An Algorithmic Study to Maximize 5G Network Throughput Based on the Markov Decision Process

Zheng Xiao*
Sichuan Technology & Business College, China

*Corresponding author email: 26442481@qq.com

Abstract. In this article, the network throughput optimization problem is investigated based on the theory of Markov decision process, combining the device-to-device direct selection problem with the finite stage discount MDP model problem. First, models for the device-to-device communication selection using MDP are built; second, the optimal mode selection strategy is derived using a finite stage backward iterative algorithm; and finally, the given mode selection strategy is evaluated by conducting a large number of simulation experiments. The results show that the MDP-based mode selection method proposed in this article has better performance in maximizing throughput and can yield better mode selection strategies with the advantage of obtaining larger system throughput.

Keywords: 5G; Markov decision process; Device-to-Device communication; iterative algorithm.

1. Introduction

In 5G network, the system employs direct communication using Device-to-Device (D2D) technology, Having the superiority of improving communication quality, spectrum utilization and system throughput, is seen as the most promising 5G technology [1~3]. In a network of D2D users, each user can send and receive communication signals, and also has routing capabilities. By definition, D2D communication technology refers to a new communication method in which two peer users communicate directly, without the need to use base station forwarding [4]. Among those proposed technologies, D2D pattern selection problem is the key focus, such as path loss, distance, channel quality, and signal-to-interference and noise ratio-based mode selection [5~9]. If the D2D user pairs are closer to each other than to the base station, then it is obvious that direct communication is preferable. However, this is not always the case, and sometimes it’s more practical to choose a theoretically undesirable approach because of factors such as unstable network. Without consideration of signal to noise ratio and minimization of system interference, the traditional method lacks generality. So, what we need to look for is a universal approach that can be adapted to most situations. We need to model the real problem and determine a general pattern selection rule. In this article we relate the D2D pattern selection problem to the Markov decision process (MDP) [10]. Therefore, this article proposes a novel way to solve the model selection problem: an algorithmic study based on the Markov decision process.

2. Markov Decision Process Model

2.1. Channel Model

If a D2D user uses a reusing cellular communication method, interference is created within the network, so that each receiving user receives interference signals from other users within the same
band, and the base station is affected too. In this article, a flat Riley decay channel model is used, with the signal amplitude at the receiving end satisfying the Riley distribution. The Riley distribution is a smooth narrowband Gaussian process with a mean of 0 and a variance of $\sigma^2$ [11]. In this channel model, we assume that the receiver is subject to an additive white Gaussian noise (AWGN). This noise is the most basic noise interference mode in wireless channels, whose amplitude obeys a Gaussian distribution (mean of zero, variance of $N_0$). In this model, we can obtain the signal-to-noise ratio $SINR$ [12].

$$SINR = \frac{P_{\text{receiver}}}{I + N_0} = \frac{P_t d_{ij}^\alpha}{I + N_0}$$  \hspace{1cm} (1)

$SINR$ represents the ratio of the signal to the noise of the device. The larger it is, the better the quality of the signal is. The $P_{\text{receiver}}$ in formula (1) above is the power received at the receiving device. $I$ is the interference that the receiving equipment gets. $N_0$ is the noise to which the receiving equipment is exposed to. $P_t$ is the transmission power from the device. In general, cellular users and D2D users do not have the same transmission power, with the former higher. In practice, in order to maximize the network throughput, power distribution is configured according to certain rules, which is outside the scope of this research and will not be covered here. To make the computing easier, we stipulate that the user's transmission power is $P_{\text{DUE}}$ uniformly. $d_{ij}$ is the distance between the transmitting device $i$ and receiving device $j$. $\alpha$ is the path loss coefficient, which represents the loss that occurs when the signal transmits through space, and is determined by two factors: the transmission nature of the channel itself, and the radiation effect of the transmission power. According to the standard path loss transmission model, in general $\alpha > 2$. $H_{ij}$ refers to the channel coefficient.

