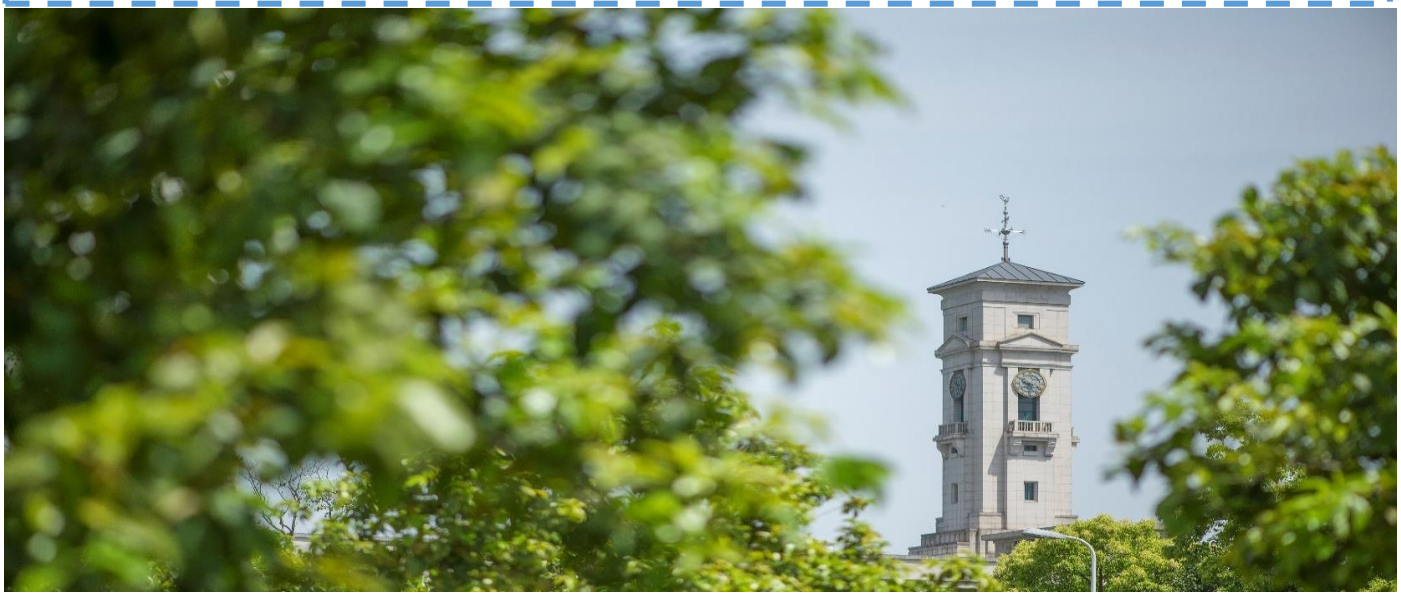
Intelligent Handover Triggering Mechanism in 5G Ultra-Dense Networks via Clustering-Based Reinforcement Learning

Liu, Q., Kwong, C.F., Wei, S., Li, L. Zhang, S.



University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, 315100, Zhejiang, China.

First published 2021

University of Nottingham
UK | CHINA | MALAYSIA

Intelligent Handover Triggering Mechanism in 5G Ultra-Dense Networks via Clustering-based Reinforcement Learning

Qianyu Liu,[a] Chiew Foong Kwong,[*,b] Sun Wei,[b] Lincan Li,[b] and Sibo Zhang [c]

[a] University of Nottingham Ningbo China, International Doctoral Innovation Centre, Ningbo, China

[b] University of Nottingham Ningbo China, Department of Electrical and Electronic Engineering, Ningbo, China

[c] Imperial College London, Department of Electrical and Electronic Engineering, London, UK

*Corresponding author: c.f.Kwong@nottingham.edu.cn

ABSTRACT

Ultra-dense networks (UDNs) are considered as key 5G technologies. They provide mobile users a high transmission rate and efficient radio resource management. However, UDNs lead to the dense deployment of small base stations (BSs) that can cause stronger interference and subsequently increase the handover management complexity. At present, the conventional handover triggering mechanism of user equipment (UE) is only designed for macro mobility and thus could result in negative effects such as frequent handovers, ping-pong handovers, and handover failures on the handover process of UE at UDNs. These effects degrade the overall network performance. In addition, a massive number of BSs significantly increase the network maintenance system workload. To address these issues, this paper proposes an intelligent handover triggering mechanism for UE based on Q-learning frameworks and subtractive clustering techniques. The input metrics are first converted to state vectors by subtractive clustering, which can improve the efficiency and effectiveness of the training process. Afterward, the Q-learning framework learns the optimal handover triggering policy from the environment. The trained Q table is deployed to UE to trigger the handover process. The simulation results demonstrate that the proposed method can ensure the stronger mobility robustness of UE that is improved by 60%–90% compared to the conventional approach with respect to the number of handovers, ping-ping handover rate, and handover failure rate while maintaining other key performance indicators (KPIs), that is, a relatively high level of throughput and network latency. In addition, through integration with subtractive clustering, the proposed mechanism is further improved by an average of 20% in terms of all the evaluated KPIs.

Keywords: handover management, Q-learning, subtractive clustering, ultra-dense networks

1. INTRODUCTION

To manage increasing demand for mobile data traffic and efficient data delivery, ultra-dense networks (UDNs) have been introduced in the fifth-generation mobile communications system (5G). UDNs involve the close deployment of small base stations (BSs) at traffic hotspots. Using this method, data traffic is mainly delivered by small BSs, which can significantly increase the system capacity, spectrum efficiency, throughput, coverage and provide ubiquitous access for user equipment (UE) [1]. When UE moving across the coverage of small BSs, the handover process needs to be performed to ensure UE's data delivery. As defined in the Third-Generation Partnership Project (3GPP) [2], the handover process is triggered by A3 event measured by UE. A3 event occur when the difference between the reference signal receiving power (RSRP) from UE serving cells and neighbouring cells is higher than a pre-determined condition, the handover hysteresis margin (HHM). When meeting the entering condition of A3 events, UE will wait for a pre-defined period, that is, the time to trigger (TTT). Subsequently, if the A3 event entering condition remains satisfied, the UE reports the A3 event to its serving base station (BS), and the handover process is then executed based on the Xn interface of the BS. The UE connection will subsequently switch to the neighbouring cell with the strongest RSRP.

On the other hand, UDNs also increase the complexity of cellular networks and introduce new challenges to handover management[3]. Traditionally, the A3 event was designed as the handover triggering mechanism for UE in macro BS systems. Therefore, the A3 event may face the following three challenges within 5G-UDNs. First, because the coverage area of small BSs is much lower than macro BSs, UE will meet the edge of the cell more frequently. UE can have many more neighbouring cells as potential handover targets in UDNs. In this situation, the A3 event's entering condition is easily satisfied, and the handover process is frequently triggered by UE even with short physical movements [4]. Since the handover process can interrupt the UE's serving link before transferring its connection to the target cell. Thus, frequent handovers can also overload core network signalling, diminish system capacity and degrade overall system performance [5]. Second, the decision-making process is easily affected by interference and frequent handovers continually occurring among serving and target cells (known as the ping-pong effect). As small BSs are deployed denser and closer to each other, this may result in much stronger inter-cell interference. As such, inter-cell interference can result in much stronger fluctuations in the RSRP that further worsen this problem. Third, the A3 event needs to adjust the handover parameters, that is, HHM and TTT, to avoid frequent handovers, ping-pong

effects, and handover failure rates. To achieve this target, the network operator needs to frequently conduct extensive measuring activities and data analysis to determine the suitable handover parameters [6]. With the increasing deployment of BSs, the network maintenance workload and complexity also significantly increase. Therefore, in A3 events, it is unrealistic to adjust HHM and TTT to optimal levels to maintain a high level of network performance at all times.

Based on the analysis above, simply implementing the current A3 event in 5G-UDNs can lead to system performance degradation. To overcome the limitation of A3 event, and increase the mobility robustness of UE in 5G-UDNs, a novel handover triggering mechanism that can precisely trigger handover process with low maintenance cost is necessary to be investigated. In this study, we integrated both advantages of Q-learning and subtractive clustering techniques to develop an intelligent handover triggering mechanism for UE in 5G-UDNs. The main contributions of this study are summarised as follows:

- First, we develop an instant handover triggering mechanism that can trigger handover processes precisely based on multiple decision criteria. The proposed mechanism has the objectives of enhancing user mobility robustness while maintaining other high-level key performance indicators (KPIs).
- Second, we proposed a Q-learning framework to achieve an optimal handover triggering policy by considering multiple network metrics, that is, the RSRP, signal to interference and noise ratio (SINR), and transmission distance. The trained Q table is utilised as a triggering mechanism of UE to decide the optimal triggering timing without additional handover conditions intelligently.
- This study utilises the subtractive clustering technique to generate state vectors from historical data. Using this method, the input metrics are systematically categorised into corresponding states with respect to the actual data distribution, which can improve the trained Q table's accuracy and effectiveness. This categorisation method can effectively process multiple fluctuating network metrics and minimise the impact of noise and interference in handover triggering decision-making.

