Holistic Mobility Management leveraging Risk Averse Reinforcement Learning

Muhammad Umar Bin Farooq*, Shahrukh Khan Kasi*, Marvin Manalastas*, Chunhui Zhu†,
Baoling Sheen†, and Ali Imran‡*
*AI4Networks Research Center, School of Electrical and Computer Engineering, University of Oklahoma, USA.
†Futurewei Technologies, USA.
‡James Watt School of Engineering, University of Glasgow, UK.
Email: {umar.farooq, shahrukhkhankasi, marvin, ali.imran}@ou.edu, {czhu, bsheen}@futurewei.com

***Abstract*—The trend towards denser base station deployment and multi-band operations in emerging cellular networks has made mobility management and handover (HO) optimization a formidable challenge. The challenge is further aggravated by the scarcity of practical multi-objective mobility management solutions optimizing both intra and inter frequency HO. This paper presents a holistic multi-objective mobility management solution for both intra and inter frequency HO employing multiple parameters of standardized HO events A2, A3, and A5. We formulate a multi-objective optimization problem to determine the optimal parameter settings that jointly optimize four key performance indicators: number of HO failures, HO latency, signaling overhead and number of radio link failures. We leverage soft actor-critic reinforcement learning (RL) to solve the multi-objective problem. To mitigate the risk of performance deterioration resulting from direct interactions between live network and RL-agent during training, this paper proposes a mobility management framework that develops and employs a digital twin (DT) as the training environment. To develop a cellular network DT for mobility management and HO optimization, we present a tri-pronged approach including realistic network deployment, realistic user mobility and 3GPP HO events. Results show that the proposed DT-trained RL solution for the multi-objective optimization can converge 7x faster than the brute force method with negligible loss in the value of the objective function. An analysis of the individual KPI values reveal a strong trade-off between HO signaling overhead and radio link failures.**

I. Introduction

Dense base station (BS) deployment and operations on motley of frequency bands have emerged as some of the most prominent solutions to meet the exploding demands for higher data rates and reliability, lower latency, support for a variety of vertical use cases, and highly dense connected devices. The shift towards denser and multi-band BS deployment is evident from the higher number of operating bands including mmWave bands in 5G compared to 4G networks. This trend is expected to continue for 6G with the utilization of THz band [1]. However, as network deployment grows more dense and operating bands expand, mobility management becomes increasingly challenging due to the associated increase in handovers and signaling overhead.

Current industry practice for HO optimization utilizes either vendor-defined gold standard based configuration and optimization parameter (COP) settings or manual knowledge-based COP settings. However, the suitability of these methods is undermined by rapidly varying network conditions, disparate user equipment (UE) mobility and requirements. To introduce some degree of autonomy in HO parameter setting, 3GPP has standardized mobility robustness optimization (MRO) under self organising networks. MRO solutions typically optimize a limited number of HO-related COPs leveraging historical data. However, the reactive nature of the majority of MRO solutions and the complex inter-connection between COPs render them inadequate for emerging cellular networks, which require a proactive and fully automated mobility management solution.

Recent application of machine learning (ML) algorithms for cellular networks have demonstrated their ability to develop proactive and automated solutions [2]. The capability of these ML models to map the impact of varying HO COPs on key performance indicators (KPI) that is not attainable with domain knowledge or even SON solutions, positions them as viable proactive mobility management enablers. Reinforcement learning is one promising approach under the umbrella of ML-based mobility management solutions. During the training phase, the RL agent explores the solution space by setting different COP values as actions on a cellular network environment in order to evaluate the reward function and learn the behavior of a cellular network with different COP combinations. While a well-trained RL agent can proactively identify COP settings that maximize the KPIs, the training phase requires an iterative, hit-or-miss approach. This idiosyncrasy of RL continues to be an impediment to its practicality, since the iterative hit-or-miss method when done in a live network raises the possibility of causing disruptions. In addition, measuring the impact of particular COP settings on a live network requires either drive tests or the exploitation of minimization of drive test (MDT) data. However, drive tests and MDT-based network measurements are both time-intensive and can significantly increase the training duration of RL. These challenges necessitate an innovative approach that can expand the applicability of RL in mobility management.

