Optimized HO
1	1	99	87	20	67
	0.7	99	93	21	72
2	1	99	84	18	66
	0.7	99	95	23	72

The original HO counts represent baseline handovers, while agent actions show the RL-decided handovers. The TTT blocked column indicates handovers suppressed by the agent per configured TTT, and the optimized HO reflects the effective handover count post RL filtering. The RL agent effectively reduces unnecessary handovers in both cases, achieving optimized HO aligned with the reward function for network stability and performance. Notably, TTT = 1 second results in fewer optimized handovers than TTT = 0.7 seconds due to stricter HO blocking. The consistent use of identical TTT values in training and testing phases validates the agent's adaptability and effectiveness across varying TTT configurations.

Figure 4–7 illustrate the temporal distribution of handover events for two cases under different TTT configurations, highlighting the RL agent's decision-making behavior in comparison to original handover instances and the TTT-based blocking mechanism.

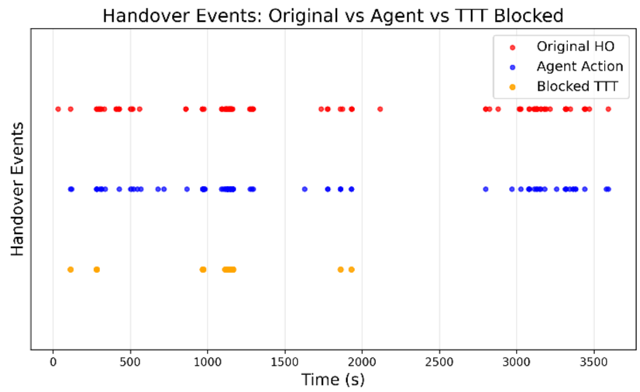

Figure 4. Temporal distribution of handover events: Case 1, TTT=1s

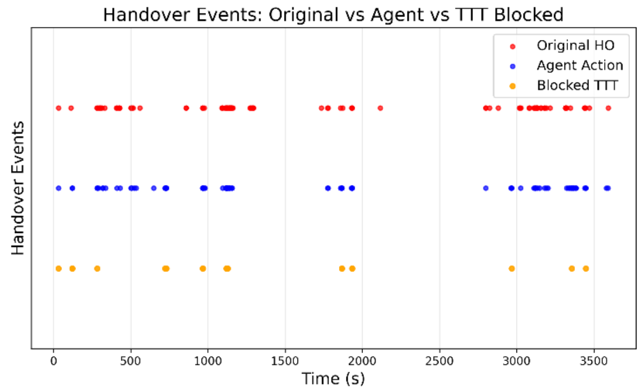

Figure 5. Temporal distribution of handover events: Case 1, TTT=0.7s

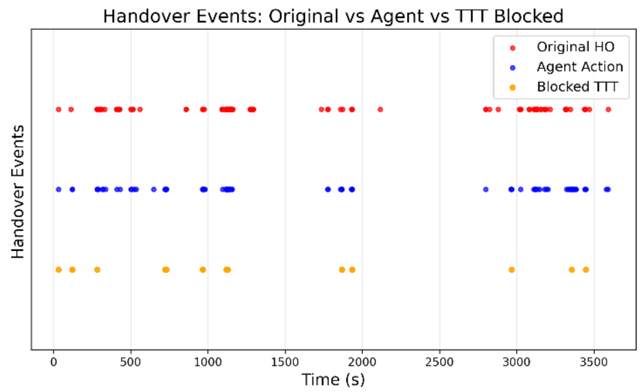

Figure 6. Temporal distribution of handover events: Case 2, TTT=1s

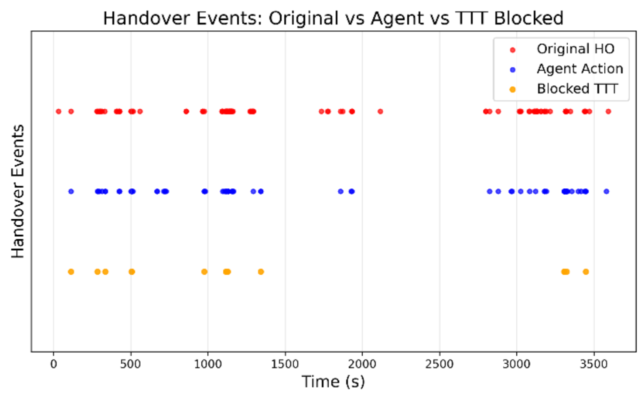

Figure 7. Temporal distribution of handover events: Case 2, TTT=0.7s

Each plot shows three event categories: original handovers, handovers accepted by the RL agent and handovers blocked by the TTT filter. The visualization shows the agent behaviour with respect to the actual handover instances and confirms that the RL agent consistently suppresses unnecessary handovers. This alignment verifies the agent's ability to incorporate external filtering (TTT) into its policy, minimizing ping-pong handovers while maintaining mobility performance. The consistent behavior across both cases and TTT settings validates the robustness and adaptability of the RL-based handover optimization framework.