To achieve the maximum overall channel capacity of the network, the overall channel capacity of the entire system $C_{\text{system}}$ needs to be computed. For the purpose of discussion, the network structure contains one cellular user and two pairs of D2D users, so $C_{\text{system}}$ is the sum of the channel capacities of the three users, as shown in Equation (2).

$$C_{\text{system}} = C_{\text{CUE}} + C_{\text{DUE1}} + C_{\text{DUE2}}$$  \hspace{1cm} (2)

$$C = BW \cdot \log_2 (1 + SINR)$$  \hspace{1cm} (3)

Formula (3) is a general formula for calculating channel capacity, where $BW$ refers to the system bandwidth and $SINR$ is the corresponding signal-to-noise ratio calculated in formula (1).

2.2. Markov Model

MDP can be analyzed in terms of five elements: decision moment/cycle, state, action set, transfer probability and reward. The set of time points at which each decision is made is $T$, while the corresponding set of system states is $S$, the set of actions is represented by $A$. At a certain point of time, assuming the existence of a state $i \in S$, then after selecting an action $a$ from the available set of actions $A(i)$ and executing it, we can immediately get a reward $r(i,a)$, and the state of the system at the next moment will be determined by the transfer probability distribution function $p(\cdot| i, a)$. Then at the next moment, we need to make another choice of action. Finally, by combining the actions at all points of time, a decision sequence, the set of choices made, is obtained. At the same time, each choice of action can result in both a timely reward and an impact in the future, thus creates an extra reward as shown in Figure 1.

![Figure 1. Diagram of the decision-making process.](image)
2.3. Algorithms of the Markov Decision Model

An iterative algorithm based on backward recursion of the expected reward for dynamic programming is used to compute the optimal value of the Markov decision model in the D2D model selection problem [13,14]. The algorithm \( f_t^* \) represents the optimal strategy at the moment \( t \), and \( \pi \) is the set of strategy sequences. At the moment \( N \), since the historic \( N \) stages of 0,1,...,N-1 in the previous cycles have been determined, there is no other decision option for the decision maker at the point, it is a fixed value. Set \( A_t^*(i_t) \) is generally defined as optimal set of actions. This backward recursive algorithm embodies the idea of an optimization principle. The optimal strategy has the nature that no matter from which initial state it starts, and no matter what initial action is taken, the strategy consisting of the remaining decision rules is the optimal strategy for the next decision moment [15].

Step 1 Let \( t = N \) and for all \( i_t \in S, u_N^*(i_N) = r_N(i_N) \).
Step 2 If \( t = 0 \), then \( \pi = (f_0^*, f_1^*, \ldots, f_{N-1}^*) \) is the optimal MDP strategy, and \( V_N^*(i) = u_0^*(i) \) is the optimal value function, then the algorithm stops. Otherwise, let \( t - 1 \Rightarrow t \), then proceed to step 3.
Step 3 For all \( i_t \in S \), compute

\[
  u_t(i_t) = \max \left\{ r_t(i_t, a) + \beta \sum_{j \in S} p_t(j|i_t) u_{t+1}^*(j) \right\},
\]

And let set

\[
  A_t^*(i_t) = \arg \max_{a \in A(i_t)} \left\{ r_t(i_t, a) + \sum_{j \in S} p_t(j|i_t, a) u_{t+1}^*(j) \right\}
\]

Take any \( f_t^*(i_t) \in A_t^*(i_t) \), then decision rule \( f_t^* \) is defined for moment of \( t=0 \)
Step 4 Return to step 2

Since the set of actions \( A \) is a finite set, the optimal solution of this Markov’s strategy must exist, and the final choice of actions at each decision moment can be obtained by the above algorithm [16], which together is the policy sequence of the mode choice we require, i.e., Policy. By searching the Policy matrix, we can know exactly what mode choice each of the two D2D pairs should make at a certain decision moment (time lag) when the system is in a certain state. At the same time, we can get an optimal expectation of reward. The flow of the algorithm is shown in Figure 2.