A. Related Work

Recent literature on mobility management focuses primarily on two broad approaches. The first approach involves proposing novel HO approaches [3]–[5] while the second involves optimizing intra and inter HO related parameters of existing 3GPP procedures [6]–[9]. Authors in [3] developed a novel two-stage handover procedure with first stage for UE clustering and second stage for learning the optimal handover controller using RL. Meanwhile, authors in [4] proposed a new centralized RL-based HO procedure that evaluates the UE's measurement report and maximizes the throughput gain opportunistically. In [5], the authors utilized RL to learn the best backup BS for HO and reduce the number of handovers without impacting the rate and reliability. While these studies show promising outcomes, the suggested solutions necessitate modification in existing handover standards, hindering a swift industrial uptake.

While previous studies proposed new HO methods, authors in [6] presented an RL-based dynamic HO optimization to simultaneously minimize HO failures (HOF) and ping pongs using time to trigger (TTT) and hysteresis of event A3. In contrast to tuning A3 parameters, authors in [7] proposed XGBoost-assisted genetic algorithm for event A5 based inter-frequency HO to optimize average reference signal received power (RSRP), average signal to interference and noise ratio (SINR) and HO success rate. Meanwhile, a method to predict inter-frequency HO failures and a proactive power-tuning algorithm to enhance HO success rate is proposed in [8]. Rather than optimizing either intra or inter frequency HO in silos, the authors in [9] proposed ML-aided simulated annealing solution for the joint optimization of intra and inter frequency HO.

The aforementioned studies have focused on optimizing either intra-frequency HO using event A3 or inter-frequency HO using event A5. However, optimization of event A3 and A5 HO parameters separately often leads to sub-optimal settings, as noted in [9]. Recognizing the interconnected nature of these 3GPP standardized HO events, we take a step further by incorporating event A2 alongside events A3 and A5. This approach enables holistic mobility management, simultaneously optimizing both intra and inter-frequency HOs. In contrast to prior research that typically optimizes a limited number of KPIs, we formulate a multi-objective problem aimed at minimizing the number of HOF, HO latency, signaling overhead, and number of radio link failures (RLF). To address the multi-objective problem, we propose a risk averse mobility management framework that utilizes RL to determine the optimal values of COPs that minimize the four KPIs. Since RL learns by repetitive process, training the RL agent directly on a live network can lead to potential network impediments, which inhibits the practical application of RL-based solutions. To overcome this limitation, we leverage digital twin of the cellular network as the training environment of the RL agent in our framework, rather than the actual network.

B. Contributions

The main contributions of this paper are given below:

1) We present a mobility management solution that holistically optimizes both intra and inter frequency HO. The proposed solution leverages five mobility COPs namely A2 threshold, A3 TTT, A3 offset, A5 TTT and A5 delta to optimize the KPIs. To design the new feature A5 delta, we employ domain knowledge and exploit the relationship between the parameters of event A5 and event A2. To the best of authors' knowledge, this is the first study to jointly optimize events A2, A3, and A5.
2) We formulate and solve a multi-objective optimization problem that jointly minimizes number of HOF, HO latency, signaling overhead and number of RLF as a function of the five COPs. To solve the multi-objective problem, we propose a risk averse mobility management framework that utilizes soft actor-critic RL algorithm.
3) To mitigate the risk of potential impairment when training RL on a live network, the risk averse framework leverages a DT of the cellular network to train the RL agent instead of training on the live network. We also highlight a three step road map to construct a DT for creating, testing, and optimizing mobility management solutions. Results indicate that the proposed DT-trained RL solution can converge 7 times faster than the brute force approach, making it suitable for rapidly changing network conditions and UE dynamics.

The rest of the paper is organized as follows: Section II presents the system model and problem formulation. In Section III, we presents the DT-aided risk averse mobility management framework, tri-pronged approach to create digital twin for developing mobility solutions and the performance evaluation of the RL algorithm. Finally, Section IV concludes the paper.

II. System Model

In this section, we present the standardized 3GPP HO events, define the four KPIs, the justification for concurrent optimization of events A2, A3, and A5, and the need for multi-objective optimization. This section also includes the multi-objective optimization problem formulation.

A. Standardized 3GPP Handover Events

The following discussion describes the 3GPP standardized HO events A3, A5 and A2 for 5G NR.