To the best of our knowledge, this is the first study to directly apply Q-learning to make handover triggering decisions in 5G-UDNs rather than optimise handover parameters. This is also the first study to utilise subtractive clustering to optimise the Q-learning framework for handover decision-making.

The rest of this article is organised as follows: Section 2 reviews some existing studies that are related to this paper. Section 3 introduces the channel model and performance metrics used in this study. The detailed proposed method is described in Section 4. The proposed triggering mechanism is compared with the other existing handover triggering mechanisms to evaluate its performance. The simulation designs and results are shown in Section 5. The study's main conclusion is summarised in Section 6.

2. RELEVANT STUDIES

2.1 Threshold comparison based handover triggering optimisation methods

To address the negative handover effects caused by A3 events, different handover management algorithms are proposed in the current literature. One way to optimise the handover triggering mechanism is adjusting the handover-related parameters, that is, HHM and TTT are adaptively based on different algorithms. References [7]–[9] reported handover optimisation methods based on threshold comparisons with specific metrics. Reference [7] proposed a handover parameter optimisation method to enhance mobility robustness across small cells. The proposed method adopts a threshold to classify categories of handover failures and then updates handover parameters according to dominant failures. To avoid handover failures due to radio link failures, Reference [8] developed a novel distributed auto-tuning algorithm based on metaheuristic algorithms that can automatically update HHM and TTT on the basis of user speed, RSRP, and SINR. In Reference [9], the authors integrated fuzzy logic into the conventional handover decision to dynamically adjust HHM and TTT. The signal levels from both serving and target cells were used as a fuzzy interference engine input to generate the adjusted margin as output. The simulation results in References [7]–[9] showed that compared with the traditional method, the proposed technique significantly reduces the number of handovers, ping-pong effect, handover failure rate, and radio link failure rate. However, some essential parameters of these proposed algorithms, that is, the thresholds, fuzzy rules, and fuzzy membership functions, rely heavily on human experience to define that is not applicable in practical situations.

2.2 Reinforcement based handover triggering optimisation methods

Since reinforcement learning and deep learning have demonstrated their powerful learning, decision-making and inference capabilities in many applications, such as [10]–[14]. As such, these techniques can be considered as effective ways to enable intelligent handover management. References [15]–[17] described reinforcement learning-based handover parameter optimisation methods. In Reference [15], the authors proposed a handover parameter tuning method that effectively detects handover events and minimises false handover triggers. To achieve optimal handover performance, the proposed method can also self-tune the handover decision parameters by defining two state variables for the Markov decision process, that is, handover decision parameters and radio state. In Reference [16], a Q-learning-based framework that can adjust HHM and TTT according to the UE speed was proposed. A multiple attribute decision-making method is then applied to choose the most suitable cell as a handover target. Reference [17] also proposed a Q-learning-based mobility robustness optimisation method to learn the most suitable HHM and TTT based on different UE speeds. The simulation results in References [15]–[17] showed that the three proposed methods can reduce the call drop rate, handover failure rate, and ping-pong handover ratio for high-speed movement UE compared to the traditional approach. However, the reinforcement learning framework in References [15]–[17] cannot define a large-scale state and action space; otherwise, the training process becomes inefficient. To scale down the size of the state–action space, References [15]–[17] attempted to categorise input metrics, for example, the speed and RSRP, into same the length range as the state vectors. This categorisation method lacks systematic methodologies to reflect actual data distribution into state vectors and thus could potentially affect the effectiveness and accuracy of the generated Q table.

These studies indicated that adjusting the HHM and TTT can effectively improve handover performance. However, the presence of HHM and TTT causes the triggering process to become not instantaneous, as the handover process is only triggered after these two pre-determined conditions. The coverage area of small BSs in UDNs is much smaller than in macro BSs. Small BSs only leave a very short time for triggering mechanisms to react and then execute the subsequent process. In this condition, an instant handover triggering mechanism can ensure the reliability and seamlessness of communications. In addition to optimising handover parameters for triggering mechanisms, some studies developed instant handover triggering mechanisms to directly trigger the handover process without any additional handover parameters and conditions.

2.3 Fuzzy logic based instant handover triggering mechanisms

In References [17] and [18], a triggering threshold called the handover factor that is generated by fuzzy logic was used to minimise the number of handovers. The RSRP, SINR, and user speed are input into the fuzzy interference system. The input metrics are processed by a group of pre-defined fuzzy membership functions and fuzzy rules to generate the output handover factor. The handover factor is distributed between 0 and 1, with 1 denoting that the probability of handover occurrence is very high. Conversely, 0 denotes that the possibility of handover occurrence is the lowest. The simulation results in References [17] and [18] showed that the handover factors can minimise unnecessary handovers and ping-pong effects. However, these three studies did not further discuss how to define an optimal membership function for each decision metric. Therefore, the reliability of these methods cannot be ensured with the changes in application scenarios.

2.4 Intelligent handover triggering mechanisms

In Reference [20], the authors proposed an adaptive fuzzy logic-based handover triggering method. The fuzzy membership functions and rules were first generated by subtraction from historical data and then tuned to the optimal level concerning different application scenarios by neural networks. Compared with conventional fuzzy logic-based handover triggering mechanisms and other traditional approaches, the proposed algorithm in Reference [20] demonstrated that it provided a significant improvement in handover performance in terms of mobility robustness and mobility load balancing. However, the approach in Reference [20] was unable to process too many metrics as input parameters; otherwise, the whole system becomes complicated and the training process is time-consuming. In Reference [21], the authors adopted model-free asynchronous advantage actor-critic (A3C) reinforcement learning techniques to learn an optimal handover method. Each network user is a local agent to interact with the environment and learn a local handover policy. The local handover policy in each UE is then uploaded and integrated as the global handover policy at the global controller. Each UE regularly copies up-to-date handover policies from a global controller to trigger the handover process and supervise the subsequent leaning process when controller updates are required. The simulation results in Reference [21] demonstrated that the proposed method can achieve better performance than existing online techniques in terms of handover rates. However, other KPIs, that is, the ping-pong handover rate, handover failure rate, and throughput, were not further evaluated in this

paper. Due to the limited computation power of UE, it is inapplicable to utilise UE as a training agent in A3C framework.

According to the aforementioned analyses, Q-learning demonstrates its powerful capabilities in handover triggering solutions. All of the Q-learning-based approaches focus on optimising handover parameters rather than instant triggering mechanisms. However, all of the Q-learning approaches lack systematic methodologies to convert radio conditions, that is, the RSRP and SINR, into state vectors. As shown in Reference [22], subtractive clustering can categorise data into corresponding groups based their distribution. This may provide a solution to define the proper state vectors for Q-learning frameworks in handover decision-making.

3. SYSTEM MODEL

In this study, we adopt two-tiered UDNs that consist of LTE-Advance and 5G networks. This two-tiered structure was widely used in many previous studies such as [18], [20], and [22]. Fig. 1 presents an example of proposed network deployment in this study. The LTE-Advance network consisting of N_m macro BSs operates under a 5 GHz frequency band. 5G networks consisting of N_s small BSs operate at million-metre wavebands. There are N_{ue} randomly moving within the deployed area with a constant velocity V_{ue}. Each UE is associated with one macro or small BS to exchange signalling. During UE movement, the UE periodically collects handover-related metrics from neighbouring BSs, that is, the RSRP, RSRQ, and SINR, and reports to its serving-based station. The proposed triggering mechanism is deployed at the UE to determine the optimal timing and then reports to its serving BS for handover execution.

