Research on Cache Strategy of Supporting D2D Communication Based on Recommendation Mechanism

Pingping Zhang\textsuperscript{1, a} Peisheng Pan\textsuperscript{1, b}

\textsuperscript{1}School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China
\textsuperscript{a}1170062339@qq.com; \textsuperscript{b}panps@njupt.edu.cn

Abstract. In D2D wireless cache network, a good cache strategy is very important, but the current cache strategy has the problems of low cache hit rate and high base station cost. Aiming at this problem, this paper proposes a smart base station cache strategy based on recommendation mechanism. Through the Q-learning algorithm to learn user mobility and file popularity information, a base station cache strategy with the goal of minimizing long-term base station cost is obtained. How to formulate the base station cache strategy and set the recommendation mechanism under different user movement scenarios in the cellular network. The simulation results show that the lower the user's mobility rate is, the smaller the optimal recommendation strength is. In the four user mobility scenarios in the simulation, the proposed cache strategy reduces the core network cost by up to 22.4% compared with the greedy cache, and it is reduced by up to 18.6% recommended strategy.

1. Introduction
Global mobile data traffic will grow a compound annual rate of 47% between 2016 and 2021, reaching 49 exabytes per month by 2021\cite{1}.The huge growth of mobile data traffic is mainly due to ultra high definition streaming video, cloud-based applications and the massive access of smart terminals\cite{2}.Device-to-Device (D2D) technology \cite{3}, as one of the key technologies of 5G communication, allows neighboring terminal to directly share data with each other under the control of a base station, forming a data sharing network and reusing the cellular network to reduce the load on the base station, improve spectrum utilization or improve system throughput \cite{4}.However, D2D communication has a geographical distance limitation, but users have a random and dynamic geographical location. Therefore, in a heterogeneous network of D2D communication, according to the real-time status of the user proposed a reasonable cache strategy to cope with the network situation of real-time updates \cite{5}.Most of the current studies, such as \cite{6}, are based on fixed network topology, but they are obviously not suitable for practical to use, especially when a large number of mobile phone users are in motion. The randomness of user behavior in mobile systems will lead to network performance degradation. This paper uses the distribution of contact time between users to model the user’s mobility. The results in \cite{7} show that website recommendations for individual products can increase the purchase rate of online products doubled, inspired by the great influence and success stories of recommendation mechanisms in life scenarios, \cite{8}\cite{9} proposed a variety of system recommendation mechanisms for user groups, allowing the system to affect file popularity, thereby affecting user request decisions for small base station coverage and reducing base station cache redundancy. Therefore, considering the impact of user mobility on the D2D communication cache network, cache strategy is
used to compare the effects of different recommendations and different user mobile scenarios. We take the long term cost of the base station as the optimization goal, consider the dynamically changing social attributes of users and propose a smart cache strategy for file popularity perception based on the recommendation mechanism. Using the Q-Learning algorithm to solve this dynamic optimization problem and simulation analysis is used to find the optimal recommendation strength and impact of recommendation strength on caching strategy.

2. System model and problem formulation

2.1. System model
Consider a scenario of a cellular heterogeneous network. A macro cell contains a macro base station, \( N_B \) small base stations, and \( n_u \) user cooperation. The set of small base stations is \( B \) and the number is \( b \). The size of the cache space \( C_b \), the set of users in the coverage area is \( u \), the user number is \( i \), and the number of users is \( n_u \). The file set is defined as \( \mathcal{F} \), the file number is \( m \), the number of included files is \( L \), and the size of each file is the same.