1) Event A3: Event A3-based HO starts when the RSRP of a UE from target BS exceeds the RSRP from serving BS by an offset for a certain time called time-to-trigger ($A3_{TTT}$).

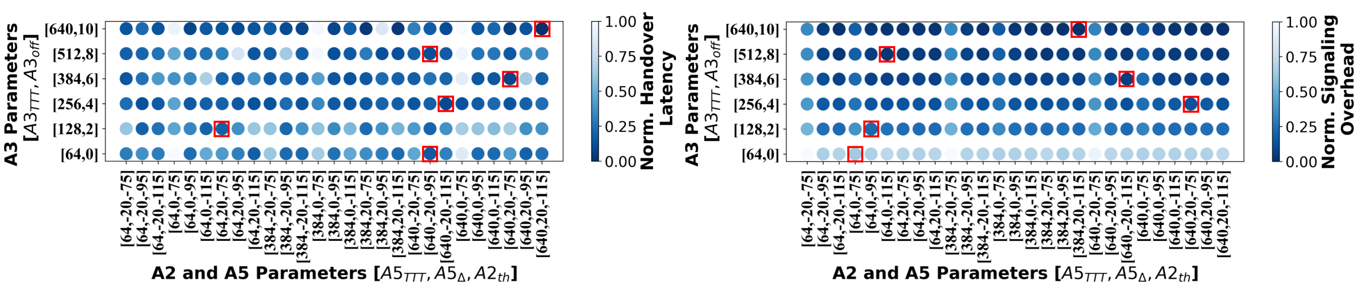
$$\eta_u^t - A3_{hyst} > \eta_u^s + A3_{off} \tag{1}$$

where η_u^t and η_u^s are the RSRP of the UE u from target BS t and serving BS s, respectively, $A3_{off}$ is the A3 offset and $A3_{hyst}$ represents A3 hysteresis.

2) Event A5: Event A5-based HO triggers when RSRP from serving BS remains below a threshold called A5-threshold1, and the RSRP from target BS remains above another threshold called A5-threshold2 for a duration of $A5_{TTT}$.

$$\begin{aligned} \eta_u^s + A5_{hyst} < A5_{th1} \\ \eta_u^t - A5_{hyst} > A5_{th2} \end{aligned} \tag{2}$$

where $A5_{hyst}$, $A5_{th1}$ and $A5_{th2}$ represent the hysteresis, threshold1 and threshold2, respectively for event A5.



(a) Normalized average HO latency

(b) Normalized total signaling overhead

Fig. 1. The variation in HO latency and signaling overhead with change in A2, A3 and A5 parameters. The red square in each row highlights the optimal KPI value for fixed A3 settings.

3) Event A2: : Event A2 is triggered when the RSRP of a UE from serving BS remains below a threshold for the duration set by $A2_{TTT}$.

$$\eta_u^s + A2_{hyst} < A2_{th} \tag{3}$$

where $A2_{hyst}$ and $A2_{th}$ are the event A2 hysteresis and threshold, respectively.

We employ event A3 and event A5 to trigger intra and inter frequency HO, respectively in accordance with the practice of major network operators [7]. Before activating an inter-frequency HO, a UE must measure the signal conditions on frequency bands other than the current operating frequency band through a 3GPP-standardized process of measurement gap (MG). We utilize event A2 to activate MG.

B. Key Performance Indicators

1) Handover Failure: The number of HOF provides a direct measurement of the HO performance. Low number of HOF indicates that the transition of UEs from one cell to another is smooth, resulting in a satisfactory quality of experience (QoE). The total number of HOF, denoted by H, is the sum of both intra and inter frequency HOF.

2) Handover Latency: HO latency represents the time a UE spends in the HO process and larger HO latency can negatively impact UE QoE. Each HO attempt will either be a success or a failure. In the event of HO success, the HO latency is measured as the time between the HO start point and the point of successful connection with the target BS. In the event that HO fails, the UE reattempts HO after a predetermined amount of time, as specified by the report interval parameter. Multiple HO failures can exacerbate the UE SINR, resulting in RLF. Hence, HO latency in the event of HO failure is the time between the HO start point and either HO success in one of the repeated attempts or UE RLF, whichever occurs first. In this paper, we optimize the average HO latency L and define it as follows:

$$L = \frac{\sum\limits_{\forall s \in \mathbb{H}_s} L_s + \sum\limits_{\forall f \in \mathbb{H}_f} L_f}{|\mathbb{H}_A|} \tag{4}$$

where L_s and L_f represent the latency for each HO success and HO failure, respectively. The sets $\mathbb{H}_s$ and $\mathbb{H}_f$ contain all the HO successes and failures, respectively and set $\mathbb{H}_A$ contains all the handovers in the network.

