Machine Learning based RATs Selection supporting Multi-Connectivity for Reliability (Invited Paper) *

Haeyoung Lee, Seiamak Vahid, and Klaus Moessner

5G Innovation Centre (5GIC), Institute for Communication Systems (ICS),
University of Surrey, Guildford, GU2 7XH, U.K.
{Haeyoung.Lee,S.Vahid,K.Moessner}@surrey.ac.uk

Abstract. While ultra-reliable and low latency communication (uRLLC) is expected to cater to emerging services requiring real-time control, such as factory automation and autonomous driving, the design of uRLLC of stringent requirements would be very challenging. Among novel solutions to satisfy uRLLC's requirements, interface diversity is widely regarded as an efficient enabler of ultra-reliable connectivity. When mobile devices are connected to multiple base stations (BSs) of different radio access technologies (RATs) and same data is transmitted via multiple links simultaneously, the transmission reliability can be improved. However, duplicate transmission of same data causes an increase in the traffic loads, leading to radio resource shortage. Considering it, efficient configuration of multi-connectivity (MC) for mobile devices is important. In this paper, the RAT selection scheme including efficient MC configuration is proposed. By adopting distributed reinforcement learning (RL), each device could learn the policy for efficient MC configuration and select appropriate RATs. Simulation results show that 20.8% reliability improvements over the single connectivity scheme is observed. Comparing to the method to configure MC for devices all the time, 37.6% improvement is achieved at high traffic loads.

Keywords: RAT selection · Multi-connectivity· Machine Learning · URLLC.

1 Introduction

Upcoming 5G networks are expected to support diversified services into three categories: enhanced mobile broadband (eMBB), massive machine-type communication (mMTC), and ultra-reliable and low latency communication (uRLLC) [1]. Especially, it is envisaged uRLLC could open the doors emerging various services, such as wireless control and automation in industrial environments [2], vehicle-to-vehicle communications, and the tactile internet, which requires to

* The work presented in this paper is partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 761745, and the Government of Taiwan.

control many objects with real-time feedback [3]. For such services, the 3GPP aims at providing uRLLC for small data payloads (e.g., 32 bytes) with an outage probability of less than 10^{-5} at millisecond level latency. While the design of uRLLC of stringent requirements would be very challenging, various novel solutions have been proposed such as flexible frame structure design with shorter transmission time intervals (TTIs) [4], pre-emptive scheduling [5], and diversity for reliability improvement [6]. Especially, diversity is widely regarded as a crucial and efficient enabler of ultra-reliable connectivity [7].

As the network evolves, multiple radio access technologies (RATs) are being integrated and jointly managed, including 3GPP and IEEE families, with the vision of heterogeneity [8]. In addition, as cells are deployed closer and more heterogeneous, multiple links of different RATs would become available to user equipments (UEs) at the same time. Based on such availability of multiple links, multi-connectivity (MC) is expected to offer enough diversity and redundancy for achieving reliability [9]. Actually, the initial goal of using MC was to improve throughput performance by splitting its traffic and sending over multiple links, overcoming the capacity limitations imposed by backhaul links. By adopting packet duplication (PD) for duplicate transmission of same data [10], MC has been additionally considered as an effective solution to satisfy the stringent reliability requirement.

In heterogeneous networks (HetNets) of multiple base stations (BSs) from different RATs, for each UE, the matter to decide the suitable RATs impact on the network performance including reliability. In [11, 12], the impact of the number of BSs involving multi-connectivity (called an active set) is investigated. In [11], it is shown that dynamic management of the active set of BSs (i.e., adding, removing, replacing based on thresholds of signal quality) can improve system performance in terms of radio link failure (RLF) and throughput compared to the use of fixed active set of BSs. In [12], the network load is also considered to decide the number of BSs. The impact of the network load to the effectiveness of multi-connectivity is investigated in [13]. With the assumption that UEs can access all BSs of SINR (signal to interference and noise ratio) higher than the pre-defined threshold, MC is proved more effective at the low traffic scenario in terms of throughput and RLF performance. In [14], the dual connectivity (DC) architecture is considered. The optimisation problem on RAT selection to maximize the sum throughput is formulated and the pair of serving macro and small cell for each UE is found. In [15], the utility based approach is studied considering user's satisfaction as well as service provider's satisfaction in terms of throughput. While the aforementioned works focus on improvement of mobility robustness or throughput performance, in [16], the reliability improvement through DC with data duplication is demonstrated. By the simulation results, it shows the required level of reliability and network traffic loads should be considered for each UE's DC configuration. However, how to configure DC for each UE is not investigated.