3.1 Channel model

According to Reference [24], a large-scale channel model for macro Eq. (1a) and small base stations Eq. (1b) in urban areas are adopted. The path loss (PL_{ij}) between UE i and BS j is defined as

$$PL_{ij-Uma} = 32.4 + 20log_{10}(f) + 30log10(d_{i,j}) + \chi \quad (1a)$$

$$PL_{ij-Umi} = 32.4 + 20log_{10}(f) + 31.9log10(d_{i,j}) + \chi \quad (1b)$$

$$d_{i,j} = \sqrt{(x_i - x_j)^2 - (y_i - y_j)^2} \quad (1c)$$

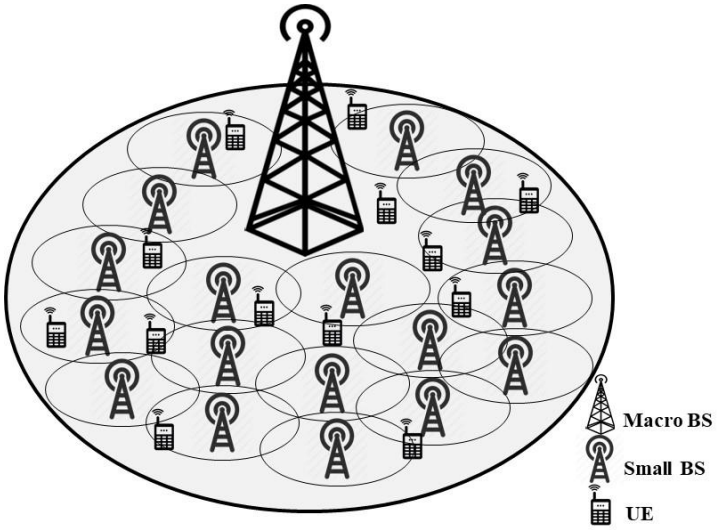


Fig. 1 Two-tiered system model

In Eqs. (1a) and (1b), $d_{i,j}$ represents the transmission distance between UE i and BSj calculated by Eq (1c). (x_i, y_i) and (x_j, y_j) in Eq. (1c) are the coordinates of UE i and BS j, f is denoted as the carrier frequency for small and macro BSs, and χ is the interference and noise modelled by Gaussian random and Rayleigh random variables. The RSRP of UE *i* is then calculated by subtracting PL_{ij} from the cell reference signal of BS *j*.

According to Reference [25], the SINR from BS j to UE i is formulated as

$$\gamma_{j,i} = 10log_{10}(\frac{P_j d_{ij}^{-\alpha}}{\sum_{o=1}^{n_{m-1}} P_o d_{io}^{-\alpha} + P_n}), \quad (2)$$

where P_j and P_o represent the transmission power of UE serving BSs and neighbouring BSs, respectively, and d_{ij} and d_{io} represent the distance between the UE to its serving and neighbouring BSs. P_n is the power spectral density of the background noise and n_m represents the number of BSs around the UE.

3.2 System measurements

Several KPIs are used in this study to quantify system performance due to different handover triggering mechanisms.

The first KPI is the average number of handovers per UE ($\overline{NOH}$), which is the essential parameter to quantify the handover frequency in the entire simulation. The average handovers per UE is formulated as

$$\overline{NOH} = \frac{\sum_{i=1}^{N_{ue}} NOH_i}{N_{ue}}, \tag{3}$$

where NOH_i is the number of UE i handovers and N_{ue} is the total amount of UE in the environment. The second KPI is the probability of ping-ping handovers (PPHOs) used to determine unnecessary handovers between two BSs. The PPHO is counted when there are continual handovers by UE between the target cell and presently serving cells within a certain interval T_p. Thus, the average PPHO probability is calculated as

$$\overline{P(PPHO)} = \frac{N_{PPHO}}{N_{HO}}, \tag{4}$$

where N_{PPHO} is the number of PPHOs that occur during the entire simulation and N_{HO} is the number of handovers during the entire simulation, respectively.

The third KPI is the probability of handover fails (HOFs). According to the analysis in Reference [26], the handover process may fail if it is triggered too early or too late. Under these two conditions, the UE may out of the target coverage area or serving cell and subsequently lead to radio link failure before completely establishing handover. Moreover, HOFs may also occur when there is UE handover to the wrong cell. When this occurs, the target cell does not have sufficient resources to maintain UE connections. Therefore, the probability of HOF is a key parameter to evaluate the reliability of the proposed handover triggering mechanism, which is formulated as

$$\overline{P(HOF)} = \frac{N_{HOF}}{N_{HO}} \tag{5}$$

The fourth KPI is the average UE throughput in the entire simulation, which can reflect the quality of network service. According to Reference [27], the system throughput is calculated using Shannon's capacity theory. The correction factor is adopted in this formula to account for the inherent implementation losses, that is, the reference symbol loss ($\mathcal{L}_{\text{ReferenceSymbol}}$) and cyclic prefix loss ($\mathcal{L}_{\text{CyclicPrefix}}$). Therefore, Shannon's capacity theory is formulated as

$$\Gamma_{total} = \xi \times B \times (log_2(1 + 10^{\gamma_{j,i}/10})) \tag{6a}$$

$$\xi = \mathcal{L}_{CyclicPrefix} \times \mathcal{L}_{ReferenceSymbol} \tag{6b}$$

$$\mathcal{L}_{CyclicPrefix} = \frac{T_{frame} - T_{CP}}{T_{frame}} \tag{6c}$$

$$\mathcal{L}_{ReferenceSymbol} = \frac{N_{SC} \times \frac{N_S}{2} - 4}{N_{SC} \times \frac{N_S}{2}} \tag{6d}$$

$$\mathcal{B} = \frac{N_{SC} \times N_S \times N_{rb}}{T_{sub}}, \tag{6e}$$

where Γ_{total} represents the sum of throughput gain by UE in bps; $\gamma_{j,i}$ represents SINR between UE i and BS j obtained from Eq. (2); ξ is the correction factor and $\mathcal{B}$ is the system bandwidth assigned to UE in Hz; T_{frame} is the interval of one orthogonal frequency division multiple access (OFDMA) frame and equals 10 ms; T_{CP} is the total cyclic prefix time of all OFDMA symbols in a frame calculated as (5.2μs + 6 × 4.69 μs) × 20 = 666.8 μs; N_{SC} is the number of subcarriers in the physical resource block (PRB), which is 12 subcarriers for both macro and small BSs; N_S is the number of OFDMA symbols within a subframe, which is 14 symbols for macro BSs and 28 symbols for small BSs; and N_{rb} is the number of PRBs assigned to the UE, which is 100 for macro BSs and 275 for small BSs. The bandwidth assigned to each PRB is the smallest unit of bandwidth that is assigned and can only be applied to one UE, and T_{sub} is the time interval of an OFDMA subframe and equals 1 ms for both macro and small BSs.

The last KPI is the network latency that directly affects network performance. According to the analysis in References [5] and [27], this study considers the network latency from BS j to UE i during time t, which is denoted as $\widehat{\Delta}_{i,j}^{t}$ and formulated as

$$\hat{\Delta}_{i,j}^{t} = \ell_{trans} + \ell_{propa} + \ell_{ho} + \ell_{deal} + \ell_{queue}, \tag{7}$$

where ℓ_{trans} is the transmission latency; ℓ_{propa} is the propagation latency; ℓ_{ho} is the handover latency; ℓ_{deal} is the packet handling latency; and ℓ_{queue} is the queuing latency, respectively. ℓ_{deal} and ℓ_{queue} are much shorter than ℓ_{trans} and ℓ_{propa}, so the last two items in Eq. (7) can be omitted. Eq. (7) is then rewritten as

$$\hat{\Delta}_{i,j}^{t} = \frac{\Theta}{r_i} + \ell_{max\,i.j} \times \frac{d_{i.j}}{d_y} + \ell_{ho} \tag{8}$$

The first item in Eq. (8) calculates ℓ_{trans}. Θ represents the transmitting packet size and is 100 kbit in this study. r_i is the UE i throughput. The second part of Eq. (8) obtains ℓ_{propa}, where $\ell_{\max i.j}$ is the maximum propagation latency from BS j to UE i, which is assumed to be 20 ms for macro BS and 10 ms for small BS, respectively. $d_{i.j}$ is the distance between UE i and its serving BS j. d_y represents the maximum transmission distance from BS j to UE i. ℓ_{ho} is assumed to be 20 ms based on our measurements from real environments.

4. PROPOSE METHOD

Fig. 2 demonstrates the proposed framework based on Q-learning and subtractive clustering. During UE movement, the UE will frequently collect handover-related metrics, that is, the RSRP, SINR, and transmission distance (d), from its serving and neighbouring BSs. These data are stored in the database as historical data. During the training stage, the historical data are used to build the Q-learning framework, which will be explained in detail in Section 4.1. To increase training efficiency, subtractive clustering is adopted to locate the clusters for each input metric and categorise input metrics into state vectors (Section 4.2). The trained Q table of the framework is used as a triggering mechanism to enable the UE to select the optimal timing and report to the BS for handover execution. The detailed handover process under the proposed triggering mechanism will be described in Section 4.3.

