

S.Zh. Aibagarov¹*, A.A. Mukhanbet¹

¹ Al-Farabi Kazakh National University, Almaty, Kazakhstan

*e-mail: awer1307dot@gmail.com

BENCHMARKING HANDOVER CONTROL IN MULTI-USER 5G NETWORKS

Abstract

Handover management is an important issue in mobile fifth-generation (5G) networks. A controller must decide when each piece of user equipment (UE) should switch its serving base station to maximize throughput while minimizing handovers, ping-pong switching, and connection failures. This paper presents a reproducible simulation environment and a systematic review of handover strategies in a multi-user scenario with four base stations and ten mobile UEs, using a Third Generation Partnership Project (3GPP) Urban Macro channel model with log-normal shadowing. We evaluate four controllers over 100 independent episodes: the 3GPP A3-event algorithm with hysteresis and time-to-trigger (TTT), an ideal Strongest-Cell rule, a No-Handover lower bound, and a learned controller trained with Proximal Policy Optimization (PPO). The results show that the standardized A3-event policy provides the best overall balance. It achieves the highest mean system throughput at 89.1 Mbit/s, the fewest handovers per episode at 21.7, and the lowest ping-pong rate at 0.098 among all methods assessed. The Strongest-Cell rule offers similar throughput but results in about 2.6 times more handovers. The learned controller shows stable training but does not reach the performance of the tuned A3 baseline based on the current reward and observation design. We discuss why the hysteresis-and-TTT logic remains a strong baseline and outline the conditions where a learned policy could outperform it. The entire environment and evaluation process are available for others to reproduce.

Keywords: 5G, handover management, mobility management, A3-event, ping-pong handover, reinforcement learning, benchmark, radio resource management.

С.Ж. Айбагаров¹, А.А. Мұханбет¹

¹ Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

КӨП ПАЙДАЛАНУШЫЛЫ 5G ЖЕЛІЛЕРІНДЕ ХЭНДОВЕРДІ БАСҚАРУДЫ САЛЫСТЫРМАЛЫ БАҒАЛАУ

Аңдатпа

Хэндоверді басқару мобильді 5G желілеріндегі негізгі мәселелердің бірі болып табылады: контроллер әрбір пайдаланушы жабдығының (UE) өткізу қабілетін барынша арттыру үшін қай уақытта қызмет көрсетуші базалық станцияны ауыстыруы керектігін анықтауы тиіс, сонымен қатар шамадан тыс хэндоверлерді, ping-pong ауысуларды және байланыс ақауларын болдырмауы қажет. Бұл жұмыста төрт базалық станция және он мобильді UE бар көп пайдаланушылы сценарий үшін лог-нормалды көлеңкеленуі бар 3GPP Urban Macro арна моделі жағдайында хэндовер стратегияларының қайта жаңғыртылатын симуляциялық ортасы және жүйелі бенчмаркі ұсынылады. Біз 100 тәуелсіз эпизод бойынша төрт контроллерді бағалаймыз: гистерезис және іске қосылу уақыты бар 3GPP A3-event алгоритмі, идеалдандырылған Strongest-Cell ережесі, No-Handover төменгі шекарасы және Proximal Policy Optimization (PPO) әдісімен оқытылған контроллер. Нәтижелер стандартталған A3-event саясаты жалпы тепе-теңдік бойынша ең жақсы көрсеткішке қол жеткізетінін көрсетеді: ол бағаланған барлық әдістер арасында ең жоғары орташа жүйелік өткізу қабілетін (89,1 Мбит/с), бір эпизодтағы ең аз хэндовер санын (21,7) және ең төмен ping-pong хэндовер деңгейін (0,098) қамтамасыз етеді. Strongest-Cell ережесі салыстырмалы өткізу қабілетіне жеткенімен, шамамен 2,6 есе көп хэндовер тудырады, ал оқытылған контроллер тұрақты оқытуға қарамастан, қазіргі сыйақы функциясы мен бақылау кеңістігі дизайны жағдайында бапталған A3 базалық стратегиясымен теңесе алмайды. Біз гистерезис және іске қосылу уақыты логикасының неліктен күшті базалық тәсіл болып қалатынын талқылаймыз және оқытылған саясат оны қандай жағдайларда жақсарту алатынын сипаттаймыз. Толық орта мен бағалау конвейері қайта жаңғыртуды қамтамасыз ету үшін жарияланған.