V. CONCLUSION

This study set out to develop and evaluate a reinforcement learning–based handover optimization framework using real-world LTE drive-test data, with the goal of reducing unnecessary handovers. Through contextual bandit modeling, tabular Q-learning, and a reward structure integrating HOM and TTT constraints, we successfully achieved this objective, demonstrating consistent suppression of suboptimal handovers across diverse train–test configurations.

While nearest-neighbour state matching enabled effective policy generalization, its reliance on static snapshot lookup limits the capture of deeper structural relationships between states. This motivates future work on state-space abstraction—such as state or subtractive clustering to achieve more compact, topology-aware representations, enhancing robustness and scalability.

REFERENCES

[1] S. A. Garba, F. U. Ambursa, Y. S. Baguda, and M. A. Shehu, "Exploring wireless connectivity and network performance: A dataset of 4G LTE user equipment measurements," *Science World Journal*, vol. 18, no. 4, pp. 558–566, Jan. 2024, doi: 10.4314/swj.v18i4.4.

[2] K. Hamid Bilal, A. Mustafa, A. Mugdam Jaheldeen, A. A. Babiker, and N. Mustafa, "Handover Drive Test," *International Journal of Engineering, Applied and Management Sciences Paradigms*, vol. 22, 2015, [Online]. Available: www.ijeam.com

[3] S. D. Yusuf, S. I. Isa, and B. J. Kwaha, "Analysis of 4G/LTE Network Performance in North-Central Nigeria: A Comprehensive Drive Test Approach," *Journal of Engineering Research and Reports*, vol. 26, no. 9, pp. 105–122, Sep. 2024, doi: 10.9734/jerr/2024/v26i91267.

[4] S. Farthofer, M. Herlich, C. Maier, S. Pochaba, J. Lackner, and P. Dorfinger, "An Open Mobile Communications Drive Test Data Set and Its Use for Machine Learning," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1688–1701, 2022, doi: 10.1109/OJCOMS.2022.3210289.

[5] Z. Shakir, A. Y. Mjhool, A. Al-Thaedan, A. Al-Sabbagh, and R. Alsabah, "Key performance indicators analysis for 4 G-LTE cellular networks based on real measurements," *International Journal of Information Technology (Singapore)*, vol. 15, no. 3, pp. 1347–1355, Mar. 2023, doi: 10.1007/s41870-023-01210-0.

[6] S. M. Kala *et al.*, "Architecture, Performance, and Usability of Mobile Cellular Network Monitoring Applications for Data-Driven Analysis," *IEEE Access*, vol. 12, pp. 88426–88444, 2024, doi: 10.1109/ACCESS.2024.3412752.

[7] H. Zhu and Y. Peng, "Research on adaptive handover scheme based on improved genetic algorithm," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 557–562. doi: 10.1016/j.procs.2020.02.022.

[8] W. K. Saad, I. Shayea, A. Alhammadi, M. M. Sheikh, and A. A. El-Saleh, "Handover and load balancing self-optimization models in 5G mobile networks," *Engineering Science and Technology, an International Journal*, vol. 42, Jun. 2023, doi: 10.1016/j.jestch.2023.101418.

[9] M. B. Patil and L. Math, "A novel approach for optimization of handover mechanism using metaheuristics algorithms," *Measurement: Sensors*, vol. 24, Dec. 2022, doi: 10.1016/j.measen.2022.100467.

[10] W. S. Hwang, T. Y. Cheng, Y. J. Wu, and M. H. Cheng, "Adaptive Handover Decision Using Fuzzy Logic for 5G Ultra-Dense Networks," *Electronics (Switzerland)*, vol. 11, no. 20, Oct. 2022, doi: 10.3390/electronics11203278.

[11] N. Zohar, "Beyond 5G: Reducing the Handover Rate for High Mobility Communications," *Journal of Communications and Networks*, vol. 24, no. 2, pp. 154–165, Apr. 2022, doi: 10.23919/JCN.2022.000001.