2.2. Mobility mode
We model the user’s mobility through the distribution of contact time between users and use the probability distribution of the contact time to represent the speed of the user's movement. Define the contact time between users as an exponential distribution that obeys the mean \( \tau \):

\[
T \sim \exp \left( \frac{1}{\tau} \right)
\]  

(1)

2.3. Recommendation mechanism model
The small base station notifies the users in the coverage area that it has cached the content and then the base station periodically broadcasts to the user key information about each cached file. Then user decides whether to send the request. Because users will change their requests for files due to system recommendation, users' requests for files are more concentrated on files that have been cached by small base stations, thereby increasing the communication success rate. Based on such a recommendation mechanism[10], the probability of a user's request for a file will linearly increase with the base station cache file ratio[11]. In this article, the increase in the request probability and the strength of the video site should be described by a linear function[12], the parameter \( k_r \) is the recommended strength, which indicates the perceived value of the system's broadcast publicity to users and the degree of user trust in advertising. After the recommendation, the user's probability distribution of file requests generates linear distortion according to the strength of the recommendation can be expressed by the following formula:

\[
\Delta_p = z_{bm}^b \cdot k_r
\]

(2)

where \( \Delta_p \) indicates that the distortion of the file request probability caused by the recommendation mechanism is directly proportional to the cache file proportion and recommendation strength of the small base station.

3. Problem formulation and solution

3.1. Problem formulation
The base station cost is defined as the sum of the base station’s incentives for user D2D communication distribution, the small base station service cost, and the core network service cost [13]. D2D communication is beneficial for small base stations and the core network to reduce service costs, so the setting of the recommendation mechanism is a trade-off between consumption and revenue. The expression of the total cost of the base station at time slot \( t \) is as follows:

\[
C_{b,t} = \omega_{\text{D2D}} \cdot C_{b_{\text{D2D}},t} + \omega_{\text{trans}} \cdot C_{b_{\text{trans}},t} + \omega_{\text{bh}} \cdot C_{b_{\text{bh},\text{trans}},t}
\]

(3)
where, $\omega_{D2D}$, $\omega_{trans}$, $\omega_{bh}$, represents D2D communication reward cost weights, small base station service cost weights, and core network service cost weights, respectively. The incentive cost in the total cost is proportional to the D2D communication volume between users, so we can get that in t time slot, the D2D communication incentive cost for users under the coverage of small base stations is requesting file $m$ in time slot $t$; $r_\omega$ represents the unit incentive distributed by the base station to users; $F^{m}_{i,j,t}$ unit represents the proportion $i$ of file $m$ that users $i$ and $j$ can successfully transmit through D2D communication:

$$F^{m}_{i,j,t} = \min\{x_{i,j}^m, \max\{0, V_{thr} - \sum_{j \in \mathcal{U}_s} x_{i,j}^m \min\{1, \frac{\tau_{i,j}}{X_{j,i}^m L_m / R}\}\}\}$$

where $x_{i,j}^m$ the proportion of files $i$ in the user $j$ cache of the time slot user and the first D2D communication, $L_m$-the file size, and $R$ the file transfer rate. The communication cost of the small base station is to remove the part of the file that has been obtained through D2D communication. The small base station cache can provide a representative for the users in the coverage area to transfer the required files

$$C_{trans} = \sum_{m=1}^{n} \sum_{i=1}^{m} r_{trans} \cdot P_{i,j}^m \cdot C_s \cdot L_m \cdot \min\{m_{b}, \max\{0, V_{thr} - \sum_{j \in \mathcal{U}_s} x_{i,j}^m \min\{1, \frac{\tau_{i,j}}{X_{j,i}^m L_m / R}\}\}\}$$

So by summing up all the small base stations in the macro cell can get the cost of all small base stations in the macro cell in the time slot

$$C_{b, all_{trans}, t} = \sum_{b \in B} C_{b_{trans}, t}$$

Similarly, we can deduce that in t time slot, the backhaul link communication cost of the core network is the cost incurred by the user to communicate with the core network in order to obtain the part of the file that cannot be obtained through D2D communication and small base station communication:

$$C_{backhaul} = \sum_{m=1}^{n} \sum_{i=1}^{m} r_{backhaul} \cdot P_{i,j}^m \cdot C_{backhaul} \cdot L_m \cdot \max\{0, \max\{0, \min\{1, \frac{\tau_{i,j}}{X_{j,i}^m L_m / R}\}\}\}$$