Considering the literature survey, in this paper, the approach on efficient RATs selection including MC configuration is proposed. Since the characteristics of UEs including the location and required QoS level could be different from each other in the network with dynamically changing traffic loads, the proposed algorithm employs distributed reinforcement learning. Each UE becomes an agent and learns the policy of RATs selection including MC configuration. While each UE learns the offset for effective MC configuration individually, appropriate RATs for UEs could be selected to minimizes the number of UEs in outage. It is shown that the proposed algorithm outperforms the single connectivity based RAT selection scheme by 20.8% in terms of reliability performance. Comparing to the RATs selection scheme where MC is always configured for all UEs, the proposed algorithm shows better performance by 37.6% by configuring MC only for UE at cell edge region.

The organization of the paper is as follows. Section 2 describes the system model including the architecture supporting MC. In Section 3, the proposed RATs selection algorithm adopting the distributed reinforcement learning mechanism is presented. Then, its performance is evaluated in Section 4. This paper is concluded in Section 5.

2 System Model

We consider packet duplication (PD) exploiting the multi-connectivity (MC) feature to improve reliability performance. With MC, a UE is able to connected to multiple BS. Since MC is an extension of dual-connectivity (DC), for simplicity, DC is considered in this paper as shown in Fig. 1. When UEs are connected to BSs, one BS acs as the master node (MN) to establish the control interface to the core network and another BS becomes the secondary node (SN). The MN and SN are assumed to be interconnected by means of Xn interface. While DC can be applicable only for UEs in Radio Resource Control (RRC) connected

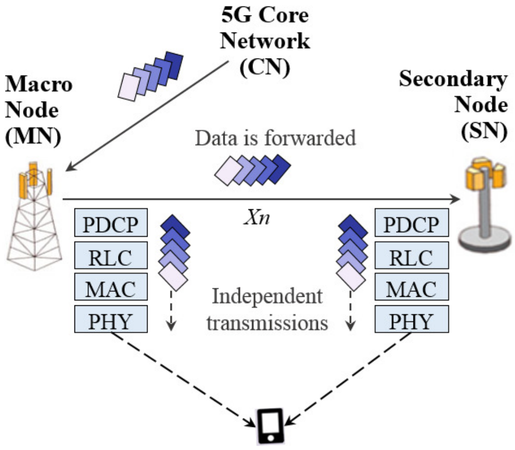

Fig. 1. Packet duplication via multi-connectivity to enhance reliability.

mode, MN can initiate DC setup. The data transferred from the core network to MN is duplicated at the packet data convergence protocol (PDCP) layer of MN and the entire data is forwarded to SN via Xn interface. Thus, the same data becomes to be transmitted via both MN and SN to the UE. For resource scheduling, both MN and SN have flexibility and no restrictions are imposed at the RLC and MAC layer. The lower layers at MN and SN work independent of each other without coordination [10]. Such data duplication process can be carried out as long as the UE remains into the coverage area of both nodes.