[12] M. Alnabhan, E. Al-qatawneh, A. Abadleh, M. Atoum, and M. Alnawyseh, "Efficient handover approach in 5G mobile networks," *Int J Adv Sci Eng Inf Technol*, vol. 10, no. 4, pp. 1417–1422, 2020, doi: 10.18517/ijaseit.10.4.11988.

[13] I. Zubeiri, Y. El Morabit, and F. Mrabti, "Genetic algorithm for vertical handover (GAfVH) in a heterogeneous networks," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4, pp. 2534–2540, Aug. 2019, doi: 10.11591/ijece.v9i4.pp2534-2540.

[14] I. Shayea, M. Ergen, A. Azizan, M. Ismail, and Y. I. Daradkeh, "Individualistic Dynamic Handover Parameter Self-Optimization Algorithm for 5G Networks Based on Automatic Weight Function," *IEEE Access*, vol. 8, pp. 214392–214412, 2020, doi: 10.1109/ACCESS.2020.3037048.

[15] S. Alraih, R. Nordin, A. Abu-Samah, I. Shayea, N. F. Abdullah, and A. Alhammadi, "Robust Handover Optimization Technique with Fuzzy Logic Controller for Beyond 5G Mobile Networks †," *Sensors*, vol. 22, no. 16, Aug. 2022, doi: 10.3390/s22166199.

[16] A. I. Mbulwa, H. T. Yew, A. Chekima, and J. A. Dargham, "Self-Optimization of Handover Control Parameters for 5G Wireless Networks and Beyond," *IEEE Access*, vol. 12, pp. 6117–6135, 2024, doi: 10.1109/ACCESS.2023.3346039.

[17] W. K. Saad, I. Shayea, B. J. Hamza, H. Mohamad, Y. I. Daradkeh, and W. A. Jabbar, "Handover parameters optimisation techniques in 5g networks," *Sensors*, vol. 21, no. 15, Aug. 2021, doi: 10.3390/s21155202.

[18] K. Jha, A. Gupta, A. Alabdulatif, S. Tanwar, C. O. Safirescu, and T. C. Mihaltan, "CSVAG: Optimizing Vertical Handoff Using Hybrid Cuckoo Search and Genetic Algorithm-Based Approaches," *Sustainability (Switzerland)*, vol. 14, no. 14, Jul. 2022, doi: 10.3390/su14148547.

[19] B. H. Prananto, Iskandar, and A. Kurniawan, "A New Method to Improve Frequent-Handover Problem in High-Mobility Communications Using RIC and Machine Learning," *IEEE Access*, vol. 11, pp. 72281–72294, 2023, doi: 10.1109/ACCESS.2023.3294990.

[20] S. Hatipoğlu *et al.*, "Machine Learning-Based Handover Performance Prediction for Beyond 5G Communication Networks," 2024. [Online]. Available: https://www.researchgate.net/publication/382248189

[21] V. O. Nyangaresi, A. J. Rodrigues, and S. O. Abeka, "ANN-FL Secure Handover Protocol for 5G and Beyond Networks," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 99–118. doi: 10.1007/978-3-030-70572-5_7.

[22] J. Wang, W. Mu, Y. Liu, L. Guo, S. Zhang, and G. Gui, "Deep Reinforcement Learning-based Satellite Handover Scheme for Satellite Communications," in *13th International Conference on Wireless Communications and Signal Processing, WCSP 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/WCSP52459.2021.9613411.

[23] S. He, T. Wang, and S. Wang, "Load-Aware Satellite Handover Strategy Based on Multi-Agent Reinforcement Learning," in *Proceedings - IEEE Global Communications Conference, GLOBECOM*, 2020.

[24] A. Abdelmohsen, M. Abdelwahab, M. Adel, M. Saeed Darweesh, and H. Mostafa, "LTE handover parameters optimization using Q-learning technique," in *Midwest Symposium on Circuits and Systems*, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, pp. 194–197. doi: 10.1109/MWSCAS.2018.8623826.

[25] Q. Liu, C. F. Kwong, S. Wei, S. Zhou, L. Li, and P. Kar, "Reinforcement learning-based joint self-optimisation method for the fuzzy logic handover algorithm in 5G HetNets," *Neural Comput Appl*, vol. 35, no. 10, pp. 7297–7313, Apr. 2023

[26] M. M. S. Shafi, K. M. Istiaque, S. S. Sowad, and M. T. Kawser, "Drive-Test-Based LTE Handover Dataset for Cellular Mobility Studies in Urban Bangladesh," *Mendeley Data*, vol. V1, 2025, doi: 10.17632/n2pvmtyn2j.1.