---

**Figure 2.** Flowchart of the backward recursive algorithm.
2.4. Simulation Experiments

In conjunction with the network structure model, the parameters for simulation experiments we set are shown in Table 1. The distance parameters in Table 1 are taken in the initial case and are subject to change. In order to simplify the model and facilitate the discussion, we take the channel coefficient $H_{ij}$ to be 1. It should also be noted that the units of power and noise are not uniform, and the conversion between them should be noted in calculation.

| Parameters                                             | Value       |
|--------------------------------------------------------|-------------|
| Radius of area ($R$)                                   | 500 m       |
| first pair of D2D source user location radius ($R_2$)  | 100 m       |
| second pair of D2D source user location radius ($R_3$) | 400 m       |
| Cellular user radius ($R_1$)                           | 300 m       |
| noise ($N_0$)                                          | -174 dbm/Hz |
| transmission power $P_{DUE}$                           | 20 dbm      |
| channel coefficient $H_{ij}$                           | 1           |
| Path loss factor $\alpha$                              | 4           |
| Band Width $BW$                                        | 10 MHz      |
| discount factor $\beta$                                | 0.9         |
| The distance between the first D2D pair ($r_1$)        | 10 m        |
| The distance between the second D2D pair ($r_2$)       | 10 m        |
| transfer probability $P_{gg}$                          | 0.8         |
| transfer probability $P_{bb}$                          | 0.2         |

3. Result

Using finite stage backward recursive iterative algorithms to combine D2D direct selection problem with finite stage discount MDP model problem together into a complete MDP problem, simulation models are built by MATLAB. Network throughput changes are observed when parameters are changed. Detailed experimental results are shown in Figures 3 to 5.

**Figure 3.** Trend plot of the impact of Order N on V.

In Figure 3, $R_1 = 300m$, $r_1 = 10m$, $r_2 = 10m$. For any stage, the trend of optimal value is consistent, and in general, the better the state of the channel, the higher the value. We are also more interested in the value of a good channel. So to simplify the image, two of the stages (1111 and 1110) were selected as representatives.

In Figure 4, $R_1 = 300m$, the distance between the two D2D pairs is changed simultaneously (the direction of motion is both at an angle of 0 degrees to the x-axis, and the direction of motion remains
constant). It is clear that, in terms of the trend, as the distance between D2D pairs increases, the value of the expected reward changes more dramatically and over a wider range when the distance between two D2D pairs changes simultaneously. It could be inferred that when there are multiple D2D pairs in the system and they are in motion at the same time, the channel capacity of the system may go extreme, which need to be investigated.

Figure 4. Trend plot of the impact of distance between D2D pairs on V.

Similarly, in Figure 5, let $r_1 = 300m$ and $r_2 = 10m$. Taking the order $N = 100$ and the time gap $= 500$ while changing the distance between the first D2D pair. It is seen from the simulation that increasing the distance will definitely lead to a decrease in the channel capacity due to the weakness of the received signal. At the same time, we can clearly see the advantage of MDP-based approach over channel-capacity-based approaches in maximizing network throughput, as the former can clearly obtain a larger system channel capacity. It was calculated that the MDP-based method is on average about 6Mbps higher than the channel capacity-based method, 7.1Mbps higher at most (at the distance of about 51m), which is a very significant.

Figure 5. Comparison of total system throughput generated from different mode selection methods.

4. Analysis

By taking network channel state into account, MDP is used to analyze the mode selection problem, and observe the impact of factors such as distance on throughput, with the aim of finding a mode selection method that can obtain the maximum channel capacity. The experimental results show that the proposed Markov decision process algorithm works better than channel capacity method.
5. Conclusion
The problem is solved by using an iterative algorithm from dynamic programming to obtain a time-dependent decision sequence with low computational complexity. When comparing simulations under multiple time slots, it is confirmed that the MDP-based model selection approach does have some advantages in optimizing network throughput. In the future, more rational options will be found to further improve the decision-making effectiveness of the algorithm.

