Computers, Materials & Continua
DOI: 10.32604/cmc.2023.028597
Article

Analysis on D2D Heterogeneous Networks with State-Dependent Priority Traffic

Guangjun Liang[1,2], Jianfang Xin[3,*], Linging Xia[1], Xueli Ni[1,4] and Yi Cao[5]

[1]Department of Computer Information and Cyber Security, Jiangsu Police Institute, Nanjing, China
[2]National and Local Joint Engineering Laboratory of Radio Frequency Integration and Microassembly Technology, Nanjing University of Posts and Telecommunications, Nanjing, China
[3]School of Intelligent Engineering, Nanjing Institute of Railway Technology, Nanjing, China
[4]Jiangsu Province Electronic Data Forensics and Analysis Engineering Research Center, Nanjing, China
[5]Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, N9B 3P4, Canada
*Corresponding Author: Jianfang Xin. Email: xinjfang@163.com
Received: 14 February 2022; Accepted: 19 April 2022

Abstract: In this work, we consider the performance analysis of state dependent priority traffic and scheduling in device to device (D2D) heterogeneous networks. There are two priority transmission types of data in wireless communication, such as video or telephone, which always meet the requirements of high priority (HP) data transmission first. If there is a large amount of low priority (LP) data, there will be a large amount of LP data that cannot be sent. This situation will cause excessive delay of LP data and packet dropping probability. In order to solve this problem, the data transmission process of high priority queue and low priority queue is studied. Considering the priority jump strategy to the priority queuing model, the queuing process with two priority data is modeled as a two-dimensional Markov chain. A state dependent priority jump queuing strategy is proposed, which can improve the discarding performance of low priority data. The quasi birth and death process method (QBD) and fixed point iteration method are used to solve the causality, and the steady-state probability distribution is further obtained.Then, performance parameters such as average queue length, average throughput, average delay and packet dropping probability for both high and low priority data can be expressed. The simulation results verify the correctness of the theoretical derivation. Meanwhile, the proposed priority jump queuing strategy can significantly improve the drop performance of low-priority data.

Keywords: Stochastic geometry; queuing theory; D2D heterogeneous networks; quasi-birth and death process

1 Introduction

Device-to-device (D2D) communication is a modern communication method that allows direct communication between nearby users, which is considered to be one of the key technologies of 5G communication system [1,2]. It provides new opportunities for proximity-based business services, including the development of social networks [3,4]. Due to D2D communication may bring many advantages, such as improved spectrum efficiency, expanded cellular coverage, improved energy efficiency, and reduced backhaul requirements [3]. Unlike self-organizing networks, D2D communication is usually established under the control of a base station (BS) [3].

In [5], Zhong et al. employed queuing theory and Poisson point process (PPP) to study the necessary and sufficient conditions for queuing stability in static Poisson networks. In [6], Zhong et al. considered the randomly generated data flow and further analyzed the packet delay in the cellular heterogeneous network. They evaluated the delay performance of three scheduling schemes: Random Selection (RS), first-in first-out (FIFO) scheduling and round-robin (RR) protocol. However, because the signal-to-interference ratio (SIR) analysis and queue status are not decoupled, only the boundary of the network delay distribution under a fixed base station can be obtained. In [7], Yang et al. assumes that the number of users in each BS is fixed. On this basis, the delay interruption of PPP distributed small cell network using RS or RR is derived.

Existing works have applied the priority jump queue model to wireless packet networks or cognitive radio networks [8–10]. According to the inherent characteristics of multiple traffic types, the system parameters are derived under the premise of satisfying Quality of Service (QoS) to optimize the network architecture. In [8], Balapuwaduge et al. established two queues for secondary users in a multi-channel wireless cognitive network with heterogeneous traffic. The two queues analyzed and stored real-time and elastic data. The work allocated priority service to high-level real-time data through channel access strategies, while establishes a queuing scheme based on delay threshold for elastic data. In addition, a continuous-time Markov chain model is established to evaluate the performance of the proposed channel allocation strategy. The correctness and accuracy of the derived theoretical model is verified through extensive simulations. In [9], Walraevens et al. proposed a spatial merging method in the packet switching network to study the queue model with limited buffer and jump priority. When a call with a low priority arrives, it is assumed that such a call is transferred to the end of the high priority call queue. The transition probability depends on the status of the heterogeneous call queue. The algorithm calculates the service quality index of this type of queuing model and analyzes the results of numerical experiments. In addition, the dynamic priority jump model is also applied to wireless body area networks. In [10], Zhao et al. considered combining the priority transmission of medical data packets in the wireless body area network with the dynamic priority queuing model. Based on the analysis of a delay-dependent dynamic priority transmission scheduling scheme, a queuing principle with buffer overflow is studied, which is approximated as the jump strategy in the priority transmission scheduling criterion, and the delay is analyzed. The tail characteristics of the distribution. So far, few papers apply the priority shifting model to the D2D cellular heterogeneous network.