Түйін сөздер: 5G, хэндоверді басқару, мобильділікті басқару, A3-event, ping-pong хэндовер, нығайтумен оқыту, бенчмарк, радиоресурстарды басқару.

С.Ж. Айбагаров¹, А.А. Мұханбет¹

¹ Казахский национальный университет имени аль-Фараби, г.Алматы, Казахстан

СРАВНИТЕЛЬНАЯ ОЦЕНКА УПРАВЛЕНИЯ ХЭНДОВЕРОМ В МНОГОПОЛЬЗОВАТЕЛЬСКИХ СЕТЯХ

Аннотация

Хэндовер это головная боль любой сети 5G. Контроллеру надо угадать момент, когда абонента (UE) пора передать другой базовой станции: слишком поздно связь просядет, слишком рано и слишком часто начнутся ping-pong-переключения и обрывы. Мы собрали симуляционную среду, которую легко повторить, и прогнали в ней несколько стратегий хэндовера. Сценарий: четыре базовые станции, десять движущихся UE, канал по модели 3GPP Urban Macro с логнормальным затенением. Сравнивали четыре контроллера на 100 независимых эпизодах 3GPP A3-event с гистерезисом и временем срабатывания, идеализированный Strongest-Cell, вариант вообще без хэндоверов и обученный контроллер на PPO. Победил старый добрый A3-event. Он вытянул 89,1 Мбит/с средней пропускной способности, 21,7 хэндовера за эпизод и ping-pong-rate 0,098 лучшие цифры по всем трём метрикам сразу. Strongest-Cell по скорости почти не отстаёт, но переключается примерно в 2,6 раза чаще. А обученный контроллер, хотя учился ровно и без срывов, обойти настроенный A3 так и не смог, по крайней мере с той функцией вознаграждения и тем набором наблюдений, что были у нас. Дальше в статье разбираемся, почему связка гистерезиса и time-to-trigger оказывается такой крепкой базой и что должно измениться, чтобы обученная политика всё-таки вышла вперёд. Всю среду и конвейер оценки мы выложили можно проверить самим.

Ключевые слова: 5G, управление хэндовером, управление мобильностью, A3-event, ping-pong-хэндовер, обучение с подкреплением, бенчмарк, управление радиоресурсами.

Introduction

Key Provisions

This study presents a reproducible simulation environment for comparing handover-control strategies in a multi-user 5G network. Four approaches were evaluated: the 3GPP A3-event algorithm, Strongest-Cell, No-Handover, and a controller trained using Proximal Policy Optimization. The A3-event algorithm achieved the best overall performance, with the highest mean throughput of 89.1 Mbit/s, the lowest number of handovers among the active controllers, and a low ping-pong rate. The results demonstrate that a well-tuned standardized handover mechanism can outperform a reinforcement-learning controller when the latter has limited observations and an insufficiently optimized reward function. Mobility is one of the main characteristics of cellular networks. As a user equipment (UE) moves within the coverage area, the network has to hand over its connection from one base station (BS) to another periodically. This process, called handover, is important for continuity of service in Long-Term Evolution (LTE) and New Radio (NR) cellular systems [1], but it is also costly. Each handover causes a short interruption of the data plane, consumes signaling resources and, if triggered too aggressively, can cause the UE to oscillate between two cells, a pathology known as the ping-pong effect [2], [3]. With the densification of fifth-generation (5G) networks and the reduction of the cell size, the number of handovers increases and the trade-off between the connection quality and the handover overhead becomes more and more delicate [4].

The Third Generation Partnership Project (3GPP) specifies a family of measurement-event-triggered handovers, of which the most commonly deployed is event A3, where handover is triggered if the reference signal received power (RSRP) of a neighbour cell is higher than that of the serving cell for a specified time duration for a given hysteresis [5], [6]. Hysteresis and time-to-trigger (TTT) parameters are used to prevent unnecessary handovers due to rapid fading; however, their fixed values prevent adaptiveness to UE speed and load, and this has stimulated many works on data-driven mobility management [7]. Learning-based methods have been proposed to adapt or substitute handover logic, most notably reinforcement learning (RL) [8].