The long-term cost of the base station obtained by summing the discount factors in the time domain is expressed by the following formula, where the discount factor $\beta$ is set to reflect the influence of the current action on the longer-term cost.

$$C_{BS, t} = \sum_{r=1}^{\infty} \beta^{r-t} C_{BS, t}$$

Because $P_{i,j} = (P_{i,j}^1, P_{i,j}^2, ..., P_{i,j}^m)$ is a function of vectors $C_{BS}$, it is not practical to directly solve the optimization problem by minimizing this parameter to maximize the system benefits. The optimization problem can be solved by minimizing its expectations \cite{8}. $Z = (z_{i,j}^1, z_{i,j}^2, ..., z_{i,j}^m)$ is set as an array of small base station cache policies, which forms a mapping of each state to each action, expressed as $\pi: P_{i,j}^m \rightarrow Z_{\pi,i,j} = (z_{i,j}^1, z_{i,j}^2, ..., z_{i,j}^m)$, so small base station cache strategy optimization problem is expressed by the following formula:

$$\min_{Z} E_{Z} \{C_{BS, t}\}$$

s.t. $z_{i,j}^m \in [0, 1], \sum_{i=1}^{L} z_{i,j}^m = C_{b}, i \in \{1, 2, ..., L\}$
The constraint is that the small base station has a non-negative nature of the cache capacity limitation and the small base station cache ratio. The long-term cost expectation of the base station can be split into the following forms:

\[ E^*[C_{BS}] = E^*[P_i, P_{ij} | P_{i,t} = P_z(m)] = E^*[P_{i,t} | P_{i,z} = P_z(m)] + \beta E^*[P_{i,t+1} | P_{i,z} = P_z(m)] + \beta \sum_{P_{i,z} \in \Pi} P \{ P_{i+1,t} = P_z(m) | P_{i,z} = P_z(m), z_{b,j} = \pi(P_z(m)) \}
\]

(11)

So the objective function that needs to be optimized is a non-linear convex optimization problem with equality constraints. In order to solve this dynamic programming optimization problem, we use reinforcement learning algorithms to solve this problem. In order to reduce complexity, we turned to the upper bound of maximizing system revenue expectations [8], so we can solve the upper bound of the system's reward expectations express as

\[ E^*[C_{BS}] = E^*[P_{i,t} | P_{i,z} = P_z(m)] + \beta \sum_{P_{i,z} \in \Pi} P \{ P_{i+1,t} = P_z(m) | P_{i,z} = P_z(m), z_{b,j} = \pi(P_z(m)) \} \]

(12)

Finally, the optimization problem is transformed into the following form:

\[ \min \ E^*[C_{BS}] \]

\[ s.t.: z_{b,j} \in [0,1], \sum_{i=1}^{L} z_{b,j} = C_k, i \in \{1,2,...,L\} \]

(13)

3.2. Solution

Combining formula (11), it can be found that the process of caching policy decision fits well with the decision process of action, state, feedback and environment in reinforcement learning[14], so we Q-Learning algorithm be used to solve this optimization problem. Since the standard Q-Learning algorithm is to solve the problem of maximizing system revenue, and our goal is to minimize the long-term base station cost, we set the system revenue Q to the inverse of the target long-term base station cost:

\[ Q(P_i(m), z^*_m) = 1/(C_{BS} + 1) \]

(14)

The process of reinforcement learning is the process in which actions and states affect each other, and the optimal action is obtained through the feedback value function acting on the network control center.

1. Action: "Action" refers to the small base station cache strategy, which is represented by an array \(Z = [z_{b,1}, z_{b,2}, ..., z_{b,k}]\).