So far, few papers have applied the priority queue model to the D2D cellular heterogeneous network. Reference [11] models two priority data transmission processes, assuming that two priority data have independent caches. There is no mutual constraint between the two caches. when the arrival rate of one type is low, it will cause a great waste of the cache space of the queue. Moreover, HP data transmission is always met first. If there is a large amount of LP data, there will be a large amount of LP data that cannot be sent. This situation will cause excessive delay and packet loss of LP data. In order to solve this problem, in our paper, we assume that the high and low priority data are stored

in the common cache space. We allocate the public cache to the high and low priority data storage space in proportion. This strategy can maximize the utilization of the public buffer area and ensure the packet loss rate of low-priority data. In order to further optimize the data transmission process of high and low priority queues, we consider adding a priority jump strategy to the priority queuing model. Priority jump strategy allow low priority data packets to be sent as high priority data with a certain jump probability. The jump probability is determined by the queue leader status of high and low priority queues, so it is called state dependent priority jump strategy.

This paper focuses on the multi-class service transmission model in the D2D underlying cellular network. Combined with queuing theory and Stochastic geometry, a spatiotemporal model is proposed to analyze the performance of cellular users in heterogeneous networks. First, we consider the transmission mode of potential D2D users in the cellular network which adopts a distance-based D2D mode selection strategy. At the same time, considering that the interference is caused by users who use non-empty buffers, we use the thinning Poisson point process to model the spatial distribution of cellular users and obtain the probability of successful transmission. Second, we consider the priority transmission status of multiple business models, which can make full use of the buffer space. The priority jump strategy with a common buffer area in the queuing theory is adopted to provide more transmission opportunities for low-priority data, thereby alleviating the starvation state of low-priority data under high load. A two-dimensional Geo/G/1 Markov chain is established to describe the queue model with priority switching strategy for each cellular user, and the quasi-birth and death (QBD) method is used to evaluate the queuing behavior. When calculating the steady-state probability distribution, the iterative solution is used to calculate the steady-state probability distribution and obtain the expression of performance indicators.

The rest of this paper is organized as follows. The Section 2 describes the system model and performance indicators. The queuing model analysis framework with priority jumping is established in the Section 3. Queue stability analysis and performance parameters are introduced in Section 4. Section 5 provides numerical and simulation results, followed by the conclusions of Section 6.

2 System Model

As shown in Fig. 1, we consider a hybrid network, including cellular users and D2D users, and focus on the user's uplink communication. Assuming that the location distribution of base stations follows independent PPPs Φ_B with the density is λ_b.They are randomly distributed in the regular hexagonal grid-like area and the area of the hexagonal cell is $\frac{1}{\lambda_b}$, where λ_b is the average number of BSs per unit area. In this paper, A is used to denote the coverage area of the hexagonal macro cell. We approximate A by a spherical disk with the same area as the hexagonal cell with a radius of R, $R = \sqrt{\frac{1}{\pi\lambda_b}}$. In addition, we assume that the cellular users are evenly distributed in the coverage area A, and the location of the cellular user constitutes a PPP Φ_C of λ_c. Cellular users communicate with the base station closest to it. For typical uplink transmission, the interference of other cellular users comes from outside area A. We make reasonable assumptions of $\lambda_c >> \lambda_b$.

This paper considers that the uplink spectrum resources in the cell are orthogonally allocated to different cellular users. Potential D2D users reuse the channel in an underlay manner. It is assumed that the cellular network is in a fully loaded state, that is, all uplink sub-channels in the cell are allocated to different cellular users. We adopt a mode selection strategy for potential D2D users. When the distance between the D2D transmitter and the target receiver is less than the preset threshold, the potential D2D user will work in D2D mode, that is, sharing channel resources with cellular users

for direct communication. At this time, D2D communication will cause interference to the cellular link of the shared channel. When the distance between the D2D transmitter and the receiver exceeds the threshold, the D2D user enters the sleep mode [12,13]. In addition, it is assumed that each D2D transmitter is equipped with a limited buffer for storing sporadic arriving data packets, in which only non-empty queues will cause interference to other links in the shared channel. Assuming that potential D2D users obey PPPs Φ_D with density of λ_d, only some users can cause interference to the cellular link. The position distribution of these users constitutes a thinned PPP Φ_D', and the density λ_d' can be expressed as

$$\lambda_d' = P\left(r_d < \mu\right)\eta\lambda_d \tag{1}$$

where r_d represents the distance from the D2D transmitter to its target receiver and μ is the preset distance threshold. η represents the probability of non-empty queue status.

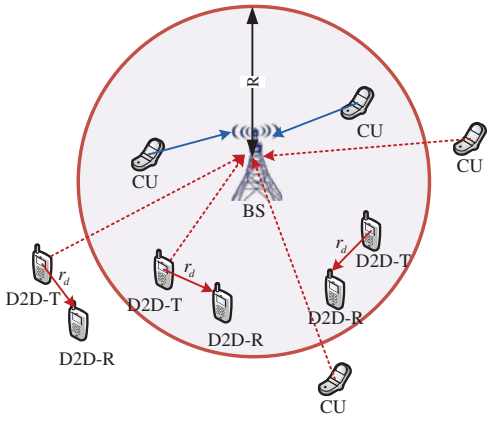

Figure 1: D2D underlying cellular heterogeneous network

We assume that each potential D2D receiver is randomly distributed around its potential transmitter, and the distance from the transmitter to the receiver r_d obeys the Rayleigh distribution. Its probability density function (PDF) is as follow:

$$f_{r_d}(x) = 2\pi\varsigma x e^{-\pi\varsigma x^2}, x \geq 0 \tag{2}$$

We consider the path loss plus block fading channel model [14], The received power of the cellular user is $p_t g d^{-\alpha}$, where p_t is the transmit power of the cellular user, g represents the small-scale fading of the communication link, which follows the exponential distribution of the normalized mean. d is the distance from the transmitter to the receiver and α is the path loss factor.