As depicted in Fig. 2, the downlink transmission of the OFDM-based heterogeneous cellular network with multiple UEs is considered. The system consists of a set of BSs including macro BSs (MBSs) and small cell BSs (SBSs) and a set of users (UEs) capable to connect to multiple networks simultaneously. While macro BSs and small cell BSs could support different RATs, they can communicate based on Xn interface for DC [9]. Based on reference signal received power (RSRP) based cell selection, UEs measure the reference signal power from each BS, and could be connected to either the largest one or two BSs. When UEs are located close to the serving BS (e.g., UE 1, UE 2, UE 4 and UE 5), a strict reliability requirement could be met with a single highly reliable link of the strong signal power. For cell-edge UEs (e.g., UE 3), the received strongest signal level would not be strong enough to fulfil their QoS requirements. In this case, multi-connectivity employing PD could be set up to improve the reliability so that UEs can be connected to two BSs, one from macro BS and another from small cell BS. Then, the macro BS becomes master node (MN) while small cell BS becomes secondary node (SN).

While the channel quality is an important factor to decide MC configuration of UEs, the network traffic load could also impact on the reliability performance. Although MC contributes to enhancing UEs' reliability, inrease in the number of UEs exploiting MC configuration would lead to increase in network traffic load [16]. Configuration of MC for too many UEs will shed the light on the benefit of

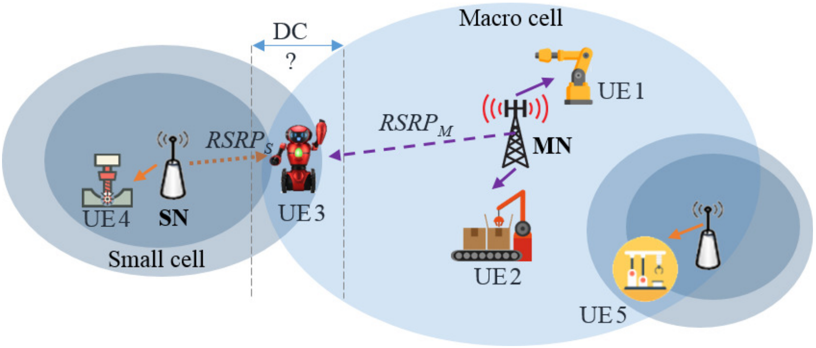

Fig. 2. Downlink transmission with multiple connectivity in heterogeneous networks.

MC. Thus, in order to decide MC configuration for UEs, the network traffic load needs to be considered as well as UEs' QoS requirements and channel quality.

In this paper, RATs selection based on MC configuration is investigated and the scenario of dual-connectivity (DC) is focused. Based on RSRP values from macro and small cell BSs, UEs's connection can be determined as follows.

- Connect to macro BS (MBS) for $rsrp_M \geq rsrp_S + \beta$
- Connect to small cell BS (SBS) for $rsrp_M \leq rsrp_S - \beta$
- Connect to MBS and SBS by DC for $rsrp_S + \beta > rsrp_M > rsrp_S - \beta$

Here, $rsrp_M$ and $rsrp_M$ denote RSRPs from MBS and SBS, respectively. In the case the cell range extension (CRE) bias is given, $rsrp'_M = rsrp_M - CRE$ can be considered instead of $rsrp_M$. In order to provide DC configuration to UEs, a DC offset value β is considered with RSRP values. Optimal DC offsets β could be changed by various factors, such as the available radio resource among BSs and by the location of UEs and BSs. Since the optimal offset values vary from one UE to another, offset values could be defined by each UE [20].

3 RATs selection with Reinforcement Learning

While the Reinforcement Learning (RL) mechanism uses experiences of agents and could§ learn automatically from the environment without any training data on field, it allows an online learning. In our work, Q-learning is chosen since it enables learning the best policy without any priori knowledge of its environment.

In RL, at time epoch t, the agent in the state s_t selects and performs an action a_t. After the action a_t, it observes the environment and receives a reward cost C for this specific action. While RL accumulates costs obtained by action, it considers instant cost as well as cumulative costs in the future. With learning, it is aimed to find the optimal policy for selecting an action in a given state that minimizes the value of total cost. In Q-learning, in order to learn this policy, an agent utilises a value-function, Q-function, $Q(s_t, a_t)$. It is defined as follows [21]:

$$Q(s,a) = E\left\{ \sum_{t=0}^{\infty} \gamma^t \, C(s_t, a_t) \mid s_0 = s, a_0 = a \right\}, \tag{1}$$

where γ, $C(s_t, a_t)$, s_0 and a_0 denote a discount factor ($0 \leq \gamma \leq 1$), the cost of the set of state s_t and action a_t, intial state, and initial action, respectively.