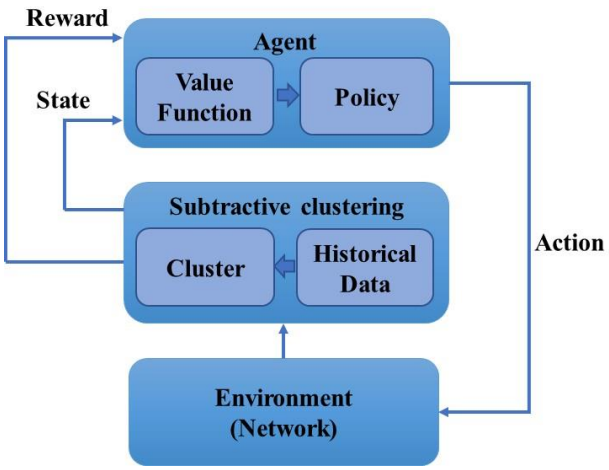


Fig. 2 Subtractive clustering-based Q-learning framework for handover triggering

4.1 Q-learning framework for handover in 5G-UDNs

Q-learning is a model-free and off-policy reinforcement algorithm that provides the optimal policy from a set of Markov decision processes. The Q-learning framework consists of agent and triple $< \mathcal{S}, \mathcal{A}, \mathcal{R} >$, where $\mathcal{S}$ and $\mathcal{A}$ represent the sets of all possible states and actions, respectively, and $\mathcal{R}$ is the reward function. When an environment is in state $s_t \in \mathcal{S}$ at time step t, $a_t \in \mathcal{A}$ is executed by the agent. The environment is subsequently subjected to a transition from s_t to $s_{t+1} \in \boldsymbol{\mathcal{S}}$, and an immediate reward $r_t \in \mathcal{R}$ is received by the agent. The main target of agent is to learn the optimal action for each state (policy) from the environment that can maximise the accumulated reward.

In this study, the environment refers to the UDNs in a specific area l, and the triple $< \mathcal{S}, \mathcal{A}, \mathcal{R} >$ in this framework is defined as

- Action $\mathcal{A}$: At time step t and area l, the action $a_{i,l,t} \in \mathcal{A}$ for UE i is set to execute the handover process or maintain the UE connection. If the agent decides to execute the handover process, the UE link switches to a new BS at time $t+1$. Otherwise, the UE will maintain its link to the previous serving BS.
- State $\mathcal{S}$: The input metrics, that is, the RSRP, SINR, and transmission distance (d), are first normalised between 0 and 1. The normalised value for each metrics x is then mapped into the corresponding cluster to find its cluster index. The states are represented by the combination of the cluster index. At time step t and area l, the states $s_{i,l,t}$ for UE i are

$$s_{i,l,t} = \left\{ \widetilde{c_{k,RSRP}^{i,l,t}}, \widetilde{c_{k,SINR}^{i,l,t}}, \widetilde{c_{k,d}^{i,l,t}} \right\}, \tag{9}$$

where $s_{i,l,t} \in \mathcal{S}$ and $\widetilde{c_{k,RSRP}^{i,l,t}}$, $\widetilde{c_{k,SINR}^{i,l,t}}$, and $\widetilde{c_{k,d}^{i,l,t}}$ are the input data at time t and area l belonging to the *k-th* cluster of the RSRP, SINR, and d, respectively. If the handover process is executed by the agent at time t, the state at $t+1$ is updated based on the RSRP, SINR, and d from a new serving BS. Otherwise, the state at $t+1$ is updated based on the input metric of the current BS.

- Reward $\mathcal{R}$: The sum of centre value of the corresponding cluster is utilised as the reward value at each time step *t*. At time step t and area l, after the agent executes an action $a_{i,l,t}$, the reward for UE i is defined as

$$r_{i,l,t} = \widetilde{x_{k,RSRP}^{i,l,t+1}} + \widetilde{x_{k,SINR}^{i,l,t+1}} + \widetilde{x_{k,d}^{i,l,t+1}}, \tag{10}$$

where $r_{i,l,t} \in \mathcal{R}$ and $\widetilde{x_{k,RSRP}^{i,l,t+1}}$, $\widetilde{x_{k,SINR}^{i,l,t+1}}$, and $\widetilde{x_{k,d}^{i,l,t+1}}$ represent the centre values of clusters $\widetilde{c_{k,RSRP}^{i,l,t+1}}$, $\widetilde{c_{k,SINR}^{i,l,t+1}}$, and $\widetilde{c_{k,d}^{i,l,t+1}}$, respectively. If the handover process is executed by the agent at time t, the reward signal is obtained from the new serving BS. Otherwise, the reward signal is obtained from the current serving BS.

After establishing the framework based on the aforementioned information, the Q-learning framework updates its value function, also known as the Q table, through several epochs. Assuming that the agent chooses an action based on policy π, the Q table is defined to represent every state–action pair. The expected total discount reward received from starting action a in state s is based on policy π. For the optimal policy π^*, the Q^{π^*} is formulated as

$$Q^{\pi^*}(s_t, a_t) = E\left[r(s_t, a_t) + \gamma * \underset{a_{t+1}}{max}\{Q^{\pi^*}(s_{t+1}, a_{t+1})\}\right], \quad (11)$$

where $\gamma \in (0,1)$ is adopted as a discount factor to balance immediate and future rewards. During the learning stage of Q-learning, the agent estimates the Q value from received rewards using the temporal difference (TD) error, which means the difference between the actual Q value ($Q(s_t, a_t)$) and its currently estimated Q value ($\hat{Q}(s_t, a_t)$). Therefore, the Q value at time $t+1$ and $\hat{Q}_{t+1}(s_t, a_t)$ is updated by adding a discount TD error to the currently estimated $\hat{Q}_t(s_t, a_t)$ as

$$\begin{aligned}\hat{Q}_{t+1}(s_t, a_t) &= \hat{Q}_t(s_t, a_t) + \eta * \left[Q(s_t, a_t) - \hat{Q}_t(s_t, a_t)\right] \\ &= \hat{Q}_t(s_t, a_t) + \eta * \left[\mathcal{R}(s_t, a_t) + \gamma * \underset{a_{t+1}}{\max} Q(s_{t+1}, a_{t+1}) - \hat{Q}_t(s_t, a_t)\right] \quad (12)\end{aligned}$$

where $\eta \in (0,1)$ is the learning rate to balance the new and old information. For example, when η=0, all of the new information is abandoned and no further Q value is update required; when η=1, all of the oldest information is discarded and the Q value is updated entirely from the latest information.

In this study, each epoch comprises 10000 simulation time steps and is equivalent to 2.7 hours of actual network time. For each epoch, the accumulated reward (R_e) is calculated as

$$R_e = \sum_{t=1}^{n} r_{i,l,t} \quad (13)$$

The training stage is terminated when the accumulated reward is converged. To achieve optimum Q values, ϵ-greedy is adopted to facilitate a trade-off between exploration and exploitation of the state–action pair. With ϵ-greedy, at each time step t, the agent performs the action with the maximum reward, that is, $a_i = arg\,\underset{a}{max}\, Q^*(s_i; a)$ with probability $1 - \epsilon$; otherwise, it will take a random action. In the initial training phase, ϵ is set to nearly 1 and gradually becomes 1 as each epoch increases. The Q-learning-related parameters are $\{\gamma = 0.9, \eta = 0.1, \text{and } \epsilon = 0.9 - 0.1\}$ in this study.

4.2 Subtractive clustering

To improve the training efficiency and obtain a small Q table, it is necessary to reduce the scale of the state–action pairs. The traditional method is to categorise the related metrics into several equal length states. However, this categorisation method cannot reflect the actual characteristics of the input metrics. For example, the RSRP is divided into five equal length states -20 to -50 dB, -50 to -80 dB, -80 to -110 dB, -110 to -130 dB, and -130 to -160 dB based on the traditional categorisation method. The actual data distribution of RSRP is concentrated between -80 and -140 dB. Therefore, the states need

to concentrate at -80 to -140 dB, rather than the other intervals, to ensure accuracy and effectiveness of the training results. In this study, we introduce a more systematic subtractive clustering technique to categorise the handover metrics into corresponding states based on the data distribution. Categorising input metrics into clusters effectively processes uncertain and imprecise data, minimising the effect of inference and noise on decision-making.

