Fair Resource Reusing for D2D Communication Based on Reinforcement Learning

Fang-Chang Kuo[1], Hwang-Cheng Wang[1], Jia-Hao Xu[1], and Chih-Cheng Tseng[2(✉)]

[1] Department of Electronic Engineering, National Ilan University, Yilan, Taiwan
[2] Department of Electrical Engineering, National Ilan University, Yilan, Taiwan
tsengcc@niu.edu.tw

Abstract. Device-to-device (D2D) communications can improve the overall network performance, including low latency, high data rates, and system capability for the fifth generation (5G) wireless networks. The system capability can even be improved by reusing resource between D2D user equipment (DUE) and cellular user equipment (CUE) without bring harmful interference to the CUEs. A D2D resource allocation method is expected to have the characteristic that one CUE can be allocated with variable number of resource blocks (RBs), and the RBs can be reused by more than one CUE. In this study, Multi-Player Multi-Armed Bandit (MPMAB) reinforcement learning method is employed to model such problem by establishing preference matrix. A fair resource allocation method is then proposed to achieve fairness, prevent wasting resource, and alleviate starvation. This method even has better throughput if there are not too many D2D pairs.

Keywords: Device-to-Device (D2D) · Resource allocation · Reinforcement learning · Multi-Player Multi-Armed Bandit (MPMAB) · Dynamic resource allocation

1 Introduction

In order to meet mobile users' soaring demands, 5G wireless communication is projected to provide enormous data rates, extremely low power consumption and delay [2–4]. Due to the scarcity of available spectrum resources, D2D communication was proposed as a new paradigm to facilitate direct communication between two devices which are in each other's proximity. The direct communication can process without the intervention of an evolved Base Station (eNB) after the session is established, and these two devices are called a D2D pair. D2D communications can improve the overall network performance, including low latency, high data rates, and system capability [6, 7]. Several direct communication modes are defined in [8], including unicast, relay, groupcast, and broadcast. Moreover, the Third Generation Partnership Project (3GPP) is targeting the availability of D2D communication to become a competitive broadband communication technology for public safety networks used by first responders [8–10].

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2021
Published by Springer Nature Switzerland AG 2021. All Rights Reserved
Y.-B. Lin and D.-J. Deng (Eds.): SGIoT 2020, LNICST 354, pp. 45–53, 2021.
https://doi.org/10.1007/978-3-030-69514-9_5

1.1 Related Work

The most charming aspect of D2D communication is that it can promote the network capacity by spectrum reusing between D2D user equipment (DUE) and cellular user equipment (CUE) [11]. The resource reusing is based on the criteria that it would not bring harmful interference to the CUEs [12].

From the prospect of system operator, in general, we may want to maximize the overall network throughput by allocating sufficient resource blocks (RBs) to some UEs with better channel conditions. On the other hand, we may expect the RBs can be scheduled fairly so that RBs would not be wasted by UEs with better channel conditions, and that UEs with poor channel conditions would not be starved.

In [5], proportional fairness (PF) algorithm is applied on D2D resource allocation. However, the authors did not consider the data rate requirements of UEs. Here, in this paper, we will take the UEs' requirements into consideration. In [5], an interference-degree-based greedy heuristic resource allocation algorithm, which is called SDF_GRA here for convenience, is explored to maximize the number of admitted D2D pairs for each CUE. SDF_GRA utilizes Conflict Graph (CG) [13] to model network interferences between devices. In Sect. 5 we will compare the performance between SDF_GRA and our proposed scheme, which is called proposed_with_AAG here in this paper.

In recent years, machine learning is very popular. Recently, some researchers tried to apply machine learning on D2D resource allocation problems [14–16]. In [17], the D2D resource allocation problem was investigated by applying a kind of Reinforcement Learning method, namely Multi-player Multi-Armed Bandit (MPMAB). This paper indicated that MPMAB algorithm is suitable for improving the capacity of D2D communication.