Following the success of deep reinforcement learning in high-dimensional control tasks [9], deep RL has been applied for learning handover protocol [10], predictive handover for sixth-generation (6G) networks [11], proportional fairness mobility management [12] and traffic steering in Open

Radio Access Network (O-RAN) architectures [13], [14]. Survey studies have cataloged a rapid uptake of learning-based mobility management over LTE and NR [4], and a vast range of prior work targets specific applications such as learning for sliced networks [15], A3-offset and TTT optimisation via Bayesian optimisation and deep reinforcement learning [14], learning of deep deterministic policy gradient (DDPG) parameters [16], based on the DDPG framework [17], and graph-convolutional models for vehicular scenarios [18]. Learning-based handover schemes often report a significant reduction in unwanted handovers and radio link failures [19], [20] and ongoing online learning research attempts to provide guarantees for smooth operation [21]. The vast range of learning techniques have however been evaluated using different simulators, mobility models and metrics, making direct comparison of results problematic and difficult to judge how a learned strategy fares against the default one. This literature often highlights the fundamental trade-off between two key mobility-management failure modes.

Conservative handover parameters (large hysteresis, long TTT) tend to cause high radio link failure due to excessive attachment to weakened cells, whilst overly aggressive settings increase the ping-pong rate and the overall handover overhead [7]. It is possible to navigate this trade-off with A3-event's static configuration parameters, but this leads to static mobility management. Many of the learned solutions in [6, 10, 14, 16] can be seen as attempting to learn a function mapping observations to this trade-off, to provide data-driven, adaptive mobility management. This has been demonstrated with various reported gains, though these are often highly sensitive to the training simulation, mobility model, and the choice of reward signal, with several authors noting that only time-dependent observation and a well-defined churn reward yield gains over simple baselines [10], [21].

This leads to a clear need for benchmarking where a learning approach is directly compared to the default setting on a fair basis. We frame our work around this benchmarking premise. We introduce one comprehensive, reproducible, multi-user environment, run a selection of handover schemes through this benchmark under identical conditions, and conduct a discussion of our results to explain the advantages of both standardized algorithms and learning-based approaches.

Our key contributions are threefold:

1) We propose and implement a fully reproducible, Gymnasium-compatible, multi-user handover benchmark environment, following the OpenAI Gym-style interface for reinforcement-learning environments [22], consisting of four cells, 10 mobile users, a 3GPP Urban Macro pathloss model with log-normal shadowing, and explicit handover interruption time.

2) We benchmark four different handover controllers: the standardized 3GPP A3-event algorithm, an idealized Strongest Cell (SC) algorithm, a No-Handover lower bound, and a Proximal Policy Optimization (PPO)-trained controller, using a collection of 100 independent episodes, measuring five key performance indicators: total system throughput, handover count, ping-pong rate, connection failure rate, and spectral efficiency.

3) We discuss the factors behind the continued performance of the standardized A3-event policy, and the necessary conditions for a learning controller to effectively outperform it. All associated code and trained models have been made publicly available.

Research methodology

System Model

Network and Mobility. The setup is a 5G downlink serving a 1000x1000 m² area with 4 base stations at (250, 250), (750, 250), (250, 750), and (750, 750) m. 10 UEs traverse randomly using the random-walk model with step sizes between 1-30 m, aiming for a mix of pedestrian and vehicular mobility. A single episode runs for 500 transmission time intervals (TTIs). Handovers result in 10 TTIs of downtime for the affected UE.

Channel Model

The Propagation uses the 3GPP Urban Macro (UMa) Non-line-of-sight path-loss model [23]. For a UE located distance d km away from base station, path loss is calculated as shown in Equation (1):

$$L_p(d) = 128.1 + 37.6 \log_{10}(d_{km}) \quad [\text{dB}] \quad (1)$$

The RSRP received by UE i from base station j combines path loss with a per-link log-normal shadowing term that is fixed within an episode, as defined in Equation (2):

$$\text{RSRP}_{i,j} = -L_p(d_{i,j}) + X_{i,j}, \quad X_{i,j} \sim \mathcal{N}(0, \sigma^2), \quad \sigma = 8 \text{ dB} \quad (2)$$

where the standard deviation of the shadowing is $\sigma = 8 \text{ dB}$. A UE is declared as a connection failure whenever the RSRP of its serving cell falls below the 3GPP coverage threshold of -110 dBm .