For *m* input metrics and each metric with *n* data points,

$$\{\overrightarrow{\boldsymbol{x_i}} = (x_{i1}, x_{i2}, \dots, x_{im}) | i \in [1, n]\} \tag{14}$$

Three input indicators for subtractive clustering are used in this study. Each metric has 5000 data points, hence n=5000 and m=3. A data point is counted as the high potential value if it has many neighbouring points. The potential value of each data points is evaluated as

$$P_i = \sum_{j=1}^{n} e^{-\alpha \|\overrightarrow{x_i} - \overrightarrow{x_j}\|^2} \tag{15a}$$

$$\alpha = \frac{4}{r_a^2} \tag{15b}$$

In Eq. (15b), r_a defines a neighbourhood's effective radius. The data outside r_a have only a limited influence on P_i.

After calculating P_i for each data point, the point with the highest P_i is located as the first cluster centre. P_i for the rest of the data points is revised based on the potential P_1^* of the first cluster centre $\overrightarrow{\boldsymbol{x_1}}^*$ as

$$P_i \leftarrow P_i - P_1^* e^{-\beta \|\overrightarrow{x_i} - \overrightarrow{x_1}^*\|^2} \tag{16a}$$

$$\beta = \frac{4}{r_b^2} \tag{16b}$$

Subsequently, P_i of the rest of the data is discounted by a function, $e^{-\beta \|\overrightarrow{x_i} - \overrightarrow{x_1}^*\|^2}$, which includes the distance between each data point to the first cluster centre $\overrightarrow{\boldsymbol{x_1}}^*$. According to this function, the data points near the first cluster centre are unlikely to be selected as the new cluster centre as its P_i is significantly discounted.

Next, the point with the highest revised P_i is then located as the new cluster centre. P_i of the rest of the points continues to decrease as new centres are found. When the kth centre of the cluster is located, P_i of the rest of the data points is updated as

$$P_i \leftarrow P_i - P_k^* e^{-\beta \|\overrightarrow{x_i} - \overrightarrow{x_k}^*\|^2}, \tag{17}$$

where $\overrightarrow{\boldsymbol{x_k}}^*$ is the kth cluster centre with potential value P_k^*. The new cluster centres continue to be found until $P_k^* < \varepsilon P_1^*$, where ε is the rejection ratio. The distance between each cluster centre is controlled by β.

If there are k clusters located for input metric m, the set of cluster is represented by $\{\boldsymbol{C_m} = (c_{1m}, c_{2m}, \dots, c_{km})\}$. Similarity, the set of cluster centres is denoted as $\{\overrightarrow{\boldsymbol{x_m}}^* = (x_{1m}, x_{2m}, \dots, x_{km})\}$ for the metric m cluster. The parameters related to subtractive clustering in this study are set as $\{\alpha = 16, \beta = 12, \text{and } \varepsilon = 0.005\}$.

The subtractive clustering-based Q-learning algorithm is described by algorithm 1.

Algorithm 1: Subtractive clustering-based Q-learning for UE i in area l

Input: *historical data, that is, the RSRP, SINR, d, etc.*
Locate the cluster for each input data using Eqs. (14)-(17)
Initialise Q(s,a) arbitrarily, $\forall\, s \in \mathcal{S}$, $a \in \mathcal{A}$, and Q(terminal_state)=0
for *each epoch,* ***do***
    Initialise $\mathcal{S}$ based on the cluster index
    for *each time step t,* ***do***
        Update the UE location
        Compute the RSRP, SINR, and d from each BS to UE_i
        Observe the input metric from the serving BS and convert it to state $s_{i,t}$ from the cluster
        Choose $\mathcal{A}$ from $\mathcal{S}$ using ϵ-greedy policy
            if *$a_{i,t}$ = execute handover process*
                Select the neighbouring BS with max(SINR) as the handover target BS_{j+1}
                Transfer the UE connection to the new BS_{j+1} and observe the reward from BS_{j+1}
            Else
                Maintain the UE connection with the current BS_j and observe the reward from BS_j
            end if
        Update the Q value using Eq. (12)
        $\mathcal{S} \leftarrow \mathcal{S}$
        until $\mathcal{S}$ is terminal
    End
End
Output: *Q table with the optimum Q value*

4.3 Handover triggering using the trained Q table

After the proposed framework learns the optimal handover policy from a specific application scenario *l*, the trained Q table is utilised by the UE as the handover triggering mechanism. During the movement of UE *i*, the measuring data, that is, the RSRP, SINR, and d at time *t*, are first converted into state vector $s_{i,l,t}$ using Eq. (9). The UE then searches corresponding action $a_{i,l,t}$ with the maximum accumulated reward from the Q table based on $s_{i,l,t}$. If optimal action $a_{i,l,t}$ for state $s_{i,l,t}$ is *execute the handover process*, then the handover process is triggered by the UE. As shown in Fig. 3, once the handover process is triggered, the UE reports the handover event to its serving BS. Subsequently, the UE serving BS selects a neighbouring BS with the highest SINR and sends a handover request to it. The radio resource control (RRC) is reconfigured after target BS acknowledges the handover request. In this phase, the connection between the UE and its serving BS is transferred to the target BS, subsequently completing the RRC reconfiguration. If optimal action $a_{i,l,t}$ for state $s_{i,l,t}$ is *maintain the UE connection*, then the UE will maintain its connection with the current serving BS.

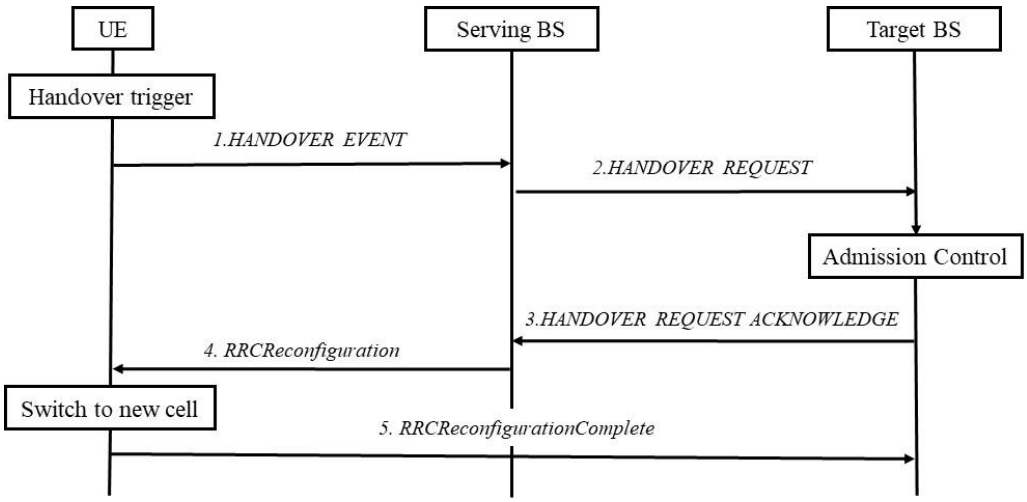


Fig. 3 The signalling after handover triggering

5. PERFORMANCE ANALYSIS

5.1 Analysis design

A simulation environment was built using MATLAB to test the mobility robustness of the UE under the proposed triggering mechanism. The environment designs are illustrated in Table 1. There are 16

small BSs and 2 macro BS deployed in a 1000 m ×1000 m scenario, and each small BS is approximately 350 m apart. The macro BS is deployed on the diagonal of the simulated environment. The 40 UEs move randomly at a speed of 30 km/h in the proposed environment.

The RSRP, SINR, and transmission distance (d) modelled by Eqs. (1) and (2) are implemented as decision criteria for the proposed triggering mechanism. The average number of handovers (NOH) per UE, the probability of PPHO, the handover failure rate, throughput, and network latency calculated by Eqs. (3-8) are adopted as KPIs to test the effectiveness of the proposed algorithm as discussed in Section 2.

The 160000 sets of data for each input metric are collected from the simulation environment to obtain a well-trained Q table. The 160000 sets cover all of the physical locations and are collected from all of the BSs deployed in this environment. The final trained Q table is utilised as the proposed handover triggering mechanism to be evaluated. There are two comparative approaches adopted, that is, traditional A3 event RSRP-based [2] and fuzzy logic-based triggering mechanisms [17] and [18]. As previously mentioned, the A3 event is based only on a single metric, that is, the RSRP that triggers the handover process. The fuzzy logic-based approach in this study also considers the RSRP, SINR, and d as input metrics. The approach based on Q-learning without clustering techniques (with states of equal lengths) is also adopted as a comparison group to show the advantage of this clustering technique. These four approaches work in the same test environment.