3) Radio Link Failure: The number of RLFs in a network can be used to measure the instances of service disruption. Improper HO parameter settings can lead to a higher number of RLF in the network, severely impacting the UE experience and leading to churn. Several timers and indicators are involved in declaring RLF. Specifically, we consider N310, T310 and N311 as the parameters for declaring an RLF. The total number of RLF in the network is denoted by R.

4) Handover Signaling: One consequence of HO is the added signaling overhead on the network. Although optimizing number of HOF, HO latency, and the number of RLF will enhance UE QoE, these factors do not directly represent the burden caused by HO to the network. In order to assess the tradeoff between UE QoE and signaling overhead as a result of HO, we incorporate signaling overhead as a KPI. We have modeled the overhead of several over-the-air signaling messages between the UE and source or target BS during the HO process using X2 interface [10]. We define total HO signaling overhead S as the additional signaling bytes transmitted over-the-air during the HO process.

C. Impact of Handover COPs on KPIs

Fig. 1 highlights the impact of handover COPs on HO latency and HO signaling using data from the digital twin setup described in sub-section III-A. Fig. 1(a) shows that the optimal value of HO latency in each row, denoted by red square, shifts as A3 settings are adjusted. This suggests that the optimal values of A2 and A5 COPs become sub-optimal if A3 COP values are altered. This trend is also apparent in Fig. 1(b) for total signaling overhead, with optimal settings shifting across each row. This observation demonstrates that optimization of A2, A3, and A5 parameters in silos, as observed in the majority of academic literature and industrial norms, may result in sub-optimal KPI values. Moreover, simultaneous analysis of Fig. 1(a) and Fig. 1(b) reveals that different sets of COPs optimize HO latency and signaling overhead. This observation implies that there is a trade-off between optimizing these KPIs, which necessitates a multi-objective KPI optimization.

D. Problem Formulation

We use the parameters from HO events A3, event A5 and event A2 to jointly optimize H, L, R, and S. We leverage domain knowledge to combine $A5_{th1}$ and $A5_{th2}$ into a new parameter $A5_{\Delta}$ and define it as $A5_{\Delta} = A5_{th1} - A5_{th2}$.

$$
\begin{array}{ll}
\min\limits_{A3_{TTT}, A3_{off}, A5_{TTT}, A5_{\Delta}, A2_{th}} & \sqrt{\alpha(\overline{H}-\overline{H_t})^2+\beta(\overline{L}-\overline{L_t})^2+\gamma(\overline{S}-\overline{S_t})^2+(1-\alpha-\beta-\gamma)(\overline{R}-\overline{R_t})^2}; \\
\text{subject to} & A5_{TTT}, A3_{TTT} \in T \\
& O_{min} \leq A3_{off} \leq O_{max} \\
& T_{min} \leq A2_{th1} \leq T_{max} \\
& A5_{th1} = A2_{th} \\
& A5_{th2} = A5_{th1} + A5_{\Delta} \\
& \alpha+\beta+\gamma \leq 1
\end{array}
\tag{5}
$$