Throughput

Given a serving RSRP and an effective noise floor $N_0 = -100 \text{ dBm}$, the instantaneous Shannon capacity delivered to UE i over the system bandwidth $B = 9 \text{ MHz}$ is given by Equation (3):

$$\gamma_i = \frac{\text{RSRP}_{i,b_i} - N_0}{10}, \quad C_i = B \log_2(1 + 10^{\gamma_i}) \quad (3)$$

where b_i denotes the serving base station of UE i and γ_i is the corresponding signal-to-noise ratio in linear scale. The capacity is set to zero for any UE currently undergoing a handover. The system throughput at each step is the sum of the per-UE capacities.

Handover Strategies

We benchmark four controllers, all acting on the same per-step observation of serving and neighbor RSRP, UE speed, and time since the last handover.

A3-event (3GPP standard). Following the 3GPP measurement-event framework, a handover from the serving cell to a neighbor j is triggered when the condition in Equation (4) is satisfied:

$$\text{RSRP}_{i,j} > \text{RSRP}_{i,b_i} + H \quad \text{for } T_{TTT} \text{ consecutive steps} \quad (4)$$

holds continuously for the time-to-trigger window. We use a hysteresis margin $H = 3 \text{ dB}$ and a time-to-trigger of 40 steps, standard values that suppress fast-fading-induced spurious handovers. When several neighbors qualify, the one with the highest RSRP is selected.

Strongest Cell (idealized). At every step each UE connects to the base station with the instantaneously highest RSRP, with no hysteresis or dwell time. This is an unrealistic upper reference for signal quality that deliberately ignores handover cost.

No Handover (lower bound). Each UE remains attached to its initial cell for the entire episode. This isolates the throughput attainable with zero mobility signaling and serves as a floor.

PPO (learned). This agent trained using Proximal Policy Optimization (PPO) [7], an on-policy actor-critic algorithm, available from Stable-Baselines3 [24]. It takes as input a 50D vector, composed of 5 features for each of the 10 UE and it outputs a MultiDiscrete action determining for each UE whether to stick to the current cell or to handover to one of the other four cells. The agent is rewarded on a combined objective, that consists of the normalized throughput on the positive term, and the handover rate, the number of UEs in interruption, and low connection quality on the negative term, as specified in Equation (5):

$$r = 0.6 \hat{C} - 0.5 \eta_{HO} - 0.1 \phi_{HO} + 0.1 Q \quad (5)$$

where $\hat{C}$ is normalized system throughput, η_{HO} the handover rate, ϕ_{HO} the fraction of UEs in handover, and Q the mean connection quality. The agent is trained for one million environment steps across three random seeds.

Experimental Setup

All four controllers are evaluated in the exact same environment for 100 independent episodes of 500 steps each; episodes seeds are fixed across all methods so that every controller faces the same mobility and shadowing realizations. We record 5 metrics per method: mean system throughput (Mbit/s), number of handovers per episode, ping-pong rate (fraction of completed handovers that bring a UE back to its previous cell within 100 steps), connection failure rate (fraction of UE-steps with serving RSRP < 110 dBm) and spectral efficiency (bit/s/Hz). PPO controller is trained using a standard Stable-Baselines3 configuration; we average over 3 random seeds for the learning curve.

The environment, trained agents and evaluation scripts are publicly available. These 5 metrics are selected to describe the full mobility management cost structure as opposed to just the system throughput: system throughput and spectral efficiency reflect the utility; handover count the overhead; ping-pong rate the pathological back and forth handovers; and connection failure rate how frequently we leave a UE with a useless RSRP. A method with good system throughput and bad number of handovers or ping pong is operationally useless since every handover consumes a network resource and disrupts the user. A single value is not enough to describe the performance; the combination of these 5 metrics under fixed episode seeds is crucial to show the trade-offs between methods as illustrated in Figures 2 and 3.