Eq. (18) is utilised to quantify the improvement of each KPI (ΔKPI) under the proposed approach. KPI_1 means the evaluated KPI value under method 1, and KPI_2 denotes the same logic. Each KPI is tested for at least 100 rounds in this study to ensure reliable evaluation results.

$$\Delta KPI = \frac{KPI_1 - KPI_2}{KPI_2} \ \% \tag{18}$$

Table 1 Simulation parameters

Parameters	***Specification***	
	Macro BS	**Small BS**
Carrier frequency (GHz)	1.5~2	28
Subcarrier spacing (KHz)	15	30
System bandwidth (MHz)	20	100
Physical resource block	100	275
Number of BSs	2	16
BS transmitted power (dBm)	49	35
Subcarriers per PRB	12	
Duration of simulation	10000 s	

Mobility model	Random direction
Number of UE	40
UE speed (km/h)	30
Propagation model:	Eq. (1)
KPIs	Eqs. (3-8)
Type of noise	AWGN, Rayleigh
Handover preparation time (ms)	10ms
Handover execution time (ms)	10 ms

5.2 Results and analysis of comparison experiments

Fig. 3 shows the training stage of Q-learning and accumulated rewards in each epoch with and without subtractive clustering. In the same training environment, the Q-learning with clustering approach (black solid line) converges at the 70th epoch, and the Q-learning only approach (green dash line) converges at the 20th epoch. After convergence, the accumulated Q-learning reward that optimised by subtractive clustering is approximately 3900, and the Q-learning only approach can receive 3400. Based on Eq. (18), the approach based on Q-learning with clustering can receive approximately 15% more rewards than the Q-learning only approach. The trained Q tables from both methods are utilised as the handover triggering mechanisms of the UE to be evaluated. The performance of these two mechanisms is shown in Fig. 5-9.

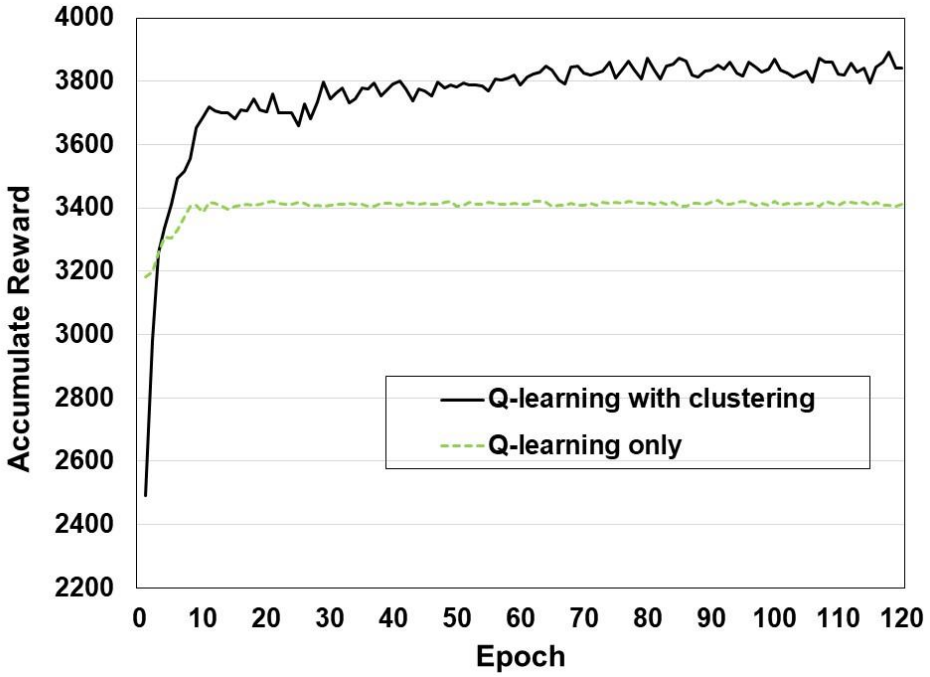


Fig. 4 Accumulated rewards in each epoch

The KPIs in Fig. 5-7 are used to evaluate the mobility robustness of the UE based on different triggering mechanisms. According to the results in Fig. 5-6, the A3 event RSRP-based triggering mechanism (blue line with circle) has the highest NOH (4250) and PPHO ratio (0.24%), as it depends on only a single metric, the RSRP, to trigger the handover process. The RSRP fluctuates due to noise and interference, which can significantly reduce the triggering decision's accuracy. The fuzzy logic (orange line with triangular) can consider multiple metrics as a decision criterion and thus has a lower NOH (2500) and PPHO ratio (0.21%) than the RSRP-based approach. Since the Q-learning-based approach has powerful learning capability and also considers multiple metrics in decision-making, the two proposed Q-learning-based approaches have the best performance. Based on the results in Fig. 5-6 and Eq. (18), the Q-learning only approach (green dash line) can significantly reduce 70–90% of handovers and approximately 60% of PPHO ratios compared with the RSRP and fuzzy logic-based approaches. Moreover, the adoption of clustering (black solid line) can reduce another 22% of NOH and 20% of PPHO ratios compared with the Q-learning only approach.