Here in this paper, we also apply MPMAB to deal with the D2D resource allocation problem, and propose a scheme where the eNB tries best to fairly allocate RBs to guarantee the data rate requirements of D2D pairs.

1.2 Organization of This Paper

The rest of this paper is organized as follows. Section 2 describes the system model of D2D communication underlaying 5G networks. The evaluation of SINR and capacity is also described. Section 3 describes MPMAB learning algorithm, which would be employed to learn the preference of D2D pairs for reusing resource blocks with CUEs. Section 4 explains the principle of the proposed scheme. The performance of the proposed scheme is evaluated by being compared with SDF_GRA scheme in Sect. 5. Finally, we conclude the main results in Sect. 6.

2 System Model

This section describes the model under consideration, the used symbols, the used formulas for calculating SINR and link capacity.

The system model under consideration is that an eNB serving some CUEs and D2D pairs, as illustrated in Fig. 1. Each D2D pair has a dedicated transmitter device and a receiver device. The eNB knows the locations of all CUEs and DUEs and tries to allocate multiple D2D pairs to reuse the uplink RBs of each CUE based on underlay mode. The symbols are as follows.

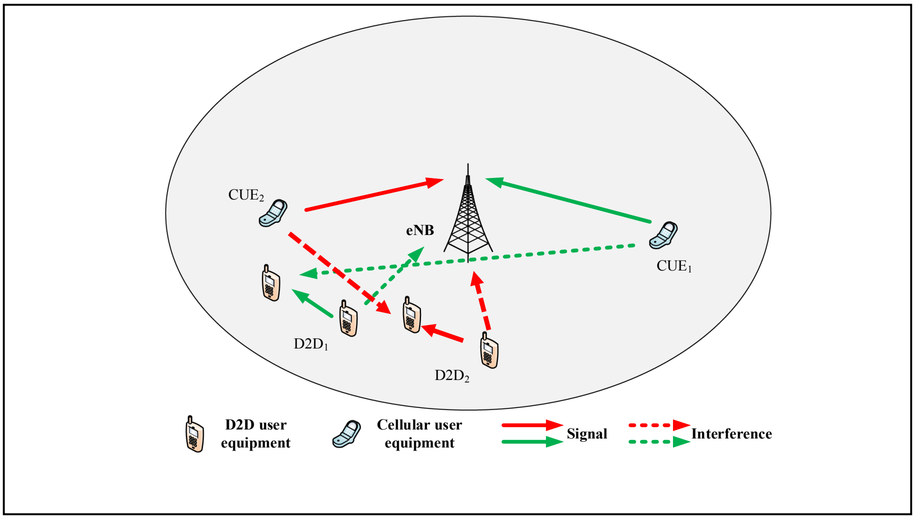


Fig. 1. System model

D_i: the i-th D2D pair with predetermined transmitting DUE and receiving DUE
$\mathrm{D}_{i,\mathrm{Tx}}$: the transmitting DUE of D_i
$\mathrm{D}_{i,\mathrm{Rx}}$: the receiving DUE of D_i
C_m: the m-th CUE
$G_{\mathrm{C}_m,\mathrm{B}}$: the channel gain from C_m to eNB
$G_{\mathrm{D}_i,\mathrm{B}}$: the channel gain from $\mathrm{D}_{i,\mathrm{Tx}}$ to eNB
$G_{\mathrm{D}_i,\mathrm{D}_i}$: the channel gain from $\mathrm{D}_{i,\mathrm{Tx}}$ to $\mathrm{D}_{i,\mathrm{Rx}}$ of D2D pair i
$G_{\mathrm{D}_i,\mathrm{D}_j}$: the channel gain from $\mathrm{D}_{i,\mathrm{Tx}}$ of pair i to $\mathrm{D}_{j,\mathrm{Rx}}$ of another pair j
$G_{\mathrm{C}_m,\mathrm{D}_i}$: the channel gain from C_m to $\mathrm{D}_{i,\mathrm{Rx}}$
P_{C_m}: the transmission power of C_m
P_{D_i}: the transmission power of the transmitting DUE of D_i