Acknowledgments
Our thanks to all of the people who have contributed to this paper. My paper is supported by the project of the Ministry of Education's Fund for Educational Research Projects (2018A03007); Sichuan Provincial Department of Education's Fund for Educational Research and Reform Projects (JG2018-1168).

References
[1] Xia N, Chen H H, Yang C S. Radio Resource Management in Machine-to-Machine Communications - A Survey[J]. IEEE Communications Surveys & Tutorials, 2017:791-828.
[2] Fang L, Zhang R, Cheng X, et al. Cooperative Content Download-and-Share (CoCoDaS): Motivating D2D in Cellular Networks[J]. IEEE Communications Letters, 2017, PP(8):1831-1834.
[3] Li, Zhuoming, Chen, et al. Fuzzy mathematics and game theory based D2D multicast network construction[J]. Journal of Systems Engineering & Electronics, 2019:13-21.
[4] Huang Yonghui,Lian Zhaotong,Guo Xianping. Risk-sensitive finite-horizon piecewise deterministic Markov decision processes[J]. North-Holland,2020,48(1).96-103.
[5] Ochia O E, Fapojuwo A O. Energy and Spectral Efficiency Analysis for a Device-to-Device-Enabled Millimeter-Wave OFDMA Cellular Network[J]. IEEE Transactions on Communications, 2019, 67(11):8097-8111.
[6] Ortiz A, Asadi A, Engelhardt M, et al. CBMoS: Combinatorial Bandit Learning for Mode Selection and Resource Allocation in D2D Systems[J]. IEEE Journal on Selected Areas in Communications, 2019: 2225–2238.
[7] Omri A, Hasna M O. A Distance Based Mode Selection Scheme for D2D Enabled Networks with Mobility[J]. IEEE Transactions on Wireless Communications, 2018, 17(99):4326-4340.
[8] Li J, Lei G, Manogaran G, et al. D2D Communication Mode Selection and Resource Optimization Algorithm With Optimal Throughput in 5G Network[J]. IEEE Access, 2019:25263-25273.
[9] Sobhi-Givi S, Khazali A, Kalbkhani H, et al. Joint mode selection and resource allocation in D2D communication based underlaying cellular networks[J]. Telecommunication Systems: Modelling, Analysis, Design and Management, 2018, 67:1-16.
[10] Łukasz Balbus,Anna Jaskiewicz,Andrzej S. Nowak. Markov perfect equilibria in a dynamic decision model with quasi-hyperbolic discounting[J]. Springer US,2020,287(2).573-591.
[11] Wang Shuang,Yang Yang,Zhang Xinli. Equilibrium solution analysis of quantum chicken game model in phase damped channel [J]. Economic mathematics,2020,37(01):70-74.
[12] Xu Y. A Distributed Mode Selection Scheme for Device-to-Device Communications[J]. Wireless Personal Communications An International Journal, 2016, 90(2):639-655.
[13] XIE Xianzhong, YAN Ke, TIAN Yu, LI Chuanwen, CHEN Jiujui. Resource allocation algorithm with interference constraint for energy-efficient D2D communication based on game theory in cognitive networks [J]. Journal of changjing university of posts and telecommunications (natural science),2020,32(01):47-56.
[14] X. Dong, B. Guo, Y. Shen, X. Duan, Y. Shen and H. Zhang. "A Self-Controllable and Balanced Data Sharing Model," in IEEE Access, vol. 7, pp. 103275-103290, 2019.
[15] Bellman R. Dynamic Programming[J]. Science, 1966, 153(3731):34-37.
[16] Greggory, J, Schell,etc. Data-Driven Markov Decision Process Approximations for Personalized Hypertension Treatment Planning.[J]. Mdm Policy & Practice, 2018.