Results of the study

Aggregate Comparison

Table 1 reports the mean and standard deviation of each metric over the 100 evaluation episodes. The 3GPP A3-event policy attains the highest mean throughput at 89.1 Mbit/s, exceeding the idealized Strongest-Cell rule by about 11% while triggering only 21.7 handovers per episode - roughly 2.6× fewer than Strongest Cell (56.2). A3-event also yields the lowest ping-pong rate (0.098) and the lowest connection-failure rate (0.427) among the active controllers. The No-Handover bound, by construction, incurs no handovers or ping-pong but sacrifices throughput (71.1 Mbit/s) and suffers frequent connection failures as UEs drift away from their initial cell. The PPO controller, under the present reward and observation design, does not match the tuned A3 baseline on the operational metrics.

Table 1. Handover-management benchmark over 100 evaluation episodes (mean $\pm$ std)

Method	Throughput (Mbit/s)	Handovers	Ping-pong rate	Conn. failure	Spec. eff. (b/s/Hz)
A3-event	89.1 $\pm$ 25.2	21.7 $\pm$ 4.2	0.098 $\pm$ 0.070	0.427 $\pm$ 0.092	9.90 $\pm$ 2.80
Strongest Cell	80.3 $\pm$ 23.4	56.2 $\pm$ 6.2	0.325 $\pm$ 0.063	0.402 $\pm$ 0.080	8.92 $\pm$ 2.60
No Handover	71.1 $\pm$ 24.9	0.0 $\pm$ 0.0	0.000 $\pm$ 0.000	0.542 $\pm$ 0.102	7.90 $\pm$ 2.76
PPO (learned)	33.3 $\pm$ 14.8	71.2 $\pm$ 7.0	0.346 $\pm$ 0.072	0.602 $\pm$ 0.064	3.70 $\pm$ 1.64

The throughput, along with 95% confidence intervals, for all four schemes are illustrated in Figure 1. It is visually and statistically clear, the four algorithms can be separated. The ranking, from high throughput to low, is: A3-event, Strongest Cell, No Handover and Learned Controller.

This demonstrates that the edge A3 has over Strongest Cell can be explained. Although they are both aimed at identifying the strong cells for tracking, they approach it in slightly different manner. Hysteresis in A3 is responsible to keep it within a good cell whereas the learned algorithm attempts to anticipate when best to handover between the currently perceived good cells.

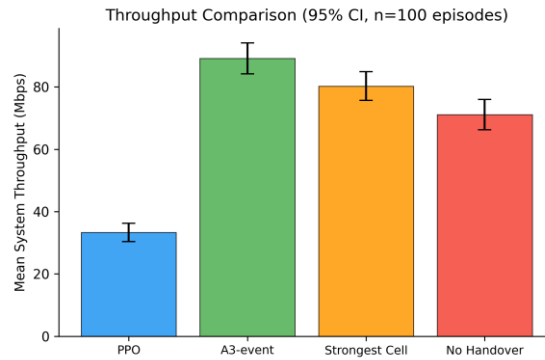


Figure 1. Mean system throughput by method, with 95% confidence intervals over 100 episodes

In Figure 2 we position each method on the throughput and handover plane, with the region we are most interested in being on the top left (good throughput, few handovers). A3-event can be seen on its own in that region, clearly showing that it has the most effective balance between operational parameters. Strongest Cell is at a similar throughput to A3-event but at much higher handover, with No Handover lying on the axis at the lower handover but with considerably lower throughput.

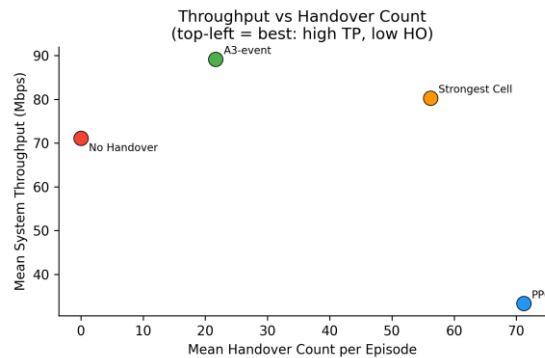


Figure 2. Throughput versus handover count. The upper-left region denotes the best trade-off

Figure 3 compares ping-pong rates. A3-event keeps ping-pong low (0.098) thanks to its dwell-time logic, whereas Strongest Cell, which reacts to every RSRP fluctuation, exhibits a high ping-pong rate (0.325) despite competitive throughput. No Handover trivially has zero ping-pong. This metric isolates the practical value of the hysteresis-and-TTT mechanism that the A3 event encodes.