2.1 Evaluation of SINR and Capacity

To assess the performance of the system, we have to calculate two kinds of SINR. The first kind is that from a CUE C_m to the eNB and is expressed as

$$SINR_{\mathrm{C}_m} = \frac{P_{\mathrm{C}_m} G_{\mathrm{C}_m,\mathrm{B}}}{\sum_{i \in \boldsymbol{U}_m} \left(P_{\mathrm{D}_i} G_{\mathrm{D}_i,\mathrm{B}} \right) + \sigma^2} \tag{1}$$

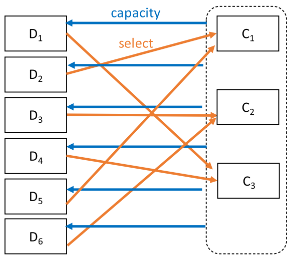
where σ^2 is the noise power and $\boldsymbol{U}_m$ is the set of indices of D2D pairs that reuse the RBs allocated to C_m. The second kind is that from the transmitting DUE of D_i to the receiving DUE of D_i and is expressed as

$$SINR_{\mathrm{D}_i} = \frac{P_{\mathrm{D}_i} G_{\mathrm{D}_i,\mathrm{D}_i}}{P_{\mathrm{C}_m} G_{\mathrm{C}_m,\mathrm{D}_i} + \sum_{j \in \boldsymbol{U}_m, j \neq i} \left(P_{\mathrm{D}_j} G_{\mathrm{D}_j,\mathrm{D}_i} \right) + \sigma^2} \tag{2}$$

In order to be more practical, in this paper, the link capacity is evaluated by the procedures in 3GPP specification 36.213 [18] and SINR values in [19]. Thus, link SINR can be mapped to CQI (Channel Quality Indicator) to get MCS (Modulation and Coding Scheme), and then TBS (Transport Block Size) table can be used to find the capacity if the number of RBs is known. On the other hand, by following the reverse procedures, the minimum SINR requirement can be obtained if the number of RBs and the data rate requirement are given.

3 Multi-Player Multi-Armed Bandit (MPMAB)

The model that we intend to apply on D2D resource allocation is MPMAB, which is an extension of Multi-Armed Bandit (MAB). With refer to the model as shown in Fig. 2(a) with 6 D2D pairs to reuse the RBs of 3 CUEs, there is a 6 × 3 matrix $\boldsymbol{Q}$ to record the average rewards corresponding to the reusing preference. Here the rewards are link capacities as defined in the previous section. Figure 2(b) is an example of the corresponding matrix $\boldsymbol{Q}$. Each element in $\boldsymbol{q}_i$ indicates the capacity/preference to the corresponding CUE. The matrix is updated after each trial until the learning process has converged. The converged criteria is similar to that presented in [1].



(a)

	C_1	C_2	C_3	
D_1	15	10	23	$\boldsymbol{q}_1$
D_2	11	25	14	$\boldsymbol{q}_2$
D_3	14	20	19	$\boldsymbol{q}_3$
D_4	17	20	23	$\boldsymbol{q}_4$
D_5	18	21	30	$\boldsymbol{q}_5$
D_6	19	18	16	$\boldsymbol{q}_6$

(b)

Fig. 2. MPMAB model of D2D reusing

4 Proposed Fair Resource Allocation Scheme

After the learning phase, the eNB will allocate RBs based on the final result of matrix $\boldsymbol{Q}$. Most of the papers concerning resource allocation employs greedy strategy so as to maximize the system capacity [5]. In this paper, we propose another criteria for allocating RBs: For the D2D pairs that are admitted to be served, the eNB tries best to fairly allocate RBs independent of their SINR values.