Additionally, as we consider the uplink transmission, power control is necessary. Uplink power control is used to adjust the transmit power to keep the base station receiving the signal power from the cellular user at a certain value, we employ the full channel inversion for uplink power control [15]. Then the transmit power of the cellular user is $\rho_{BS} d_{k,BS}^{\alpha_C}$, where ρ_{BS} is the minimum required power at the BS. $d_{k,BS}$ is the distance from cellular user k to its serving base station. α_C represents the path loss factor of the cellular link.

According to the above assumptions, cellular users k is subjected to intra-layer interference from cellular users in other cells I_{agg}^C, as well as inter-layer interference I_{agg}^D from D2D users sharing channels

inside and outside the cell due to non-orthogonal spectrum access strategies. Then the interference received by the serving base station of cellular user k can be expressed as

$$I_{agg}^{C} + I_{agg}^{D} = \sum_{i \in \Phi_C \setminus \{k\}} \rho_{BS} d_{k,BS}^{\alpha_C} d_{i,BS}^{-\alpha_C} \; g_{i,BS} + \sum_{j \in \Phi'_D} P_d g_{j,BS} d_{j,BS}^{-\alpha_D} \tag{3}$$

where $g_{i,BS}$ and $g_{j,BS}$ respectively represent the fading power gain of the interference link from cellular user i and the D2D user j to the BS. These two parameters are assumed to be independent and identically distributed. Considering that the cellular link and the D2D link may experience different propagation channels, α_C and α_D are defined as the path loss factors of cellular and D2D link. $d_{i,BS}$ and $d_{j,BS}$ respectively represent the distance from CU i and D2D users j to the BS. P_d is the D2D transmit power.

Considering the interference limited system, we can get the expression of the signal to interference and noise ratio of cellular user i to the BS as:

$$SINR^k = \frac{g_{k,BS} \rho_{BS}}{I_{agg}^{C} + I_{agg}^{D} + \delta^2} \tag{4}$$

where $g_{k,BS}$ is the fading power is gain between CU k and BS, and δ^2 is the noise power.

3 Queuing Model and Priority Jump Strategy

In this section, we build a dynamic priority queuing model with a common buffer area, and discuss state-dependent priority jump strategies.

3.1 Queuing Model

We use the two-dimensional Geo/G/1 Markov chain to establish a priority queue model for cellular users, where Geo represents the arrival process that obeys the geometric distribution, and G represents the service time, which is a non-negative random variable. Taking into account the actual interference received by cellular users in the D2D underlying cellular network, the service process is determined by the SINR value of each time slot. For the convenience of analysis, this paper takes two types of priority data as examples. For the case of multiple types of priority, the derivation process is similar. By setting the jumping probability $\gamma_i(j)$, the relationship between high and low priority queue is expressed.

(1) Data packet arrival process

It is assumed that the data packets arrival of the cellular user obeys the Bernoulli distribution with the parameter α, in which the high priority (HP) service data is generated with probability p_{HP}, and the low priority (LP) service data is generated by p_{LP}, $p_{LP} = 1 - p_{HP}$. Then HP data packet arrival probability is ap_{HP}, and the LP data packet arrival probability is $a(1 - p_{HP})$.

(2) Service process

Assuming that the cellular user is sending at the maximum rate, the sending rate is $R(t) = B\log_2\left(1 + SINR^k(t)\right)$. We set the time slot length is T_0 and each data packet has a fixed length L_0. Then the lowest average sending rate to send a data packet is $R_0 = \frac{L_0}{T_0}$. The expression of probability that the sending rate exceeds the lowest average sending rate is the probability of successful transmission of a data packet, then

$$b = P(R(t) > R_0) = P\left(B\log_2\left(1 + SINR^k(t)\right) > \frac{L_0}{T_0}\right) = P\left(SINR^k(t) > 2^{\frac{L_0}{BT_0}} - 1\right) \tag{5}$$

The complementary cumulative distribution function (CCDF) of the cellular link SINR can be expressed as

$$P\left[SINR^k > 2^{\frac{L_0}{BT_0}} - 1\right] = E_r\left[P\left[SINR^k > 2^{\frac{L_0}{BT_0}} - 1 \,|r\right]\right]$$

$$= \int_0^R P\left[SINR^k > 2^{\frac{L_0}{BT_0}} - 1 \,|r\right] f_r(r)\, dr$$

$$= \int_0^R \frac{2}{R^2} e^{-\frac{2^{\frac{L_0}{BT_0}} - 1}{\rho_{BS}} \cdot \delta^2} \cdot \mathcal{L}_{I_{agg}^D}\left(\frac{2^{\frac{L_0}{BT_0}} - 1}{\rho_{BS}}\right) \cdot \mathcal{L}_{I_{agg}^C}\left(\frac{2^{\frac{L_0}{BT_0}} - 1}{\rho_{BS}}\right) rdr \tag{6}$$

where $\mathcal{L}_{I_{agg}^D}(s)$ and $\mathcal{L}_{I_{agg}^C}(s)$ are the Laplace transform of I_{agg}^C and I_{agg}^D, respectively. Their expressions are as follow:

$$\mathcal{L}_{I_{agg}^D}(s) = \exp\left(-2\pi\lambda_d \int_0^\infty \left(1 - E_{g_d}\left[\frac{1}{1 + sP_d v^{-\alpha_D}}\right]\right) vdv\right) \tag{7}$$