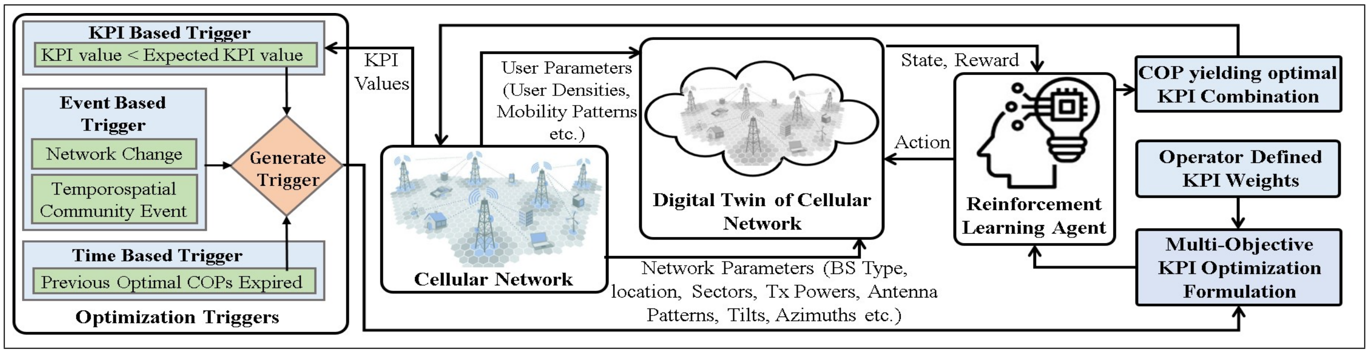

KPI Based Trigger
KPI value < Expected KPI value
Event Based Trigger
Network Change
Temporospatial Community Event
Generate Trigger
Time Based Trigger
Previous Optimal COPs Expired
Optimization Triggers
KPI Values
User Parameters (User Densities, Mobility Patterns etc.)
Cellular Network
Digital Twin of Cellular Network
State, Reward
Action
Reinforcement Learning Agent
COP yielding optimal KPI Combination
Operator Defined KPI Weights
Multi-Objective KPI Optimization Formulation
Network Parameters (BS Type, location, Sectors, Tx Powers, Antenna Patterns, Tilts, Azimuths etc.)

Fig. 2. The proposed risk averse mobility management framework with digital twin-based environment of reinforcement learning.

Moreover, we are aware that the value of $A2_{th}$ is generally equal to or higher than $A5_{th1}$. This happens because a UE cannot measure other frequency bands until event A2 is initiated, and hence, inter-frequency HO utilizing event A5 cannot be triggered prior to triggering event A2. We leverage this fact to set $A5_{th1} = A2_{th}$ and $A5_{th2} = A5_{th1} + A5_{\Delta}$. This intelligent domain knowledge aware settings of parameters allows the merging of three COPs ($A5_{th1}$, $A5_{th2}$ and $A2_{th}$) into two COPs ($A5_{\Delta}$ and $A2_{th}$), hence reducing the search space of the optimization problem.

Eq. (5) shows the multi-objective optimization problem formulation for joint minimization of H, L, R and S as a function of A3 parameters ($A3_{TTT}$ and $A3_{off}$), A2 parameter $A2_{th}$, A5 parameter ($A5_{TTT}$) and $A5_{\Delta}$ defined earlier. We have formulated the objective as a target minimization problem to minimize the difference of each KPI with a target KPI value defined by the operator. The parameters α, β, γ and $(1-\alpha-\beta-\gamma)$ are the operator-defined weights of H, L, S and R, respectively, and can be leveraged to set the importance of each KPI. Meanwhile, H_t, L_t, S_t and R_t are the normalized target values for H, L, S and R, respectively and the overline on KPI represents the normalized value of that KPI. The normalization eliminates the bias towards larger KPI values and ensures that the KPI weights control the importance of each KPI. The first three constraints confine the values of the optimization variables to the 3GPP-defined ranges. The set T contains the range of values of $A3_{TTT}$ and $A5_{TTT}$, O_{min} and O_{max} are the minimum and maximum values of $A3_{off}$, respectively, while T_{min} and T_{max} are the minimum and maximum values of $A2_{th}$, respectively. The fourth and fifth constraints characterize the intelligent domain-knowledge aware relationship between $A5_{th1}$, $A5_{th2}$, $A2_{th}$ and newly proposed $A5_{\Delta}$, respectively. Finally, the last constraint ensures that the sum of the four weights equals 1.

III. Digital Twin-Aided Risk Averse Mobility Management Framework

Training an RL agent directly on a live cellular network carries the risk of substantial KPI degradation and hence, the cellular network operators are reluctant to adopt RL-based solutions. The proposed framework in Fig. 2 aims to address this challenge by first creating a digital twin representation of the cellular network. During training, the RL model will utilize a DT as the environment instead of the actual cellular network. Only after the RL agent has been trained on the DT, it will be deployed in the actual cellular network. The network deployment parameters, geographical conditions, UE parameters, as well as the standardized cellular network processes can be utilized to create a DT of the cellular network as presented in sub-section III-A. There can be three major triggers for the optimization process: KPI-based, event-based, and time-based triggers as highlighted in 2. The multi-objective KPI optimization process commences upon the occurrence of any of these three triggers.

