Mobility Robustness Optimization for Self-Organizing Networks: A Deep Reinforcement Learning Approach

Kihoon Kim, Eunsok Lee, and Sangheon Pack
School of Electrical Engineering, Korea University, Seoul, Korea.
Email: {rlgns1109, tinedge, shpack}@korea.ac.kr

***Abstract*—Self-organizing network (SON) has emerged as a promising solution to manage dynamic and complex wireless/mobile networks without excessive manual intervention. As a key function of SON, mobility robustness optimization (MRO) aims to prevent radio link failure (RLF) for user equipment (UE) by adjusting handover-related parameters. Although there are several existing studies on MRO, previous studies have limitations in coping with various UE mobility patterns. Additionally, since parameter adjustments are performed on a per-cell basis, detailed adjustments based on individual UE situations become unfeasible. To overcome these limitations, we propose a deep reinforcement learning (DRL)-based MRO scheme where parameter adjustment sets are evaluated by a multi-agent double deep Q-network (MA-DDQN). Evaluation results demonstrate that the proposed DRL-based MRO scheme can reduce RLFs by up to 45. 7% compared to other MRO schemes and completely prevent handover ping-pong events.**

***Index Terms*—Self-Organizing Networks, Mobility Robustness Optimization, Radio Link Failure, Handover Ping-Pong, Deep Reinforcement Learning.**

I. Introduction

Wireless networks have become increasingly complex due to the coexistence of various wireless access networks. In such a complex network environment, passive network management requires significant resources. To address this challenge, a self-organizing network (SON), which enables automatic network management, has emerged [1].

Among SON technologies, mobility robustness optimization (MRO) is a function that aims to provide a stable connection to user equipment (UE). If proper handovers do not occur when the UE is moving, the signal strength of the cell may be reduced, resulting in radio link failure (RLF) [2], and the connection between the cell and UE is lost. For stable connections, it is important to prevent RLFs, and MRO achieves this by adjusting the relevant parameters. The key parameters of MRO are 1) the handover margin (HOM) and 2) time-to-trigger (TTT) [3]. HOM adjusts the threshold of signal strength differences between two cells for the occurrences of the handovers, and TTT is the minimum time required to execute the handover after meeting the handover condition.

Several studies on MRO function have already been carried out. In [4], the author suggests a heuristic algorithm that adjusts the HOM of each cell, based on the RLFs that occur in each cell during a specific period. However, since parameter adjustments are executed only on the basis of the RLF occurrences, it is difficult to immediately respond to dynamic mobility patterns.

To solve this problem, [5] presents a Q-learning-based algorithm so that the central agent learns how to adjust the parameters corresponding to each mobility pattern and applies them before RLFs occur. But in [5], each mobility pattern was classified only by the speed of UEs, so it could not fully respond to the various mobility patterns. Additionally, since both [4] and [5] adjust each parameter in a cell basis, there is a limitation to preventing RLF, an event that occurs in individual UE.

In this paper, we propose a deep reinforcement learning (DRL)-based MRO scheme where parameter adjustment sets are evaluated by a multi-agent double deep Q-network (MA-DDQN). By assigning each MA-DDQN agent to each UE, it is possible to learn more accurately the mobility patterns of UEs and adjust the parameters to prevent RLFs, an event that occurs for each UE. The proposed scheme uses more information to distinguish mobility patterns, which allowed more detailed and accurate parameter adjustment for each mobility pattern. Also, TTT and HOM are individually adjusted for each UE, allowing them to be fine-tuned based on the specific situations of each UE.

The remainder of this paper is organized as follows. In Section II and Section III, the system model and the proposed formulation of the DRL-based MRO scheme are described, respectively. After that, the evaluation results are given in Section IV and followed by the concluding remarks in Section V.

II. System model

Fig. 1 illustrates the system model of the DRL-based MRO scheme. The system model consists of three components: UE, eNB, and the MRO controller.

Within a system environment, UEs continuously gather measurements during their movement and transmit these measurements to the connected eNB at a particular time step. The eNBs then relay these measurements to the central MRO controller. The MRO controller is designed as a multi-agent system, with DDQN agents assigned to each UE. Each DDQN agent is responsible for learning the UE assigned to it. After

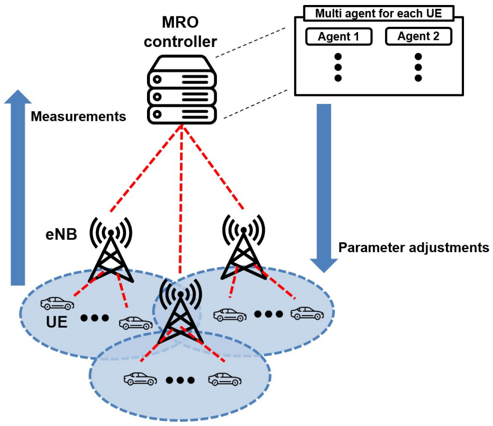

Fig. 1: System model.

learning based on measurements, the DDQN agents determine parameter adjustments. Subsequently, the MRO controller transmits these parameter adjustments to the respective UEs through the eNBs, and the UEs then apply the received adjustments.