In order to reach this goal, we propose applying Allocate As Granted (AAG) concept presented in [20], where eNB tries to fairly allocate enough RBs to all CUEs that have been granted with declared data rate requirements. For a data rate $R_{\mathrm{D}_i}^{grant}$ in bits per second (bps), it may be converted to the number of bits $B_{\mathrm{D}_i}^{\mathrm{grant}}$ to be transmitted per Transmission Time Interval (TTI, 1 ms). Based on Exponentially Weighted Moving Average (EWMA) [21], we define the average allocated capacity $\overline{B}_{\mathrm{D}_i}(t)$ corresponding to D_i at the t^{th} TTI as

$$\overline{B}_{\mathrm{D}_i}(t) = (1 - \lambda_{\mathrm{D}_i}) \cdot \overline{B}_{\mathrm{D}_i}(t-1) + \lambda_{\mathrm{D}_i} \cdot B_{\mathrm{D}_i}(t), \tag{3}$$

where $B_{\mathrm{D}_i}(t)$ is the capacity (in bits per TTI) provided by the RBs allocated to D_i at the t^{th} TTI. The coefficient λ_{D_i} is a constant smoothing factor between 0 and 1. Our objective is that eNB tries best to fairly allocate enough RBs for D2D pairs to make $\overline{B}_{\mathrm{D}_i}(t) \geq B_{\mathrm{D}_i}^{\mathrm{grant}}$. Before allocating RBs for the served D2D pairs at a general TTI, the eNB calculates the priority of each D2D pair as

$$p_i(t) = \frac{B_{\mathrm{D}_i}^{\mathrm{grant}} - \overline{B}_{\mathrm{D}_i}(t-1)}{B_{\mathrm{D}_i}^{\mathrm{grant}}}. \tag{4}$$

The larger the metric is, the higher the priority for the D2D pair to be allocated with RBs.

The procedure of the allocating process at the t^{th} TTI is shown in Fig. 3 and is explained as follows.

i. Find the D2D pair with the highest priority metric.
ii. Checks whether this D2D pair can reuse the RBs with its most favorite CUE.
iii. The constraint for the reusing is that the CUE and the D2D pairs reusing the same RBs can respectively meet their SINR requirements. If fail, then reset the preference order to zero and try the next most favorite CUE.
iv. If yes, reset the corresponding priority metric to lowest value so that the eNB can try to allocate for the D2D pair with the next highest priority metric.

5 Performance Evaluation

In this section, we are going to compare the performance of GRA_SDF with that of our proposed scheme, which is indicated as proposed_with_AAG.

5.1 Simulation Environment

The simulation environment is an eNB serving some CUEs and D2D pairs, which are randomly distributed in the coverage of radius 500 m. According to [1], the minimum distance between DUE and eNB is set to 35 m. Both the learning process and the RB allocating process are executed at the eNB. Some assumptions are listed below.

i. The uplink resource of CUEs is allocated based on semi-persistent scheduling [18], rather than dynamic scheduling.