Formula (7) is established based on the independent and identical distribution of g_d and the probability generating function (PGFL) from PPP.

$$\mathcal{L}_{I_{agg}^C}(s) = \exp\left(-2\pi\lambda_c \int_R^\infty (1 - E_{g_c}\left[\frac{1}{1 + s\rho_{BS} d_{k,BS}^{\alpha_C} d_{i,BS}^{-\alpha_C} g_c}\right]) vdv\right) \tag{8}$$

The derivation of $\mathcal{L}_{I_{agg}^C}(s)$ is mostly the same as $\mathcal{L}_{I_{agg}^D}(s)$, except that the value range of the integral variable v refers to the distance between a cellular user in other cells and its serving base station. By substituting Eqs. (6)–(8) into Eq. (5), the probability that the CU transmission rate is greater than is obtained, that is, the closed expression of the probability of successful transmission.

3.2 Priority Jump Queuing Strategy

In this discrete-time queuing system, we assume that each cellular user can only send one data packet in one time slot. In the transmission process of multiple data types, services such as video or telephony, which require high delay, need to be transmitted first and set as HP data. Data services like short messages have lower latency requirements and are set as LP services. The priority queuing model used in this paper always meets the transmission of HP data first. If there is a large amount of LP data, there will be a large amount of LP data that cannot be sent. This situation will lead to excessive delay and packet loss of LP data. We consider adding a priority jump strategy to the priority queuing model to allow low priority data packets to be sent as high priority data with a certain jump probability. The jump probability is determined by the number of data packets in the high and low priority queues. it is called state dependent priority jump strategy.

As shown in Fig. 2, we consider a model with a common buffer in which two types of data traffic are stored. The buffer size is L. In this model, the state-dependent priority jump strategy is adopted and there are the following definitions.

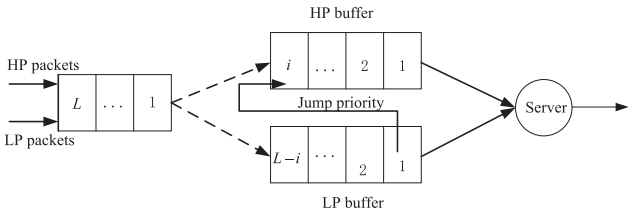

Figure 2: Structure of queuing system with priority jump

First, as long as there is high priority data in the queue, high priority data will be sent first, and LP data will be sent when the HP buffer is empty. Secondly, if there is only high priority data in the queue and the queue is not full, the packets will enter the queue with probability 1 and to be transmitted. If the queue is full, the arriving HP data will be discarded with probability 1. Third, when the number of data packets in the buffer of the HP queue is i, $i < L$ and the number of data packets in the LP queue is $j, j < L - i$,The newly arrived low-priority data can be transmitted as high-priority data with probability $\gamma_i(j)$. There is no need to wait for transmission in the low-priority queue. Fourth, if the common buffer area is full, the low priority data that arrives is discarded with probability 1.

In this queuing system, we consider a State-dependent-based priority jump strategy as follow

$$\gamma_i(j) = \frac{i + j + \beta_2 \cos(\pi i/I) - \cos(\pi j/J)}{\beta_2 I + J} \tag{9}$$

where $\beta_2, \beta_2 > 1$ is the adjustment parameter.

4 QBD Process and Matrix Geometry Method

4.1 Quasi-Birth and Death Process Method

In order to analyze the transmission behavior of high and low priority data under the jump strategy respectively, we regard the two types of data entering the same buffer as two virtual queues. We use a two-dimensional Markov chain to model the data transmission of cellular user. The system state can be represented by the number of data packets (i,j) in the two queues in the system. The state process is defined as $\{X(t), t \geq 0\}$ where i represents the number of HP data packets, and j represents the number of LP data packets. Fig. 3 shows state transition diagram. The behavior of queuing is evaluated by QBD. We rearrange (i,j) the state of the queue length and treat it as a one-dimensional process. Definition, where is named level and is phase.