A. Digital Twin Creation for Mobility Management

To realistically mimic the conditions of a cellular network in a DT, we focus on accurately modeling three crucial aspects of cellular networks, which include network deployment, UE mobility patterns, and 3GPP-compliant HO events. Although additional considerations are necessary to develop a complete DT of cellular network, the aforementioned three steps can serve as a good starting point, particularly for mobility management solutions.

1) Network Deployment: The first step entails acquiring the network deployment parameters for the geographical area of interest. These deployment parameters are readily accessible to network providers and include, but are not limited to, BS location, BS type (macro or small), sectors of each BS, BS

TABLE I
NETWORK DEPLOYMENT PARAMETERS

Parameter Description	**Value**
Simulation Area	938m×697m
Number of Sites	Macro Cells (MC): 2; Small Cells (SC): 7
Cell Sectors	MC: Tri-sectored; SC: Omni-directional
Transmission Frequency	MC: 870 MHz; SC: 3300 MHz
Transmission Bandwidth	MC: 10 MHz; SC: 20 MHz
Beamforming Model	MC: 64T64R 90° 24dBi Low & Mid-bands; SC: 32T32R 360° 17 dBi Mid-band
Pathloss Model	Aster Propagation (Ray-tracing)
Geographic Information	Ground and building heights, land use maps
Shadowing	Clutter-dependent shadowing
Total Active Users	60

TABLE II
ACTIONS PERFORMED BY THE REINFORCEMENT LEARNING AGENT

COPs	**Values**
$A5_{TTT}$	[64, 128, 192, 256, 384, 512, 640] ms
$A2_{th}$	[-75 to -115] dBm
$A5_{\Delta}$ $(A5_{th1} - A5_{th2})$	[-20 to 20] dBm
$A3_{TTT}$	[64, 128, 192, 256, 384, 512, 640] ms
$A3_{off}$	[0 to 10] dB

height, antenna patterns, tilts, azimuths, MIMO configurations, transmit power, operating frequency bands, etc. Once these characteristics are known for a certain region, RSRP maps can be generated using the 3GPP-standardized MDT report data from that region [11]. In addition, the bandwidth of different frequency bands and the scheduling algorithms employed by operators can be incorporated into the DT.

As these deployment parameters and MDT RSRP maps are not publicly available, we use a popular tool among network operators for radio access network planning, and optimization [12]. This tool employs an advanced ray-tracing approach to accurately model signal propagation. Using this tool, we create a 5G network comprised of macro and small cells to record MDT-based RSRP traces in a geographic area in Manhattan, New York, USA. A log-normal distribution is used to model the shadowing with a standard deviation that varies depending on the type of clutter. We utilize real 3D antenna patterns for both type of cells. Finally, the site deployment (i.e., base station location, tilt, and azimuth) is optimized using the tool's automatic cell planning tool (ACP) feature combined with our domain expertise. Table I presents the simulation parameters.

2) UE Mobility Patterns: Realistic UE mobility patterns can be generated using microscopic traffic simulators such as SUMO [13]. In this paper, we leverage SUMO to deploy mobile UEs on the roads in the geographic region of network deployment. Unlike conventional mobility models that are typically used to develop mobility management solutions, SUMO assures that the generated UE mobility patterns correspond to the mobility trend of the UEs in a real network.

3) 3GPP Compliant Handover Events and Procedures: Creating a digital twin for mobility management necessitates a thorough implementation of the 3GPP-defined HO events and parameters discussed in sub-section II-A as well as HO signaling messages. In addition, the 3GPP defined counters and timers for RLF as highlighted in sub-section II-B are vital to gauge the instances of service interruption due to HO settings. To fulfill these requirements, we exploit a 3GPP state-of-the-art system level simulator named SyntheticNET [14], which has been calibrated against real network measurements to ensure authenticity.