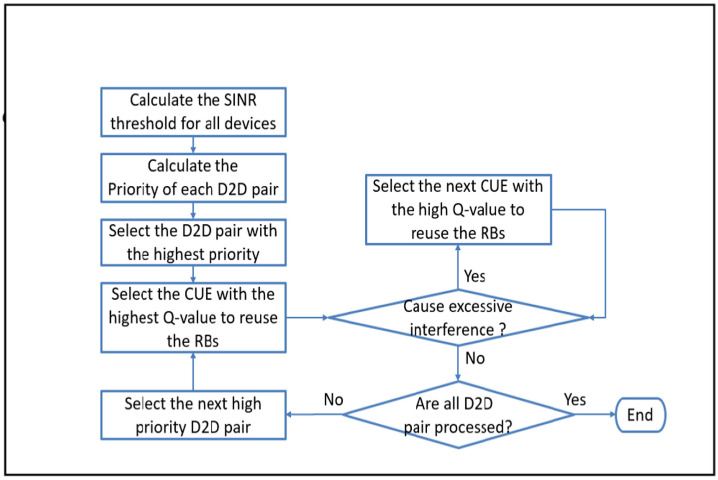


Fig. 3. Flow chart of AAG criterion

ii. Both the CUEs and D2D pairs can declare data rate requirements to the eNB.
iii. For simplicity, every CUE is allocated with the same number of RBs.
iv. Each CUE can share the allocated RBs with multiple D2D pairs.
v. Each D2D pair can reuse the RBs of only one CUE.
vi. If a D2D pair is to reuse the RBs allocated to a specific CUE, it reuses all of the RBs.
vii. Each CUE is allocated with 2 RBs. As a result, by checking the TBS table in [18], the maximum data rate of each CUE is 1.48 Mbps, which is also set as the CUEs' data rate requirement that must be satisfied.

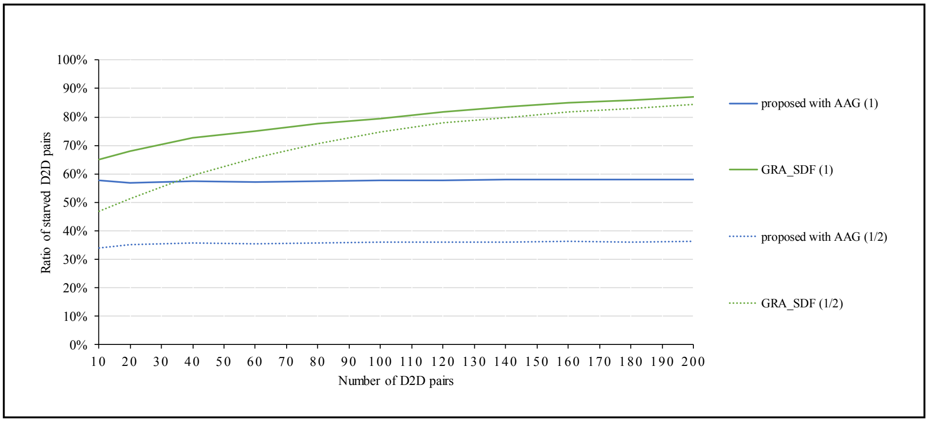


Fig. 4. Ratio of starved D2D pairs

5.2 Starved D2D Pairs

Before checking capacity, let us observe the ratio of starved D2D pairs as shown in Fig. 4. A D2D pair is regarded as starved if it gets no RBs during a simulation run.

In a whole, there are always D2D pairs that get no resource to reuse and become starved because they are located at bad positions. For a D2D pair to get higher data rate, higher SINR is required. Then, less D2D pairs can join the reusing, and more D2D pairs starve as a result. It is predictable that GRA_SDF results in higher ratios when there are more D2D pairs in the coverage. While, by introducing MPMAB to learn preference and AAG concept to fairly allocate RBs to all D2D pairs, the ratios are lower and almost constant.

It seems that it is not encouraged to have too many D2D pairs in the coverage for the GRA_SDF scheme. However, Fig. 5, which shows the numbers of non-starved D2D pairs corresponding to Fig. 4, reveals that the more the D2D pairs are in the coverage, the more are allocated to reuse RBs with CUEs. Because proposed_with_AAG tries to allocate for every D2D pair that meets SINR requirement, the numbers of allocated D2D pairs exhibit a linear increase with the numbers of D2D pairs.