The transition probability matrix $\mathbf{P}$ is given in the form of a block matrix, as follows:

$$\mathbf{P} = \begin{bmatrix} \mathbf{B_0} & \mathbf{C_0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A_0} & \mathbf{B_1} & \mathbf{C_1} & \mathbf{0} & \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \mathbf{A_1} & \mathbf{B_2} & \mathbf{C_2} & \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \mathbf{A}_{L-1} & \mathbf{B}_{L-1} & \mathbf{C}_{L-1} \\ \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} & \mathbf{A}_L & \mathbf{B}_L \end{bmatrix} \tag{10}$$

where the sub-matrix represents the transfer between levels, the sub-matrix on the diagonal $\mathbf{B}_i$ represents the set of events in which the level (HP queue length) remains unchanged after the data packet arrives and is served in a time slot, while the phase (LP queue length) plus 1, minus 1, or remains unchanged, $L_i \to L_i, i \in [0, L]$. The expressions of the elements in all sub-matrix $\in \mathbb{R}^{L\times L}$ are as follows:

$$\mathbf{B}_0 = \begin{bmatrix} B_{0,0}^{0,0} & B_{0,0}^{0,1} & 0 & \cdots & \cdots & 0 \\ B_{0,1}^{0,0} & B_{0,1}^{0,1} & B_{0,0}^{0,1} & 0 & \cdots & 0 \\ 0 & B_{0,2}^{0,1} & B_{0,2}^{0,2} & B_{0,0}^{0,1} & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_{0,L-1}^{0,L-2} & B_{0,L-1}^{0,L-1} & B_{0,L-1}^{0,L} \\ 0 & \cdots & \cdots & \cdots & B_{0,L}^{0,L-1} & B_{0,L}^{0,L} \end{bmatrix} \tag{11}$$

where $B_{0,k}^{0,k}$ represent the situation of the high priority queue length is 0, and the low priority queue remains unchanged, expressed as:

$$B_{0,k}^{0,k} = \begin{cases} 1-a, k=0 \\ (1-a)(1-b) + ba(1-\mu_{HP})(1-\gamma_0(k)), 1 \le k \le L-1 \\ 1-b+ba(1-\mu_{HP})(1-\gamma_0(L)), k=L \end{cases} \tag{12}$$

The high priority queue length is 0, and the phase (LP queue length) plus 1 or minus 1, expressed as:

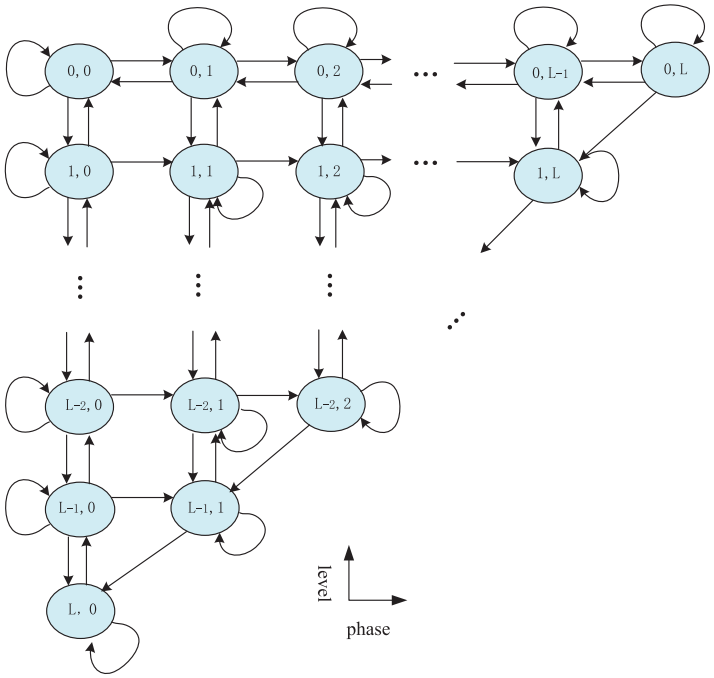

Figure 3: State transition diagram of priority queuing model with common buffer

$$B_{0,k+1}^{0,k}=\begin{cases}(1-a)\,b, 0\le k<L-1\\ b, k=L-1\end{cases} \tag{13}$$

$$B_{0,k-1}^{0,k}=\begin{cases}a\,(1-\mu_{HP})\,(1-\gamma_0\,(0))\,, k=1\\ a\,(1-\mu_{HP})\,(1-\gamma_0\,(k-1))\,(1-b)\,,\,, 1\le k\le L-2\end{cases} \tag{14}$$

For $1\le i\le L-1$, because the high and low priority data share the same buffer, that is $i+j\le L$, so the element in $\mathbf{B}_i$ exceeds the line of $L-i$ is 0.

$$\mathbf{B}_i=\begin{matrix}(i,0)\\(i,1)\\\vdots\\(i,L-i-1)\\(i,L-i)\\\vdots\\(i,L)\end{matrix}\begin{bmatrix}B_{i,0}^{i,0} & B_{i,0}^{i,1} & 0 & 0 & \cdots & 0 & 0\\ 0 & B_{i,1}^{i,1} & B_{i1}^{i,2} & 0 & \ddots & \ddots & \vdots\\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & B_{i,L-i-1}^{i,L-i-1} & B_{i,L-i-1}^{i,L-i} & 0 & 0\\ 0 & \ddots & \ddots & 0 & B_{i,L-i}^{i,L-i} & 0 & 0\\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots\\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0\end{bmatrix} \tag{15}$$

where

$$B_{i,k}^{i,k}=\begin{cases}(1-a)\,(1-b)+b\,[a\mu_{HP}+a\,(1-\mu_{HP})\,\gamma_i\,(k)]\,, 0\le k<L-i\\ (1-a)\,(1-b)\,, k=L-i\end{cases} \tag{16}$$

$$B_{i,k-1}^{i,k}=(1-b)\,[a\,(1-\mu_{HP})\,(1-\beta_i\,(k-1))]\,, 0\le k\le L-i \tag{17}$$