In this DRL-based MRO system, the A3 event-based handover is used. Each agent learns to make adjustments to the parameters that lower the probability of RLFs for measurement. When parameter adjustments are applied, the RLFs are prevented in advance.

III. DRL-based MRO scheme

In this section, we present the DRL-based MRO scheme, which aims to find optimal parameter adjustments to prevent RLFs in moving UEs. For this, each agent uses measurements from the assigned UE as the state s_t, and computes optimal parameter adjustments a_t to maximize the reward r_t. The reward r_t is an indicator of the handover performance achieved by a_t. Consequently, the DRL-based MRO scheme comprises three elements: state, action, and reward. First, we use four information as a state.

$$s_t = \{RSRP^t_{Cur},\ SINR^t_{Avg},\ Distance^t,\ BestCell^t\} \quad (1)$$

Current reference signal received power $RSRP^t_{Cur}$ and average signal-to-interference-plus-noise-ratio $SINR^t_{Avg}$ are used to indicate the strength and quality of the signal that the UE receives from the serving cell. Meanwhile, the distance to eNB, $Distance^t$, provides the approximate location of the UE from the base station. Finally, $BestCell^t$ is a value indicating whether the RSRP value received from the currently connected cell is the maximum. Agents combine this information to identify location and mobility patterns and then perform parameter adjustments to prevent RLFs based on this information. Our system model employs two handover parameters as actions in this process.

$$a_t = \{HOM^t,\ TTT^t\} \quad (2)$$

The occurrences of handovers can be controlled by manipulating these values, and the pair of parameters is associated with the A3-based handover condition. The condition of the a3 handover is given as $RSRP^t_T - RSRP^t_S > HOM^t$ [6].

$RSRP_T$ and $RSRP_S$ are values received from target cells and serving cells, respectively. When the difference between the two values exceeds the HOM, a handover is triggered. Therefore, increasing this HOM, the occurrences of handovers can be prevented. Conversely, by decreasing it, handovers can be induced to occur. TTT is the time required before handover is executed after the above A3 condition is satisfied. By manipulating this, it is possible to adjust the timing of handover. These two parameters are individually adjusted by the agents to prevent RLFs in each UE. Finally, we define reward as a performance indicator of parameter adjustments to prevent RLFs. The formula of reward is presented as follows.

$$r_t = -(w_1 \cdot N^t_R + w_2 \cdot N^t_P) + w_3 \cdot Throughput^t \quad (3)$$

In (3) N^t_R and N^t_P mean the number of RLFs and handover ping-pongs (PPs), respectively. PP refers to the frequent handover between two adjacent cells in a short period of time. PP is also considered as a reward because it can disrupt the stability of the communication environment. In the case of throughput, the occurrences of handovers can be prevented at all by adjusting the parameters to lower N^t_P. As a result, the throughput of the entire system can be reduced, so the throughput is also considered. Finally, w_1, w_2, and w_3 are the weights of each element. In most cases, $w_1 > w_2$, since RLFs are more fatal to stable communication than PPs [7]. In this work, w_1 = 0.4, w_2 = 0.2 and w_3 = 0.4.

Parameter	Value
Number of eNBs	9
Number of UEs	8
Number of coverage holes	8
Mobility model	Waypoint mobility
Distance between eNBs (m)	200
UE velocity (m/s)	5, 10, 30 ,60
Total step / step time (s)	60 / 0.5
HOM (dB)	0.0 - 30.0
TTT (ms)	0 - 5120

TABLE I: Simulation parameters.

IV. Simulation Results

For performance evaluation, we use ns3-gym, which combines the ns-3 network simulator and the OpenAi Gym reinforcement learning toolkit [8]. In the simulation scenario, we initially deploy eNBs in a grid pattern, with coverage holes-where signal strength is rapidly diminished-randomly positioned between the eNBs. Subsequently, we established a square pathway within the overlapping coverage area of multiple eNBs. On this square boundary, UEs are randomly placed and move at randomly allocated velocities. Table 1 summarizes the simulation parameters.