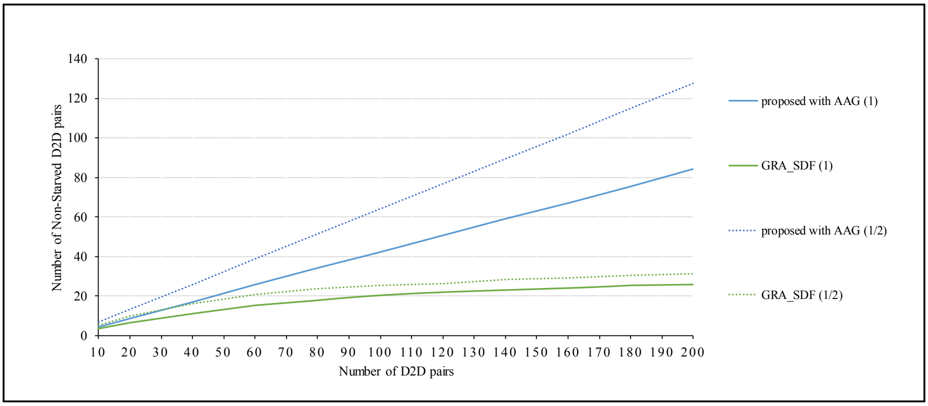


Fig. 5. Numbers of non-starved D2D pairs

5.3 Fairness of Satisfaction Degree

In the next, we are going to evaluate fairness of satisfactory based on Jain's fairness measure

$$\text{Fairness of satisfactory} = \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N \cdot \sum_{i=1}^{N} x_i^2}, \tag{5}$$

where N is the number of D2D pairs, x_i is the satisfactory degree corresponding to D2D pair i. Figure 6 shows the fairness of satisfactory. Note that all of the D2D pairs, including the starved D2D pairs, are counted in. As a result, it's not easy to reach the ideal value, one. This figure reveals that lower data rate requirements result in higher fairness. For each of the data rate requirements, fairness values of SDF_GRA decrease dramatically with the numbers of D2D pairs, while proposed_with_AAG keeps high and quite stable. From this point of view, the proposed_with_AAG also performs very well.

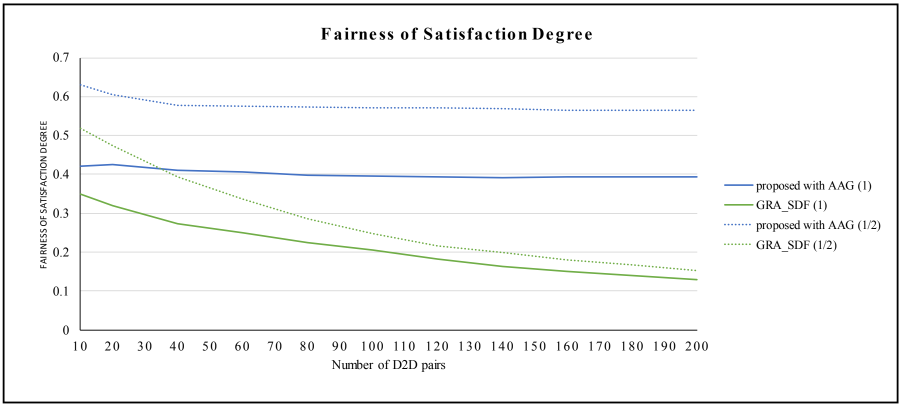


Fig. 6. Fairness of satisfaction degree

6 Conclusion and Future Work

This paper aims at proposing a D2D resource allocation method with the following characteristics: (i) One CUE can be allocated with multiple RBs depending on the declared data rate requirement. (ii) The RBs of a CUE can be reused by more than one D2D pair so as to effectively utilize the radio resource. We propose to apply AAG criterion to allocate RBs for D2D pairs fairly, rather than greedily. Simulation results reveal that this method can achieve fairness and alleviate starvation. In the future, we may combine power control to the proposed scheme.