While it is really difficult to obtain the optimal policy by solving (1), RL could be exploited to find the optimal policy by using Q-table updates. In Q-table, each table entry, $Q(s_t, a_t)$, is associated with a state-action pair and the Q-learning algorithm maintains Q-table of values that represent the goodness of taking a particular action when in a given state. It is enough to converge this learning if all Q-values of the sets of states and actions are continued to be updated. Q-learning realizes (1) by updating Q-table as follows.

$$Q(s_t, a_t) \leftarrow (1-\rho) Q(s_t, a_t) + \rho[C_{t+1} + \gamma \min_{a \in A} Q(s_{t+1}, a)], \tag{2}$$

where ρ is the learning rate of the range $0 \leq \rho \leq 1$ indicating what extent the learned Q-value will override the old one. When $\rho = 0$, the agent never learns. When $\rho = 1$, the new knowledge of the most recent Q-value is only considered. C_{t+1} represents the delayed cost, which is obtained for an action a_t taken. As the value of the discount factor γ in [0,1] is higher, the future cost $\min_{a \in A} Q(s_{t+1}, a)$ is weighted more than the delayed cost C_{t+1}. By updating Q-table in (2), the agent learns the optimal policy for selection an action.

In this paper, the state, the action, and cost are defined as follows.

- **State**: The state of time epoch t is defined with the received power from BSs as:

$$s_t = \{rsrp_M,\ rsrp_S\} \text{ where } s_t \in S, \tag{3}$$

  where $rsrp_M$ and $rsrp_S$ denote the reference signal received power (RSRP) from MBS and SBS, respectively. When there are multiple MBSs and SBSs, one MBS and SBS of the strongest RSRP can be selected. To make Q-table small and to convergence faster, two power values are quantized. S denotes the set of all states.

- **Action**: The action of time epoch t is defined as:

$$a_t = b_i \text{ where } b_i \in A \tag{4}$$

  where b_i denotes the DC configuration offset value β and A is the set of all possible offset values (i.e., all possible actions).

- **Cost**: The cost of time epoch t is defined as:

$$c_t = n, \tag{5}$$

  where n denotes the number of UEs in outage.

Each UE monitors the level of RSRP from BSs and selects one MBS and SBS of the strongest signal power. In other words, each UE observes its state. The received power value is quantized to manage Q-table size small and to convergence faster and each UE compares these quantized signal powers with its Q-table's states. If the UE cannot find the received powers from its Q-table, the new state of received powers is added to Q-table. Among those sets whose received powers are equal to the received powers, UEs can choose an action a_t based on ε-greedy exploration and exploitation policy [21]. In ε-greedy policy, at every decision epoch, a UE in state s_t explores with probability $\varepsilon(s_t)$, and stored Q-values is exploited with probability $1 - \varepsilon(s_t)$ as follows.

$$a_t = \begin{cases} \min_{a \in A} Q_t(s_{t+1}, a) & , \text{ probability } 1 - \varepsilon(s_t) \\ rand(a) & , \text{ probability } \varepsilon(s_t). \end{cases} \tag{6}$$

The exploration rate $\varepsilon(s_t)$ is defined by using $\lambda(s_t, a_t)$ which is the number of visits of state-action pair (s_t, a_t), as follows.

$$\varepsilon(s_t) = \frac{1}{\log\big(\sum_{a_t \in A} \lambda(s_t, a_t) + 3\big)}. \tag{7}$$