The four algorithms for comparison are as follows: 1) No_MRO, 2) MRO_ABC, 3) QMRO, and 4) proposed scheme. No_MRO is an algorithm that does not manipulate any parameters. MRO_ABC and QMRO are algorithms presented in [4] and [5], respectively.

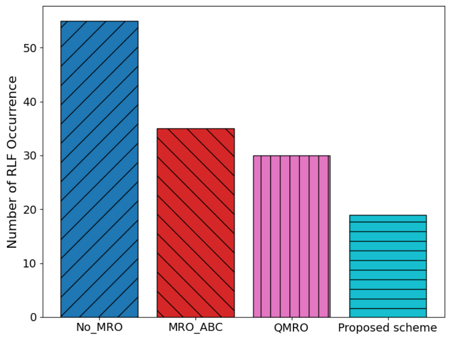

Fig. 2: Number of RLFs occurrences

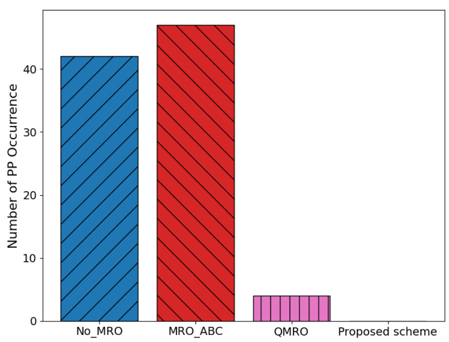

Fig. 3: Number of PPs occurrences

In Fig. 2, the largest number of RLFs occurs in No_MRO, followed by MRO_ABC and QMRO. The proposed scheme shows the smallest RLFs. Compared to No_MRO, the RLFs in the proposed scheme decreased by 34%, and 54% and 63% respectively compared to MRO_ABC and QMRO. and also in Fig. 3, the proposed scheme showed the best performance, and PPs do not occur at all. In the case of MRO_ABC, the RLFs decreased, but the PPs increased, which seems to have resulted in more PPs due to parameter adjustments to reduce RLFs inducing more handovers. The QMRO and proposed scheme decreased in both RLFs and PPs. However, the proposed scheme showed better performance, since individual agents individually adjusted the UE parameters.

V. Conclusion

To ensure stable communication with mobile UEs, it is crucial to minimize RLFs. For this purpose, it is necessary to induce appropriate handovers. We propose a DRL-BASED MRO SCHEME in which each DDQN agent learns and determines the handover parameter adjustments. The evaluation results show that the proposed DRL-based MRO scheme can reduce RLFs by up to 45. 7% better compared to other MRO schemes and completely prevent PPs.

Acknowledgment

This work was supported in part by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-01659,5G Open Intelligence-Defined RAN (ID-RAN) Technique based on 5G New Radio) and ITRC (Information Technology Research Center) support program (IITP-2023-2021-0-01810).

References

[1] P.Muñoz, R. Barco, and I. D. L. Bandera, ”On the Potential of Handover Parameter Optimization for Self-Organizing Networks,” *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1895-1905, June 2013.

[2] Z. Wei, ”Mobility Robustness Optimization based on UE mobility for LTE system,” in *Proc. IEEE WCSP 2010*, October 2010.

[3] A. Alhammadi, M. Roslee, M. Y. Alias , I. Shayea, S. Alraih, ”Dynamic handover control parameters for LTE-A/5G mobile communications,” in *Proc. IEEE RTUWO 2018*, December 2018.

[4] K. Kitagawa, T. Komine, T. Yamamoto, and S. Konishi, “A Handover Optimization Algorithm with Mobility Robustness for LTE systems,” in *Proc. IEEE PIMRC 2011*, September 2011.

[5] S. S. Mwanje, and A. M. Thiel, “Distributed Cooperative Q-Learning for Mobility Sensitive Handover Optimization in LTE SON,” in *Proc. IEEE ISCC 2014*, June 2014.

[6] A. Klein, N. P. Kuruvatti, J. Schneider, and H. D. Schotten, ”Fuzzy Q-Learning for Mobility Robustness Optimization in Wireless Networks,” in *Proc. IEEE GC Wkshps 2013*, December 2013.

[7] M. L. Marí-Altozano , S. S. Mwanje , S. L. Ramírez , M. Toril , H. Sanneck, and C. Gijón, ”A Service-Centric Q-Learning Algorithm for Mobility Robustness Optimization in LTE,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 3, pp. 3541-3555, September 2021.

[8] P. Gawlowicz and A. Zubow, “ns-3 meets OpenAI Gym: The Playground for Machine Learning in Networking Research,” in *Proc. ACM MSWiM 2019*, November 2019.