References

1. 3GPP. TR 36.843 Study on LTE device to device proximity services; Radio aspects (2014)
2. Osseiran, A., et al.: The foundation of the mobile and wireless communications system for 2020 and beyond: Challenges, enablers and technology solutions. In: 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), pp. 1–5. IEEE (2013)
3. Ericsson, L.: More than 50 billion connected devices. White Paper **14**(1), 124 (2011)
4. Agiwal, M., Roy, A., Saxena, N.: Next generation 5G wireless networks: aA comprehensive survey. IEEE Commun. Surv. Tutor. **18**(3), 1617–1655 (2016)
5. Sun, H., Sheng, M., Wang, X., Zhang, Y., Liu, J., Wang, K.: Resource allocation for maximizing the device-to-device communications underlaying LTE-Advanced networks. In: 2013 IEEE/CIC International Conference on Communications in China-Workshops (CIC/ICCC), pp. 60–64. IEEE (2013)

6. Osseiran, A., et al.: Scenarios for 5G mobile and wireless communications: the vision of the METIS project. IEEE Commun. Mag. **52**(5), 26–35 (2014)
7. Mustafa, H.A.U., Imran, M.A., Shakir, M.Z., Imran, A., Tafazolli, R.: Separation framework: an enabler for cooperative and D2D communication for future 5G networks. IEEE Commun. Surv. Tutor. **18**(1), 419–445 (2015)
8. Asadi, A., Wang, Q., Mancuso, V.: A survey on device-to-device communication in cellular networks. IEEE Commun. Surv. Tutor. **16**(4), 1801–1819 (2014)
9. Doumi, T., et al.: LTE for public safety networks. IEEE Commun. Mag. **51**(2), 106–112 (2013)
10. Lin, X., Andrews, J.G., Ghosh, A., Ratasuk, R.: An overview of 3GPP device-to-device proximity services. IEEE Commun. Mag. **52**(4), 40–48 (2014)
11. Hakola, S., Chen, T., Lehtomäki, J., Koskela, T.: Device-to-device (D2D) communication in cellular network-performance analysis of optimum and practical communication mode selection. In: 2010 IEEE Wireless Communication and Networking Conference, pp. 1–6. IEEE (2010)
12. Peng, T., Lu, Q., Wang, H., Xu, S., Wang, W.: Interference avoidance mechanisms in the hybrid cellular and device-to-device systems. In: 2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 617–621. IEEE (2009)
13. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. The MIT Press, Cambridge (2011)
14. Luo, Y., Shi, Z., Zhou, X., Liu, Q., Yi, Q.: Dynamic resource allocations based on q-learning for D2D communication in cellular networks. In: 2014 11th International Computer Conference on Wavelet Actiev Media Technology and Information Processing (ICCWAMTIP), pp. 385–388. IEEE (2014)
15. Zhang, Y., Wang, C.-Y., Wei, H.-Y.: Incentive compatible overlay D2D system: a group-based framework without CQI feedback. IEEE Trans. Mob. Comput. **17**(9), 2069–2086 (2018)
16. Ren, H., Jiang, F., Wang, H.: Resource allocation based on clustering algorithm for hybrid device-to-device networks. In: 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6. IEEE (2017)
17. Neogi, A., Chaporkar, P., Karandikar, A.: Multi-Player Multi-Armed Bandit Based Resource Allocation for D2D Communications. arXiv preprint arXiv:1812.11837 (2018)
18. 3GPP. TS 36.213: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (2016)
19. Ghosh, A., Ratasuk, R.: Essentials of LTE and LTE-A. Cambridge University Press, Cambridge (2011)
20. Kuo, F.-C., Ting, K.-C., Wang, H.-C., Tseng, C.-C.: On demand resource allocation for LTE uplink transmission based on logical channel groups. Mobile Netw. Appl. **22**(5), 868–879 (2017). https://doi.org/10.1007/s11036-017-0860-7
21. Lucas, J.M., Saccucci, M.S.: Exponentially weighted moving average control schemes: properties and enhancements. Technometrics **32**(1), 1–12 (1990)
22. Jain, D.M.C.R., Hawe, W.R.: A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Digital Equipment Corporation (DEC) Research Report TR-301 (1984)