2. Status: "Status" refers to the distribution of file popularity in the system. Cache action will affect the status of the system and is represented by an array \(P = [p_i(1), p_i(2), ..., p_i(m)]\).

3. Feedback value function: long-term base station cost.

\[ z' = \max_{P} Q(P_i(m'), z') \]

(15)

The goal of the Q-learning algorithm is to obtain a stable iterative Q table to guide the optimal base station cache strategy for each time slot in the future. For a given Q value and status, the action according to the formula (15) be updated to select the action that maximizes the long-term system reward" expect. This behavior belongs to the system's "development" process. But the known Q value may not be the optimal value. "exploration" will try different actions and compare whether they are better than the
actions that have been tried. Bring better system benefits, "development" will try to prove the most effective behavior from past experience, we use ε greedy method [15], where the development and exploration can be compromised by a greedy factor ε as follow:

$$z = \begin{cases} \arg \max Q(p_i(m), z), & \varepsilon \\ \text{or randomly,} 1 - \varepsilon \end{cases}$$ (16)

Set the maximum number of iterations and start the iteration. The single iteration process is: a random number of 0-1 is generated. If ε > g, the file is cached according to formula (15), that is, when the action is updated, select the one that can maximize the Q value. The action generates an array; otherwise, randomly select files from the cache file set for caching according to formula (16), and at the end of the iteration, update the Q table according to formula (17).

$$Q(p_i(m), z) \leftarrow (1 - \varepsilon)Q(p_i(m), z) + \varepsilon[R(p_i(m), z) + \beta \max Q(p_i(m)', z')]$$ (17)

The setting $p_i(m)' = p_i(m)$ enters the next iteration process. By accomplishing the Q-learning algorithm, the cache strategy can obtain the expected minimize long-term base station cost.

Algorithm Q-Learning based Caching Strategy

Step1: At the beginning of current timeslot, calculate file popularity distribution $p_i(m)$
Step2: Set the maximum number of iterations $\nu$
Step3: Produce a number $l = (0,1)$ randomly, if $l > \varepsilon_g$, update $Z$ according to formula (16), otherwise, randomly choose $Z$
Step4: At the end of current timeslot, calculate cache policy $Z_{m,b,t}$ and the states in the next interval $P^{n,m}$
Step5: Update the Q-value n line with formula (3)
Step6: $p_i(m)' = p_i(m)$ enters the next iteration process
Step7: Judge the deviation between the current Q table and the previous Q table, when it is less than the iteration accuracy, terminate the iteration process

4. Results

Consider a macro cell contain a macro base station, 8 micro base stations, and 40 user collaboration scenarios. The file set contains 8 files, each of which has the same size, and the number of files in the cache file set is 4. According to the literature [11], set the initial parameter of the Zipf in the recommendation mechanism to 0.4: the learning rate $a$ in the Q-learning algorithm is 0.02, Q-learning algorithm related parameters reference [8], user mobility and communication model related parameters reference [13]. Considering the impact of the proposed caching strategy on the cost of long-term base stations under different user mobility scenarios, three user mobility scenarios are taken: the average user communication time is 5s, 50s, 100s, 200s.

Figure 1 show that no matter what kind of user movement scenario, when the recommended strength is greater than 0, the long-term base station cost is always less than the recommended strength is 0. In general, with the stubbornness of recommendation, the long-term base station cost will first decrease and then increase. Therefore, in different mobile scenarios, there is an optimal value of the recommended strength to minimize the long-term base station cost.