For the boundary conditions $i=L$ of the sub-matrix,

$$\mathbf{B}_L=\begin{bmatrix}(1-a\mu_{LP})\,(1-b) & 0 & 0 & \cdots & 0\\ 0 & 0 & \ddots & \cdots & \vdots\\ 0 & \ddots & \ddots & \ddots & 0\\ \vdots & \ddots & 0 & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 0\end{bmatrix} \tag{18}$$

The sub-matrix $\mathbf{C}_i$ indicates $L_{i-1}\to L_i, i\in[1,L]$, which contains all the events of adding a HP data packet. The added data packet may be an arriving HP data packet, or a data packet from the LP queue with probability $\gamma_i\,(j)$. The boundary condition $\mathbf{C}_0$ represents the state transition situation of $L_0\to L_1$, since the initial queue length of the HP queue is 0, the LP queue length will $j\to j-1$.

$$\mathbf{C}_0=\begin{matrix}(0,0)\\(0,1)\\\vdots\\(0,L-1)\\(0,L)\end{matrix}\begin{bmatrix}C_{0,0}^{1,0} & 0 & 0 & \cdots & 0\\ C_{0,1}^{1,0} & C_{0,1}^{1,1} & 0 & \cdots & \vdots\\ 0 & \ddots & \ddots & \ddots & \vdots\\ \vdots & \ddots & C_{0,L-1}^{1,L-2} & C_{0,L-1}^{1,L-1} & 0\\ 0 & \cdots & \cdots & C_{0,L}^{1,L-1} & 0\end{bmatrix} \tag{19}$$

where

$$C_{0,k}^{1,k}=\begin{cases}a\mu_{HP}+a\,(1-\mu_{HP})\,\gamma_0\,(0)\,, k=0\\ [a\mu_{HP}+a\,(1-\mu_{HP})\,\gamma_0\,(k)]\,(1-b)\,, 1\le k\le L-1\end{cases} \tag{20}$$

$$C_{0,k}^{1,k-1} = \left[a\mu_{HP} + a\left(1-\mu_{HP}\right)\beta_0\left(k\right)\right]b_{LP}, 1 \le k \le L \tag{21}$$

For $1 \le i \le L-2$, the HP queue length $i \ge 1$. Based on the principle that HP data is sent first, there is no transfer of $j \to j-1$, the elements except the diagonal are all 0.

$$\mathbf{C}_i = \begin{array}{c}(i,0)\\(i,1)\\\vdots\\(i,L-i-1)\\(i,L-i)\\\vdots\\(i,L)\end{array}\begin{bmatrix} C_{i,0}^{i+1,0} & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & C_{i,1}^{i+1,1} & 0 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & C_{i,L-i-1}^{i,L-i-1} & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & C_{i,L-i}^{i,L-i} & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 \end{bmatrix} \tag{22}$$

where

$$C_{i,k}^{i+1,k} = \left[a\mu_{HP} + a\left(1-\mu_{HP}\right)\gamma_i\left(k\right)\right]\left(1-b\right), 0 \le k \le L-i-1 \tag{23}$$

For boundary conditions of $i = L-1$,

$$\mathbf{C}_{L-1} = \begin{array}{c}(L-1,0)\\(L-1,1)\\\vdots\\(L-1,L-1)\\(L-1,L)\end{array}\begin{bmatrix} a\mu_{HP}\left(1-b\right) & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{24}$$

The sub-matrix $\mathbf{A}_i$ below the diagonal indicates that the HP queue adds a data packet after a service period, namely $L_i \to L_{i-1}, i \in [1, L]$. For $0 \le i \le L-1$, $\mathbf{A}_i$ means that the packet with no probability is added to the HP queue.

$$\mathbf{A}_i = \begin{array}{c}(i+1,0)\\(i+1,1)\\\vdots\\(i+1,L-1)\\(i+1,L)\end{array}\begin{bmatrix} b\left(1-a\right), & a\left(1-\mu_{HP}\right)\left(1-\gamma_0\left(k-1\right)\right)b & 0 & \cdots & 0 \\ 0 & b\left(1-a\right) & a\left(1-\mu_{HP}\right)\left(1-\gamma_1\left(k-1\right)\right)b & \cdots & \vdots \\ 0 & \ddots & \ddots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & b\left(1-a\right) & a\left(1-\mu_{HP}\right)\left(1-\gamma_{L-1}\left(k-1\right)\right)b \\ & & & 0 & 0 \end{bmatrix} \tag{25}$$

4.2 Matrix Geometry Method to Solve Two-dimensional Markov Chain

The steady-state solution of the quasi-birth-death process can be obtained by solving the linear balance equations, which is the column matrix of steady-state probability. In order to use the regular structure of the block matrix, we will divide the quasi-birth-death process into:

$$\Pi = [\Pi_0, \Pi_1, \cdots, \Pi_{L_{\max}}] \tag{26}$$

where

$$\Pi_i = [\Pi_{i,0}, \Pi_{i,1}, \Pi_{i,2}, \cdots \Pi_{i,L_{\max}}], i \in [0, L_{\max}] \tag{27}$$