In (7), $\varepsilon(s_t)$ in $(0, 1)$ has a logarithmic decay. This approach aims to control the frequency of exploration so that the best-known action is taken at most of the times. Exploring is not stopped to enhance the long-term learning performance, but rather decreased gradually over time. For convergence, the learning rate $\rho(s_t, a_t)$ is set by using $\lambda(s_t, a_t)$ as follows.

$$\rho(s_t, a_t) = \frac{1}{\sqrt{\lambda(s_t, a_t) + 3}}. \tag{8}$$

According to above definition, each UE decides the appropriate offset value for MC configuration that minimize the number of UEs in outage. Then, each UE is connected to selected RATs by comparing RSRPs from BSs with the MC configuration offset. Afer BSs allocate resource to UEs, BSs calculate the number of UEs in outage UEs and send information to UEs. For resource allocation, resource block (RB), the block of subcarriers, is considered as the basic radio resource unit [9] and it is assumed that one RB is allocated to each UE.

```
Result: $\beta$, the offset value for MC configuration for each UE
Initialisation: Q-table with a very high number;
Learning procedure: while Q-table converges do
    1. UE selects the MBS and SBS of the strongest signal power;
    2. UE compares the (quantized) received powers with Q-table's states;
    if no equal received powers on Q-table then
        3. UE adds new received powers to its own Q-table;
    end
    4. Calculate $\varepsilon(s_t)$ and generate a random number $\delta$ in [0,1];
    if $\delta \leq 1 - \varepsilon(s_t)$ then
        5. UE chooses one value that has the lowest Q-value;
    else
        6. UE chooses one value randomly;
    end
    7. UE uses chosen offset value $\beta$ as an action;
    8. UE compares $rsrp_M$ and $rsrp_S$ added by $\beta$;
    10. UE decides RATs to be connected;
    11. BSs allocate RBs to each UE;
    12. BSs calculate the number of outage UEs and pass it to UEs;
    13. Each UE updates the chosen set's Q-value based on (2).
end
```

Algorithm 1: The proposed RAT selection algorithm

After resource allocation, BSs send the information on the number of UEs in outage to UEs, and each UE updates Q-value based on (2).

The procedures of the proposed algorithm are explained in Algorithm 1. While Step 1 to 10 and 13 are conducted by the UE side, Step 11 and 12 are carried out by the BS side. Repeating the above steps makes Q-table values of all sets of states and actions converge, and then agents can make right actions.

4 Performance Evaluation

We evaluate the performance of the proposed RAT selection algorithm via simulation. Table 1 shows the initial configuration parameters. In order to compare the performance of the proposed algorithm, two reference schemes are considered: 1) single connectivity based RAT selection (labeled 'SC') where one BS of the strongest RSRP is selected, and 2) dual connectivity based RATs selection (labeled 'DC-Always') where DC is always configured to all UEs. The average number of UEs in outage is considered as the performance indicator. Considering the practical scenario, traffiic of short message size [18] and path loss model for open production space is chosen [19]. Furthermore, as interval of DC configuration offset, we use 2 dB for Q-learning to make Q-table small. The maximum value of the offset is set to 32 dB, thus the actions have 17 levels.

Firstly, we investigate the impact of the DC offset value β on the reliability performance as depicted in Fig. 3. With two reference schemes, DC based algorithms are studied using different DC offset values of $4, 10, 20, 30$ dB. While all DC based algorithms produce better performance than the algorithm 'SC', it is observed that increase in a DC offset value contributes to enhancement of

Table 1. Simulation Parameters

Parameters	Values
No. of MBS/SBS	1/1
Transmit power	MBS: 10 dBm, SBS: 0 dBm
Carrier Frequency	3.5 GHz
Channel Bandwidth	5 MHz
No. RBs	25
Noise density	-174 dBm/Hz
No. UEs	2 0-30 (uniform dist.)
UE traffic [18]	40 bytes, 1 ms of message inter-arrial time
Path Loss model [19]	LoS: $32.45 + 20\log_{10}(d_{3D}) + 20\log_{10}(f) + X_\sigma, \sigma = 3$ nLoS: $32.45 + 24.7\log_{10}(d_{3D}) + 20\log_{10}(f) + X_\sigma, \sigma = 5.17$ (where $dref = 10\text{m}, K = dref^{1.5}, d_{3D} \geq 1\text{m}$) For $d_{2D} \leq dref$, $pr_{LoS} = 1$ For $d_{2D} > dref$, $pr_{LoS} = \left(-\frac{d_{2D}-dref}{K}\right)$
Resource allocation	Round Robin
DC offset value β	$[0, 2, ..., 32]$ dB