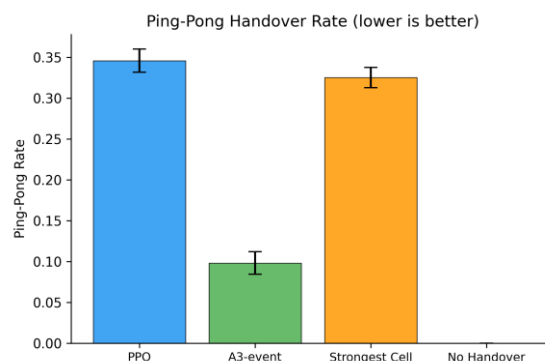


Figure 3. Ping-pong handover rate by method (lower is better)

On the Learned Controller

Figure 4 shows the PPO training curve averaged over three seeds. The mean episode reward rises from roughly -44 to about +23 over one million steps, confirming that the agent does learn to improve its own reward signal and that the poor operational scores are not an artifact of undertraining. The

gap between rising reward and weak operational metrics points instead to the difficulty of the control problem as posed: with a 50-dimensional observation, a joint MultiDiscrete action over ten UEs, and a handover that is penalized both directly and through its interruption cost, the agent converges to a conservative-yet-churning policy that neither matches A3's selectivity nor its timing. This is a known challenge in learning-based mobility management - the reward must be carefully shaped, and partial observability of fast fading limits what an instantaneous policy can achieve [10], [12]. The honest reading of these results is that a well-tuned standardized baseline is hard to beat, and that surpassing it requires either richer temporal observations, a factored or per-UE action design, or explicit reward shaping around the ping-pong and interruption costs.

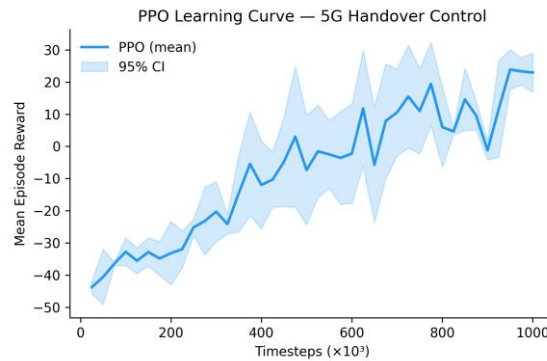


Figure 4. PPO training reward versus timesteps, averaged over three seeds (shaded area: 95% confidence interval)

Discussion

The central, well-supported finding of this benchmark is that the 3GPP A3-event algorithm provides the best throughput-handover balance of the strategies evaluated, combining the highest throughput with the fewest handovers and the lowest ping-pong rate. The comparison with Strongest Cell quantifies precisely what the A3 hysteresis-and-TTT logic buys: nearly the same signal tracking at a fraction of the handover and ping-pong cost. The No-Handover bound shows that some mobility signaling is necessary - staying put degrades both throughput and connection reliability as UEs move. Finally, the learned controller serves as a cautionary data point: reinforcement learning is not automatically superior to a tuned standard, and a fair benchmark must include such a baseline before claiming improvement. These observations align with recent literature reporting that the benefit of learned mobility management depends strongly on observation design and reward engineering [10], [14], [16].

Conclusion

We provided a reproducible benchmark of handover-management strategies for a multi-user 5G network - the 3GPP A3-event algorithm, an idealized Strongest-Cell rule, a No-Handover lower bound, and a PPO-trained controller - in one Gymnasium environment under a 3GPP Urban Macro channel model. The standard A3-event policy achieves the best compromise across 100 episodes of evaluation, exhibiting the best throughput (89.1 Mbit/s), the fewest number of handovers (21.7 per episode) and the smallest rate of ping-pong handovers (0.098). The benchmark reveals the advantages of hysteresis-and-TTT logic and furnishes a realistic baseline on which a learned controller could be evaluated. The environment will be extended with interference-aware signal-to-interference-plus-noise ratio (SINR), realistic correlated mobility traces, more flexible learned controllers (with richer state spaces, factorized actions, and penalties on interruption and ping-pong), to study under which conditions a controller learned policy could overcome the performance of the A3 policy.

Acknowledgment

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No.BR24993211).