From the linear balance equation, in order to solve the steady-state probability Π from (26), the transition probability matrix $\mathbf{P}$ is required. The analysis of the Section 3.1 shows that it depends on the arrival rate, transmission probability and jump probability. Among them, the D2D user density of the non-empty queue and the potential D2D mode selection determine the following parameters, which are the average arrival rate of HP and LP data, the probability of successful transmission and the jump probability. In particular, the probability of successful transmission depends on the interference from non-empty queue users, and the interference is related to the steady-state probability distribution of HP and LP. It can be clearly seen that there is an interactive relationship between the steady-state distribution of the queuing model and the successful transmission probability calculated by the thinned PPP. We use the fixed-point iteration method to solve the causal relationship to obtain the steady-state probability distribution, as shown in Algorithm 1.

Algorithm 1: A fixed point iteration method to calculate the steady-state probability distribution.

Initialization Set ε, $i = 1$ and, $\Pi_0^{(0)} = [1, 0, 0, \cdots 0]$.
Step 1: Use $\Pi^{(i-1)}$ to determine the probability of a non-empty queue among potential D2D users.
Step 2: Calculate the probability of successful transmission b according to $\lambda_d{}' = P(r_d < \mu)\eta\lambda_d$,
Step 3: Substituting b and jump probability, calculate the transition probability matrix $\mathbf{P}$,
Step 4: Substitute $\mathbf{P}$ into $\Pi^{(i)} = \Pi^{(i-1)}\mathbf{P}$, $\Pi^{(i-1)}\mathbf{1} = 1$ to get $\Pi^{(i-1)}$,
Step 5: Update $\Pi^{(i-1)}$ by using $\Pi^{(i)}$, and obtain $\Pi^{(i)} = \left[\Pi_0^{(i)}, \Pi_1^{(i)}, \cdots, \Pi_{L_{\max}}^{(i)}\right]$, $i = i + 1$
Step 6: Steps 1–5 until $\left|\Pi^{(i)} - \Pi^{(i-1)}\right| \leq \varepsilon$ return $\Pi^{(i)}$;
Step 7: Obtain the steady-state probability distribution.

4.3 Performance Parameters

After the steady-state probability distribution is obtained by Algorithm 1, performance parameters such as average queue length, average throughput, average delay and packet dropping probability can be expressed as the following expressions. The expressions of the average captain of HP and LP are given as:

$$Q_{HP} = \sum_{i=0}^{L} \sum_{j=0}^{L-i} (i - \gamma_i(j)) \Pi_{i,j} \tag{28}$$

$$Q_{LP} = \sum_{j=0}^{L} \sum_{i=0}^{L-j} (j + \gamma_i(j)) \Pi_{i,j} \tag{29}$$

Given the steady-state probability distribution and the probability of successful transmission, the average throughput of HP and LP queues can be derived as:

$$Thr_{HP} = \left(1 - \sum_{j=0}^{L} \Pi_{0,j}\right) \cdot b \tag{30}$$

$$Thr_{LP} = \left(1 - \sum_{i=0}^{L} \Pi_{i,0}\right) \cdot \sum_{j=0}^{L} \Pi_{0,j} \cdot b \tag{31}$$

In addition, Little's law can be used to evaluate the average waiting time Del_{HP} and Del_{LP} required before sending HP and LP packets in the queue.

$$Del_{HP} = \frac{\sum_{i=0}^{L} \sum_{j=0}^{L-i} (i - \gamma_i(j)) \Pi_{i,j}}{\left(1 - \sum_{j=0}^{L} \Pi_{0,j}\right) \cdot b} \tag{32}$$

$$Del_{LP} = \frac{\sum_{j=0}^{L} \sum_{i=0}^{L-j} (j + \gamma_i(j)) \Pi_{i,j}}{\left(1 - \sum_{i=0}^{L} \Pi_{i,0}\right) \cdot \sum_{j=0}^{L} \Pi_{0,j} \cdot b} \tag{33}$$

Furthermore, we can give the closed-form expressions of the packet dropping probability as:

$$p_{drop}^{HP} = a\mu_{HP} \sum_{j=0}^{L} \Pi_{L,j} \tag{34}$$

$$p_{drop}^{LP} = a\mu_{LP} \sum_{i=0}^{L} \Pi_{i,L} \tag{35}$$

5 Numerical and Simulation Results

In this section, we use MATLAB to analyze and verify the correctness of the mathematical expressions of the above performance parameters, and explore the system performance by changing the parameters in Tab. 1. In the following numerical simulations, unless otherwise specified, we consider a simulation area 20 $km \times 20$ km where the base station is located in a circular area with $R = 500$ m, and the cellular users are evenly distributed among them. The D2D transmitter is distributed uniformly in accordance with the density of PPP. All links experience independent and identically distributed Rayleigh fading with unit variance. In order to show the impact of changes in the proportion of HP data cache in the total cache on performance, we define high priority cache ratio as ϕ. The simulation results are carried out by using MATLAB and repeat for over 10,000 iterations.