B. RL-based Optimization

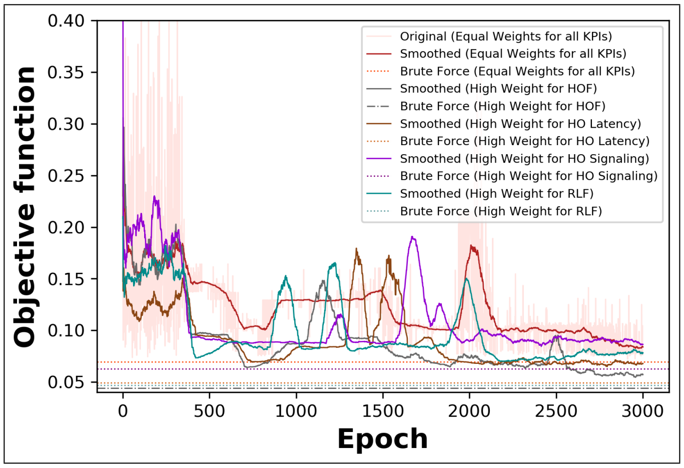


Fig. 3. Convergence of objective function with epochs of soft actor-critic RL using different KPI weights of eq. (5).

Once the digital twin is established, it can be used to securely train the RL agent for developing the mobility management solution. To design the mobility management solution, we are utilizing state-of-the-art soft actor-critic RL algorithm [15] in this study. The choice of the soft-actor critic RL is determined by its capacity to efficiently explore a vast action space with a high sampling efficiency. We define the state space, action space, and reward function for the soft actor-critic RL before comparing the performance with brute-force optimization.

State Space: The parameters of events A2, A3, and A5 influence RSRP, SINR, and throughput of the network [7]. Due to the impact of these optimization parameters on the network, they are prime candidates for characterizing the state of the DT environment. The state vector is represented as follows:

$$S_t = [\eta_t, \phi_t, \zeta_t] \tag{6}$$

where η_t, ϕ_t and ζ_t are the average values of RSRP, SINR and throughput, respectively.

Action Space: The RL agent actions are specified to select the values of optimization variables used in eq. (5) with pre-defined ranges for each COP, such that: $a_t = [A3_{TTT}, A3_{off}, A5_{TTT}, A5_{\Delta}, A2_{th}]$ Table II lists the allowable actions for each of the five COPs.

Reward Function: The scaled value of the objective function defined in eq. 5 is used as the reward for the RL model.

Fig. 3 depicts a performance comparison between the soft actor-critic approach and the brute force method with different KPI weights. The original raw values of the objective function returned by the RL algorithm in each epoch are shown for the scenario with equal weight of all KPIs. To enhance readability, we also incorporate a smoothed line to reflect the objective

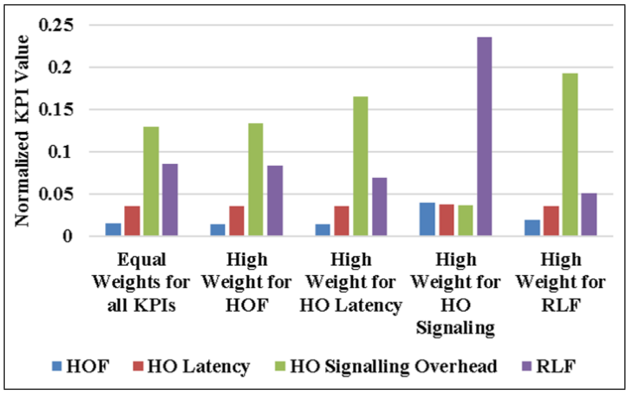


Fig. 4. Variation in individual KPIs with different KPI weights of eq. (5).

function value averaged over 50 epochs. For a scenario where KPIs have equal weights ($\alpha = \beta = \gamma = 0.25$), RL converges in less than 2500 epochs, whereas brute force requires 21,000 iterations. We observe a similar pattern when a greater emphasis is placed on decreasing the number of HOF (α = 0.7, $\beta = \gamma$ = 0.1), HO latency ($\alpha = \gamma$ = 0.1, β = 0.7), signaling overhead ($\alpha = \beta$ = 0.1, γ = 0.7) and the number of RLF ($\alpha = \beta = \gamma$ = 0.1) with convergence times between 2500 to 3000 epochs. However, the faster convergence is accompanied by a modest reduction in the objective function value. This demonstrates that RL has 7 times faster convergence than the brute force while returning near-optimal values.