References

- [1] S. Sesia, I. Toufik, and M. Baker, *LTE The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. Wiley, 2011. URL: <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470978504>.
- [2] M. Tayyab, X. Gelabert, and R. Jäntti, "A survey on handover management: From LTE to NR," *IEEE Access*, vol. 7, pp. 118907-118930, 2019. DOI: 10.1109/ACCESS.2019.2937405.
- [3] D. López-Pérez, I. Güvenç, and X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," *IEEE Communications Magazine*, vol. 50, no. 12, pp. 70-78, 2012. DOI: 10.1109/MCOM.2012.6384454.
- [4] M. Mollé et al., "A survey of machine learning applications to handover management in 5G and beyond," *IEEE Access*, vol. 9, pp. 45770-45802, 2021. DOI: 10.1109/ACCESS.2021.3067503.
- [5] 3GPP, "NR; Overall description; Stage-2," 3GPP TS 38.300, V18.0.0, 2024. URL: https://www.3gpp.org/ftp/Specs/archive/38_series/38.300/.
- [6] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description," 3GPP TS 36.300, V17.0.0, 2022. URL: https://www.etsi.org/deliver/etsi_ts/136300_136399/136300/17.00.00_60/ts_136300v170000p.pdf.
- [7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017. DOI: 10.48550/arXiv.1707.06347.
- [8] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018. URL: <https://mitpress.mit.edu/9780262039246/reinforcement-learning/>.
- [9] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, 2015. DOI: 10.1038/nature14236.
- [10] C. Lee and J. Schulman et al., "A deep reinforcement learning-based approach for adaptive handover protocols in mobile networks," *arXiv preprint arXiv:2401.14823*, 2024. DOI: 10.48550/arXiv.2401.14823.
- [11] I. Panitsas, et al., "Predictive handover strategy in 6G and beyond: A deep and transfer learning approach," in *Proc. IEEE GLOBECOM*, 2024. URL: <https://arxiv.org/abs/2404.08113>.
- [12] A. Prado, et al., "Enabling proportionally-fair mobility management with reinforcement learning in 5G networks," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 6, pp. 1681-1696, 2023. DOI: 10.1109/JSAC.2023.3273705.
- [13] A. Lacava, et al., "Programmable and customized intelligence for traffic steering in 5G networks using Open RAN architectures," *IEEE Transactions on Mobile Computing*, vol. 23, no. 4, pp. 2882-2897, 2024. DOI: 10.1109/TMC.2023.3266642.
- [14] Intelligent handover management enabled by O-RAN and deep reinforcement learning, *ResearchGate preprint*, 2024. DOI: 10.1109/VTC2024-FALL63153.2024.10757909.
- [15] Handover management in a sliced 5G network using deep reinforcement learning, *ResearchGate preprint*, 2024. DOI: 10.1109/IWCMC61514.2024.10592506.
- [16] Advanced handover optimization using deep reinforcement learning in 5G networks, *Journal of King Saud University Computer and Information Sciences*, 2025. DOI: 10.1007/s44443-025-00124-0.
- [17] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. on Learning Representations (ICLR)*, 2016. URL: <https://arxiv.org/abs/1509.02971>.
- [18] GCN-based throughput-oriented handover management in dense 5G vehicular networks, *arXiv preprint arXiv:2505.04894*, 2025. DOI: 10.48550/arXiv.2505.04894.
- [19] Hybrid artificial intelligence scheme for vertical handover in heterogeneous networks, in *Proc. 8th Int. Conf. on Graphics and Signal Processing (ICGSP)*, 2024. DOI: 10.1145/3694875.3694884.
- [20] Predictive handover mechanism for seamless mobility in 5G and beyond networks, *ResearchGate preprint*, 2025. DOI: 10.1049/cmu2.12878.
- [21] Smooth handovers via smoothed online learning, *arXiv preprint arXiv:2501.08099*, 2025. DOI: 10.48550/arXiv.2501.08099.
- [22] G. Brockman et al., "OpenAI Gym," *arXiv preprint arXiv:1606.01540*, 2016. DOI: 10.48550/arXiv.1606.01540.
- [23] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP TR 38.901, V17.0.0, 2022. URL: https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/.
- [24] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-Baselines3: Reliable reinforcement learning implementations," *Journal of Machine Learning Research*, vol. 22, no. 268, pp. 1-8, 2021. URL: <https://jmlr.org/papers/v22/20-1364.html>.