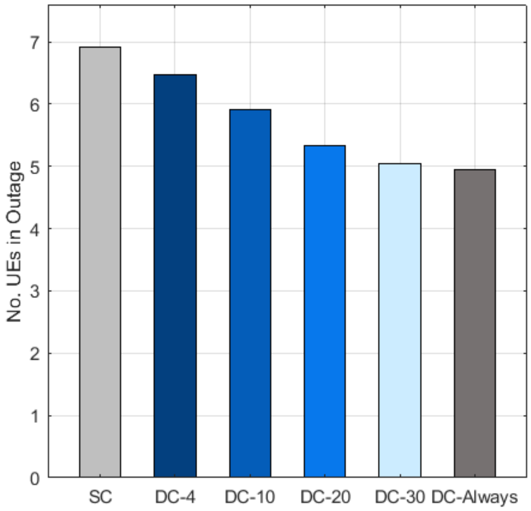

Fig. 3. Comparison of the average number of outage UEs from algorithms, SC, DC with various offset (4dB, 10dB, 20dB, 30dB), and DC-Always at low traffic loads

the reliability performance. However, the algorithm 'DC-Always' is shown to be superior. In this case, while 20 UEs uniformly distributed are assumed, BSs do not have difficulty in allocating RBs to all UEs. Since increase of traffic loads from DC configuration does not lead to overload BSs, configurating DC for all UEs could enhance their reliability.

Fig. 4 shows the performance of the proposed approach adopting distributed reinforcement learning with two reference schemes. While 30 UEs are considered in this simulation, the increase in data traffic from all UE's DC configuration can cause resource shortage which spoils the benefit in reliability. Thus, the 'DC-Always' scheme becomes insuperior to the 'SC' scheme by 26.9%. In the proposed algorithm, 'DC-Learning', while each UE learns the optimal policy in DC configuration depending on its location and the traffic loads, the algorithm tends to configure DC for UEs effectively. Compared to the 'SC' scheme, the proposed algorithm could achieve the gain of 20.8% in reliability. For larger number of UEs, the gap between the performance of 'SC', 'DC-Always', 'DC-Learning' could become more conspicuous.

5 Conclusion

In this paper, investigation into RATs selection in a multi-RAT network supporting multi-connectivity is provided. Considering different characteristics of UEs, distributed machine learning is applied so that each UE could learn the policy to configure MC and select appropriate RATs. With the simulation results, it is shown that the proposed approach is able to achieve better reliability

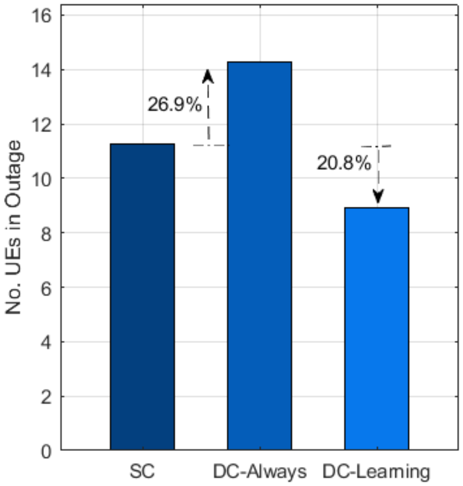

Fig. 4. Comparison of the average number of outage UEs from algorithms based on SC, DC-Always, and the proposed algorithm adopting ML at high traffic loads

performance compared to the single connectivity based RAT selection. While UEs could select MC configuration autonomously considering their characteristics and network traffic load, the proposed algorithm is shown to be superior to the mechamism to configure MC for all UEs all the time. In this paper, all UEs are assumed to have the same QoS requirements. In future research, multiple UEs of heterogeneous traffic will be considered with the resource allocation method to satisfy different QoS requirements of heterogeneous UEs.