Although Fig. 3 provides insights into the comparison with brute force, a deeper analysis can demonstrate the influence of KPI weights on the individual value of KPIs. For an in-depth examination, we illustrate the normalized value of four KPIs with varying KPI weights in Fig. 4. A relatively low value for the number of HOF and HO latency can be observed with changing the KPI weights. This highlights that numerous COP combinations can result in low values for both the number of HOF and HO latency. Furthermore, it also demonstrates a weak tradeoff of number of HOF and HO latency with HO signaling overhead and the number of RLF. This signifies that either HO signaling or the number of RLF can be minimized without incurring very high penalty for HOF and HO latency. Fig. 4 also shows a strong tradeoff between the signaling overhead and the number of RLF, since minimizing one of them increases the other significantly. This happens because, in a bid to reduce the signaling overhead, the optimization engine attempts to set the COP combinations that will postpone the HO, resulting in an RLF. These insights can be particularly useful for operators when determining the KPI weights.

IV. Conclusion

With the recent trend towards denser BS deployment and multi-band operation, mobility management and HO optimization have become major bottlenecks. This paper presents a novel mobility management solution that minimizes four KPIs including the number of HOF, HO latency, signaling overhead, and the number of RLF as a function of events A2, A3, and A5 parameters. We formulate and solve a multi-objective optimization problem using soft actor-critic RL. To address the challenge of RL training on a live network, we present a framework for DT-assisted mobility management that trains the RL agent on the digital twin of cellular network. This article also outlines a three-pronged strategy for developing a digital twin for mobility management. Results reveal that the proposed RL solution trained on DT can converge to near-optimal values 7 times faster than brute force. An analysis of the impact on individual KPI values with varying KPI weights reveal that HOF and HO latency generally have lower values and a strong trade-off exists between minimizing HO signaling overhead and RLF.

References

[1] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[2] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless communications*, vol. 24, no. 5, pp. 175–183, 2017.

[3] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover control in wireless systems via asynchronous multiuser deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4296–4307, 2018.

[4] V. Yajnanarayana, H. Rydén, and L. Hévizi, "5G handover using reinforcement learning," in *2020 IEEE 3rd 5G World Forum (5GWF)*, pp. 349–354, IEEE, 2020.

[5] S. Khosravi, H. Shokri-Ghadikolaei, and M. Petrova, "Learning-based handover in mobile millimeter-wave networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 663–674, 2021.

[6] W. Huang, M. Wu, Z. Yang, K. Sun, H. Zhang, and A. Nallanathan, "Self-adapting handover parameters optimization for sdn-enabled udn," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6434–6447, 2022.

[7] M. U. B. Farooq, M. Manalastas, W. Raza, S. M. A. Zaidi, A. Rizwan, A. Abu-Dayya, and A. Imran, "A data-driven self-optimization solution for inter-frequency mobility parameters in emerging networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 570–583, 2022.

[8] M. Manalastas, M. U. B. Farooq, S. M. A. Zaidi, A. Abu-Dayya, and A. Imran, "A data-driven framework for inter-frequency handover failure prediction and mitigation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6158–6172, 2022.

[9] M. U. B. Farooq, M. Manalastas, S. M. A. Zaidi, A. Abu-Dayya, and A. Imran, "Machine learning aided holistic handover optimization for emerging networks," in *ICC 2022 - IEEE International Conference on Communications*, pp. 710–715, 2022.

[10] Y. Ren, J.-C. Chen, and J.-C. Chin, "Impacts of S1 and X2 Interfaces on eMBMS Handover Failure: Solution and Performance Analysis," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 6599–6614, 2018.

[11] 3rd Generation Partnership Project (3GPP) Technical Specification Group Radio Access Network, "Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Measurement Collection for Minimization of Drive Tests (MDT); Overall Description; Stage 2," in *3GPP TS 37.320 version 16.8.0 Release 16*, March 2022.

[12] "Atoll." [Online]. https://www.forsk.com/. Accessed: 5 December, 2022.

[13] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wiessner, "Microscopic traffic simulation using sumo," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2575–2582, 2018.

[14] S. M. A. Zaidi, M. Manalastas, H. Farooq, and A. Imran, "SyntheticNET: A 3GPP compliant simulator for AI enabled 5G and beyond," *IEEE Access*, 2020.

[15] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, *et al.*, "Soft actor-critic algorithms and applications," *arXiv preprint arXiv:1812.05905*, 2018.