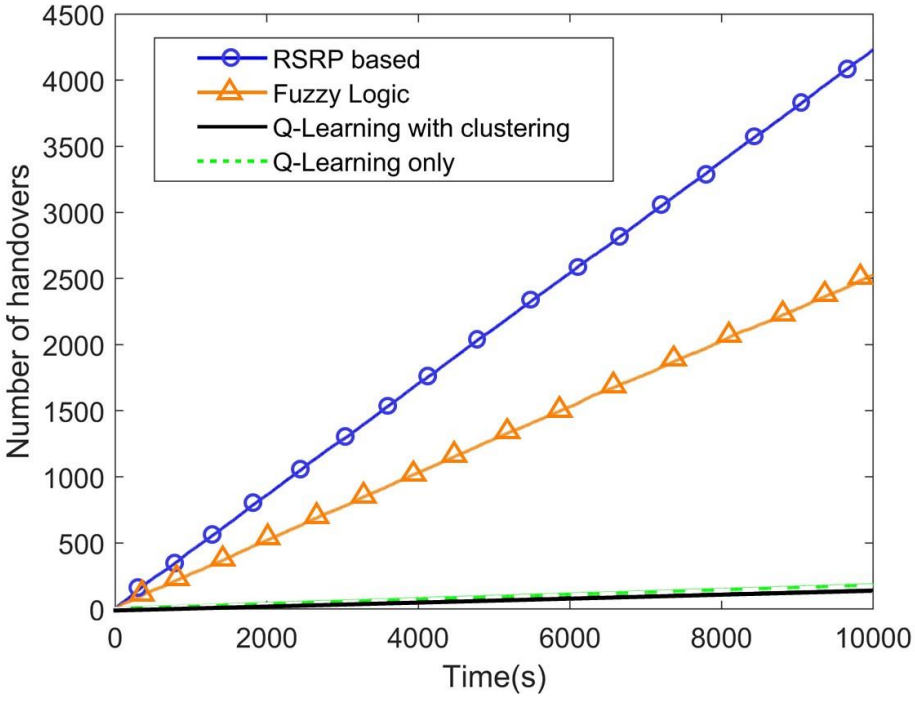


Fig. 5 Number of handovers by the different triggering mechanisms

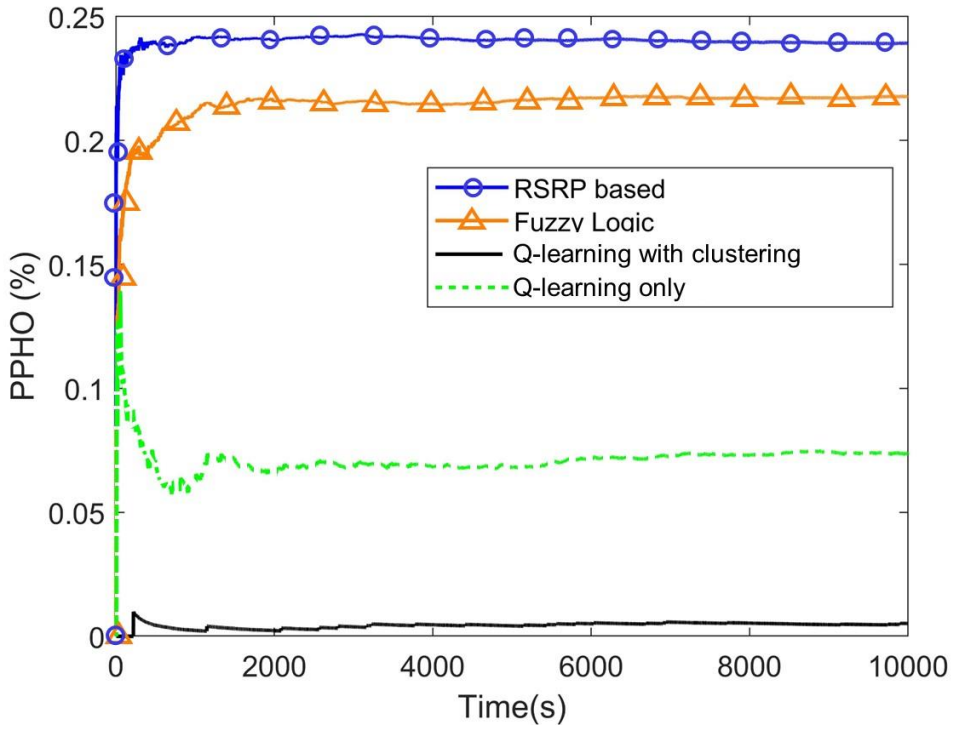


Fig. 6 PPHO of the different triggering mechanisms

As indicated in Fig. 7, the A3 event RSRP-based triggering mechanism (blue line with circle) has the second-lowest handover failure rate at 0.5%. This is because the RSRP is a key factor in determining the handover failure rate, and the A3 event RSRP-based approach will continue switching the UE connection to the neighbouring BS with the highest RSRP. As such, the A3 event RSRP-based approach can ensure a low handover failure rate. The fuzzy logic (orange line with triangular) and Q-learning only approaches (green dash line) have relatively high handover failure rates of 5% and 1%, respectively. These two approaches consider multiple metrics to trigger the handover and weaken the weight of the RSRP in decision-making. The fuzzy membership functions used in fuzzy logic are not well designed for each input metric, which can cause the input to incorrectly convert to the corresponding level. The input metrics are also not well categorised into state vectors in the Q-learning only approach, which can degrade the effectiveness of the trained Q table. Therefore, these two factors degrade the handover failure rate. However, Q-learning with clustering (black solid line) outperforms the three other approaches, with a nearly zero handover failure rate of 0.1%. Compared with the Q-

learning only approach (green dash line), the adoption of subtractive clustering can reduce the handover failure rate by approximately 75% based on Eq. (18).

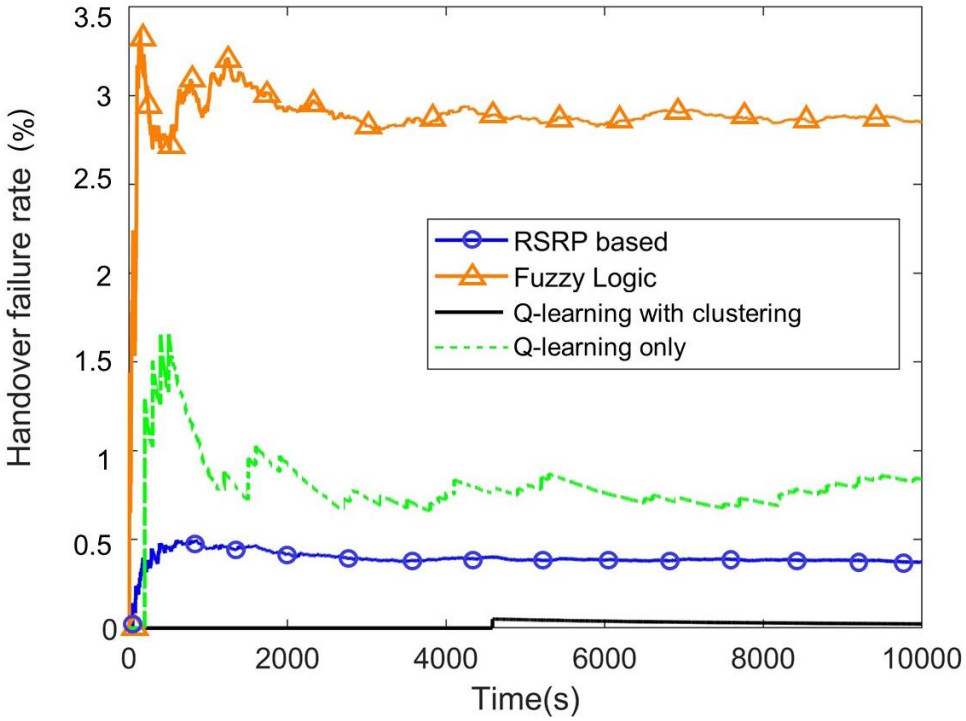


Fig. 7 Handover failure rates of the different triggering mechanisms

The KPIs in Fig. 8-9 are used to evaluate the quality of service (QoS). Some existing approaches focus only on the improvement in mobility robustness but result in a degradation of other aspects, such as load balancing and QoS. Thus, the objective of this study is to increase the mobility robustness while maintaining the other KPIs at relatively high levels.

Fig. 8 shows the network latency under the different handover triggering mechanisms. The A3 event RSRP and fuzzy logic-based approach have relatively high network latency of 16.7 ms and 15.2 ms, respectively. These two methods lead to many unnecessary handovers, which causes the accumulation of handover latency. The Q-learning only triggering mechanism also has a relatively high latency of 14.3 ms. Although the Q-learning only approach has fewer handovers, they primarily occur at the edge of coverage. This can result in a high propagation latency. Q-learning with clustering outperforms the other three approaches again and has the lowest average network latency of 10.1 ms. Compared with the Q-learning only approach, subtractive clustering can further reduce handover latency by approximately 27.3%.

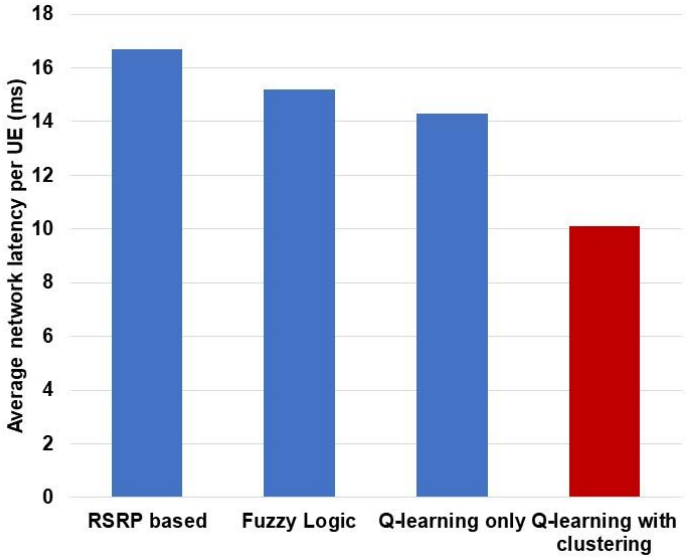


Fig. 8 Average network latency of the different triggering mechanisms

As shown in Fig. 9, the RSRP-based approach has the highest sum throughput because the RSRP is also one of the key factors determining the system throughput. As such, fuzzy logic has the lowest throughput as it considers other metrics during decision-making. The throughput of the two Q-learning-based approaches is slightly lower than the traditional methods but remains at a relatively high level. Compared with the Q-learning only approach, subtractive clustering can increase throughput by approximately 9.7%.

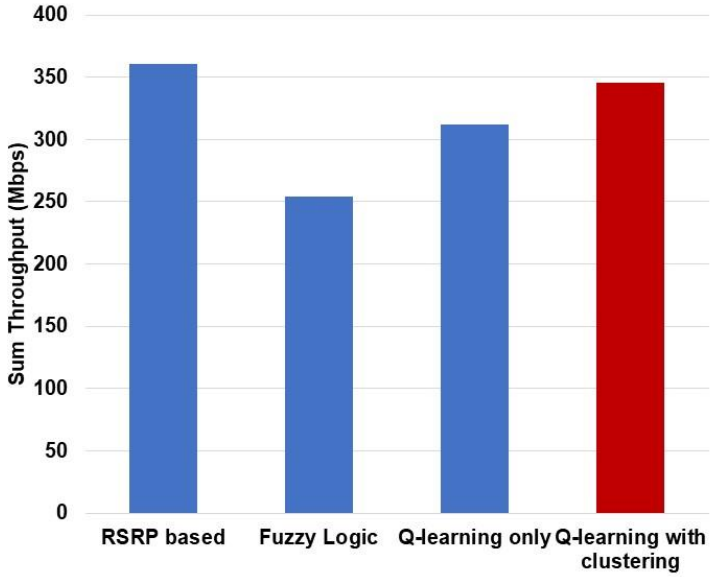


Fig. 9. Total throughput of the different triggering mechanisms

Based on the simulation results, the proposed Q-learning with clustering-based approach outperforms the other three approaches in terms of the NOH, PPHO ratio, handover failure rate, and network latency while maintaining a relatively high level of system throughput. This good performance is due to the following advantages: first, the Q-learning framework has a powerful ability to learn the optimal policy from different environments. After obtaining the optimal policy, the trained Q table executes the action with the maximum reward based on the states it faces. Second, because subtractive clustering has a strong ability to process uncertain and imprecise information, the adoption of clustering minimises the effect of noise and inference during decision-making. Moreover, subtractive clustering can locate clusters for each input metric from the historical data. This approach ensures that input metrics can systematically be categorised as state vectors with respect to their actual data distribution. Therefore, the subtractive clustering technique ensures the accuracy and effectiveness of Q-learning to achieve training targets. The trained Q table can precisely and intelligently trigger the handover process based on the states it faces.