References

1. ITU-R: IMT Vision - Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond. Rec. M.2083-0, Sept. (2015)
2. Clear5G, Deliverable 1.1 System Specifications and Business Perspectives. (2018) http://Clear5G.eu/
3. Pocovi, G. et al.: Achieving Ultra-Reliable Low-Latency Communications: Challenges and Envisioned System Enhancements. IEEE Network, 32(2), 8–15 (2018)
4. Pedersen, K. I. et al.: A flexible 5G frame structure design for frequency-division duplex cases. IEEE Commun. Mag., 54(3), 53-59 (2016)
5. Esswie, A. A. and Pedersen, K. I.: Opportunistic Spatial Preemptive Scheduling for URLLC and eMBB Coexistence in Multi-User 5G Networks. IEEE Access, vol. 6, 38451–38463 (2018)
6. Ji, H. et al.: Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects. IEEE Wireless Commun.ications, 25(3), 124–130 (2018)
7. Sutton, G. J. et al.: Enabling Technologies for Ultra-Reliable and Low Latency Communications: From PHY and MAC Layer Perspectives. IEEE Commun. Surv. Tut. (2019)

8. Andrews, J. G. et al.: What Will 5G Be? IEEE JSAC, 32(6), 1065–1082 (2014)
9. 3GPP: NR; Overall Description; Stage-2. Tech. Spec. 38.300, v15.3 (2018)
10. Rao, J. and Vrzic, S.: Packet Duplication for URLLC in 5G: Architectural Enhancements and Performance Analysis. IEEE Network, 32(2), 32–40 (2018)
11. Tesema, F. B. et al.: Evaluation of Adaptive Active Set Management for Multi-connectivity in Intra-frequency 5G Networks. In: Proc. IEEE Wireless Commun. and Networking Conf. (2016)
12. Ba, X. et al.: Load-Aware Cell Select Scheme for Multi-connectivity in Intra-frequency 5G Ultra Dense Network. IEEE Commun. Letters, 23(2), 354–357 (2019)
13. Alexandris, K. et al.: Utility-Based Resource Allocation under Multi-Connectivity in Evolved LTE. In: Proc. 86th IEEE Vehic. Tech. Conf. (2017)
14. Shi, Y. et al.: Dual Connectivity Enbled User Assocation Appraoch for Max-Throughput in the Downlink Heterogeneous Network. Wireless Personal Commun., 96(1), 529–542 (2017)
15. Escudero-Garzás, J. J. et al.: An Analysis of the Network Selection Problem for Heterogeneous Environments with User-Operator Joint Satisfaction and Multi-RAT Transmission. Wireless Commun. and Mobile Computing (2017)
16. Mahmood, N. H. et al.: Reliability Oriented Dual Connectivity for URLLC services in 5G New Radio. In: 15th Int'l Symp. on Wireless Commun. Systems (2018)
17. 3GPP: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2. TS 36.300, V14.4.0 (2017)
18. 3GPP: Study on Communication for Automatic in Vertical Domains. Technical Report 22.804 v16.2.0 (Rel 16) (2018)
19. 3GPP: Scenarios, Frequencies and New Field Measurement Results from two Operational Factory Halls at 3.5 GHz for various Antenna Configurations. Nokia, 3GPP TSG RAN WG1 Meeting, R1-1813177 (2018)
20. Kudo, T., Ohtsuki, T.: Cell range expansion using distributed Q-learning in heterogeneous networks. EURASIP Journal on Wireless Communications and Networking, 1(61), 1687–1499 (2013)
21. Sutton, R. S., Barto A. G.: Reinforcement learning - An introduction. 2nd edn. The MIT Press, 2017.