Observing the overall trend can be seen that the higher the user's movement rate, the bigger the cost of the core network is, because the D2D communication proportion in the scenario where the user's movement rate is high decreases, and the larger burden falls on the small base station and the core network. Through the simulation in Figure 2, it can be concluded that compared with the recommended cache strategy in article[9], in the four user movement scenarios, the proposed cache strategy can reduce the core network cost by 18.6% at the highest; compared with the greedy cache strategy, the maximum reduce the core network cost by 22.4%.
6

5. Conclusion
In this paper, we propose a smart caching strategy based on a recommendation mechanism and perform simulation analysis. The goal is to minimize the long-term base station cost. The optimization problem is described as a dynamic programming problem and then using the Q-learning algorithm to sense the file popularity distribution, and the problem is transformed into a base station cache strategy optimization problem that dynamically senses changes in users and file information. By analyzing the results of the iteration and comparing the results with the reference cache strategy in different recommendation strengths and different user mobility scenarios, in the simulation scenario, the recommendation mechanism introduction be compared with the greedy cache strategy, the proposed cache strategy can reduce the core network cost by up to 22.4% and the recommended cache strategy [9] can reduce the core network cost by up to 18.6%. It can be found that the cache strategy proposed in this paper is a solution that can better meet the system's low energy consumption requirements. In the future, we will focus on how to apply the proposed method to real-time networks.

References
[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update,2016-2021
[2] QIAN Zhi-hong, WANG Xue. Reviews of D2D technology for 5G communication networks [J]. Journal on Communications,2016,37(7):1-14.
[3] JIANG Jing-jie, ZHANG Sheng-kai, LI Bo, et al. Maximized Cellular Traffic Offloading via aprl N Device-to-Device Content Sharing[J]. IEEE Journal on Selected Areas in Communications,2016,34(01):82-91
[4] H. Mustafa, M. Imran, M. Shakir, et al. Separation Framework: An Enabler for Cooperative and D2D Communication for Future 5G Networks [J]. IEEE C Communications Surveys Tutorials, 2016,18(1):419-445.
[5] CAO Yuan. Global smart phone traffic expected to double by 2021. [J]. Computer and Network,2017,43(12):16
[6] K.Shanmugam, N.GoIrezaei, A. Dimakis, A. Molisch, and G.Caire, “Femtocaching: Wire less content delivery through distributed cachinghelpers.” IEEE Trans.Information Theory, vol.59, no. 12, PP.8402-8413, Dec.2013.
[7] Senercal, S., Nantel, J. The influence of online product recommendations on consumers online choices [J]. J. Retailing, 2004, 80(2): 159 - 169.
[8] McNee, S. M., Riedl, J., Konstan, J. A. Being accurate in not enough: how accuracy metrics have hurt recommenfer systems[A]. //2006 ACM CHI Conference on Human Factors in Computing systems[C], New York: ACM Press, 2006:1-5.
[9] Guo, K., Yang, C., Liu, T. Caching in Base Station with Recommendation via Q-learning [A]. //2017 IEEE Wireless Communications and Networking Conference (WCNC) [C], Washington D.C.: IEEE Computer Society, 2017:1-6.
[10] Chatzieleftheriou, L.E., Karalipoulos, M., Koutsopoulos, I. Caching-aware recommendations: Nudging user preferences towards better caching performance[A]. in IEEE Conference on Computer Communications[C], Washington Computer Society, 2013:1017-1021.
[11] YU Fu-xi. The Impact of Online Video Advertising on Consumers' Intention to Buy[J]. Modern communication,2016, 22:45-47
[12] TAN Xia-ning. Research on Key Technologies in Wireless Cache Network[D]. University of Electronic Science and Technology of China,2017:1-84
[13] I.Paralardo, G.Quer, BD. Rao, and M. Zorzi,”Caching strategies in heterogeneous networks D2D, small BS and macro BS communications”, in IEEE ICC, 2016.
[14] N. Golovin and E. Rahm, Reinforcement learning architecture for web recommendations, in null, 2004, p. 398.
[15] Chen, Z., Liu, Zhou, B.et al. Caching incentive design in wireless D2D networks: A Stachelberg game approach[A]. //2016 IEEE International Conference on Communications (ICC)[C], Washington D.c.: IEEE Computer Societf,2016:1-6.