6. CONCLUSION

In this study, we proposed an intelligent handover triggering mechanism based on the Q-learning and subtractive clustering techniques to address the challenges of handovers in 5G-UDNs. In the proposed framework, Q-leaning can learn the optimal triggering policy from different application scenarios. The proposed framework's trained Q table can be used in UE to precisely trigger the handover process based on the RSRP, SINR, and transmission distance. To further enhance the proposed approach's performance, we adopted subtractive clustering to ensure accuracy and effectiveness of the training process. According to the simulation results, compared with the A3 event RSRP-based and fuzzy logic-based approaches, the proposed solution can effectively reduce the NOH, PPHO ratio, and handover failure rate while maintaining a high level of network latency and system throughput. The evaluation also indicated that the adoption of subtractive clustering techniques can further enhance the proposed approach's performance by approximately 20% in terms of all of the evaluated KPIs. Moreover, the proposed solution has a low maintenance cost, as it can intelligently trigger the handover process without any additional handover parameters or conditions such as HHM and TTT.

An effort will be made in future works to develop energy efficient handover mechanism based on machine learning technique, which should reduce power consumption while retaining the mobility robustness for UE.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from the International Doctoral Innovation Centre (IDIC), Ningbo Education Bureau, Ningbo Science and Technology Bureau, and the University of Nottingham. This study was also supported by the Ningbo Natural Science Programme, project code 2018A610095.

REFERENCES

[1] D. Calabuig *et al.*, "Resource and Mobility Management in the Network Layer of 5G Cellular Ultra-Dense Networks," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 162–169, 2017.

[2] the 3GPP Organizational Partners, *Radio Resource Control (RRC) protocol specification, document TS 38.331*. 2018.

[3] Y. Li, B. Cao, and C. Wang, "Handover schemes in heterogeneous LTE networks: challenges and opportunities," *IEEE Wirel. Commun.*, vol. 23, no. 2, pp. 112–117, Apr. 2016.

[4] Y. Zhou, C. Shen, and M. Van Der Schaar, "A Non-stationary online learning approach to mobility management," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 2, pp. 1434–1446, 2019.

[5] H. S. Park, Y. Lee, T. J. Kim, B. C. Kim, and J. Y. Lee, "Handover Mechanism in NR for Ultra-Reliable Low-Latency Communications," *IEEE Netw.*, vol. 32, no. 2, pp. 41–47, 2018.

[6] M. Tayyab, X. Gelabert, and R. Jantti, "A Survey on Handover Management: From LTE to NR," *IEEE Access*, vol. 7, no. 1, pp. 118907–118930, 2019.

[7] M. T. Nguyen, S. Kwon, and H. Kim, "Mobility Robustness Optimization for Handover Failure Reduction in LTE Small-Cell Networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4672–4676, 2018.

[8] A. Alhammadi and G. S. Member, "Auto Tuning Self-Optimization Algorithm for Mobility Management in LTE-A and 5G HetNets," *IEEE Access*, vol. 8, pp. 294–304, 2020.

[9] K. Da Costa Silva, Z. Becvar, and C. R. L. Frances, "Adaptive Hysteresis Margin Based on Fuzzy Logic for Handover in Mobile Networks With Dense Small Cells," *IEEE Access*, vol. 6, pp. 17178–17189, 2018.

[10] Z. Xiong, Y. Zhang, D. Niyato, R. Deng, P. Wang, and L. C. Wang, "Deep reinforcement learning for mobile 5G and beyond: Fundamentals, applications, and challenges," *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 44–52, 2019.

[11] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, and S. Serikawa, "Motor anomaly detection for unmanned aerial vehicles using reinforcement learning," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2315–2322, 2018.

[12] H. Lu, Q. Liu, D. Tian, Y. Li, H. Kim, and S. Serikawa, "The Cognitive Internet of Vehicles for Autonomous Driving," *IEEE Netw.*, vol. 33, no. 3, pp. 65–73, 2019.

[13] Y. Zhang, Y. Li, R. Wang, M. S. Hossain, and H. Lu, "Multi-Aspect Aware Session-Based Recommendation for Intelligent Transportation Services," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–10, 2020.

[14] C. Lee, H. Cho, S. Song, and J. Chung, "Prediction-Based Conditional Handover for 5G mm-Wave Networks: A Deep-Learning Approach," *IEEE Veh. Technol. Mag.*, vol. 15, no. 1, pp. 54–62, Mar. 2020.

[15] S. Chaudhuri, I. Baig, and D. Das, "Self organizing method for handover performance optimization in LTE-advanced network," *Comput. Commun.*, vol. 110, pp. 151–163, 2017.

[16] T. Goyal and S. Kaushal, "Handover optimization scheme for LTE-Advance networks based on AHP-TOPSIS and Q-learning," *Comput. Commun.*, vol. 133, no. September 2018, pp. 67–76, 2019.

[17] S. S. Mwanje, L. C. Schmelz, and A. Mitschele-Thiel, "Cognitive Cellular Networks: A Q-Learning Framework for Self-Organizing Networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 13, no. 1, pp. 85–98, 2016.

[18] K. Vasudeva, S. Dikmese, I. Guvenc, A. Mehbodniya, W. Saad, and F. Adachi, "Fuzzy logic game-theoretic

approach for energy efficient operation in HetNets," *2017 IEEE Int. Conf. Commun. Work. ICC Work. 2017*, pp. 552–557, 2017.
[19] K. C. Silva, Z. Becvar, E. H. S. Cardoso, and C. R. L. Francês, "Self-tuning handover algorithm based on fuzzy logic in mobile networks with dense small cells," *IEEE Wirel. Commun. Netw. Conf. WCNC*, vol. 2018-April, pp. 1–6, 2018.
[20] C. F. Kwong, T. C. Chuah, S. W. Tan, and A. Akbari-Moghanjoughi, "An adaptive fuzzy handover triggering approach for Long-Term Evolution network," *Expert Syst.*, vol. 33, no. 1, pp. 30–45, 2016.
[21] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover Control in Wireless Systems via Asynchronous Multiuser Deep Reinforcement Learning," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4296–4307, 2018.
[22] M. S. Chen and S. W. Wang, "Fuzzy clustering analysis for optimizing fuzzy membership functions," *Fuzzy Sets Syst.*, vol. 103, no. 2, pp. 239–254, 1999.
[23] M. Alhabo and L. Zhang, "Multi-criteria handover using modified weighted TOPSIS methods for heterogeneous networks," *IEEE Access*, vol. 6, no. c, pp. 40547–40558, 2018.
[24] the 3GPP Organizational Partners, *Study on channel model for frequencies from 0.5 to 100 GHz,document TR 38.901*. 2019.
[25] Y. Bai and L. Chen, "Hybrid spectrum arrangement and interference mitigation for coexistence between LTE macrocellular and femtocell networks," *Eurasip J. Wirel. Commun. Netw.*, vol. 2013, no. 1, 2013.
[26] K. Vasudeva, M. Simsek, D. Lopez-Perez, and I. Guvenc, "Analysis of Handover Failures in Heterogeneous Networks with Fading," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6060–6074, 2017.
[27] K. E. Suleiman, A. E. M. Taha, and H. S. Hassanein, "Handover-related self-optimization in femtocells: A survey and an interaction study," *Comput. Commun.*, vol. 73, pp. 82–98, 2016.
[28] F. Jiang, Z. Yuan, C. Sun, and J. Wang, "Deep Q-Learning-Based Content Caching With Update Strategy for Fog Radio Access Networks," *IEEE Access*, vol. 7, pp. 97505–97514, 2019.