Table 1: Simulation parameters

Density of D2D user pairs (λ_d)	10 users/km^2
Density of BS (λ_b)	1 BS/km^2
Data packet size (L_0)	100 bits
Time slot duration (T_0)	1 ms
Transmit power of D2D (p_D)	20 dBm

(Continued)

Table 1: Continued

Additive white gaussian noise power	−104 dBm
D2D-TX to D2D-RX transmission distance (r_d)	25 m
Buffer capacity (L)	50 packets
The receiver sensitivity of BS (ρ_{BS})	−80 dBm
The path loss factor of D2D link (α_D)	4
The path loss factor of CU link (α_C)	3.5

Fig. 4 is a three-dimensional graph of the LP queue length in the cache with the average arrival rate and the ratio of the HP data cache to the total cache capacity under the priority jump queuing strategy (PJQS) scheme. We set $p_{HP} = 0.2$. It can be seen from the figure that with the increase of the data arrival rate, the captain of the LP gradually increases. When the high-priority cache occupies a relatively small amount, that is, when the low-priority data has more space to enter the queue, the LP queue length grows faster and the curve is steeper. On the contrary, when the high-priority queue cache space increases, for example, the proportion is 80%, most of the low-priority data cannot enter the cache, resulting in a reduction of the queue length to 10. Then the LP packet dropping probability will increase sharply.

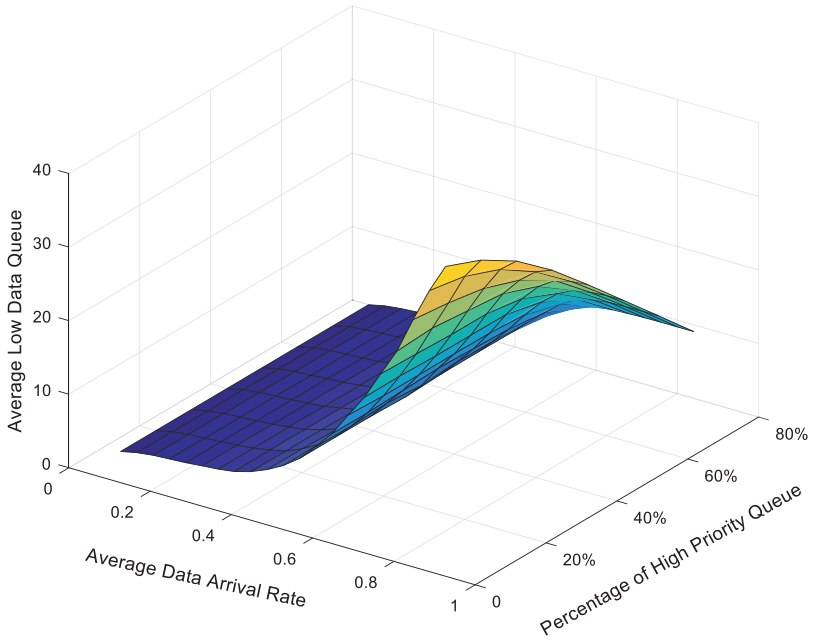

Figure 4: The LP data queue length with PJQS scheme

Fig. 5 shows the curve of queue length of the LP queue with the data arrival rate under different HP buffer ratios. We set $p_{HP} = 0.2$. It can be seen from the small backlog of low-data queue that PJQS has better performance than non-PJQS overall. We compare the curves where is $\phi = 0.5$ and $\phi = 0.8$. If the HP cache ratio ϕ is low, more LP data enters the cache, and the queue length is larger. When $\phi = 0.5$, and the data arrival rate reaches 0.6 under the no-jumping strategy, the LP queue quickly saturates, because the HP queue helps alleviate the transmission pressure of a large amount of LP data.

When the arrival rate is close to 1 and $\phi = 0.8$, all curves tend to overlap which is because the limited LP data cache space. When the data arrival rate is large, the PJQS has no effect on the queue length.

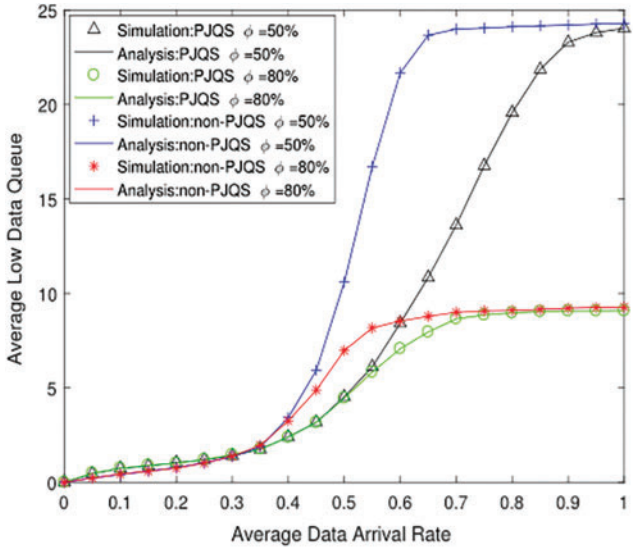

Figure 5: The LP data queue length under different ϕ

Fig. 6 is a figure of LP data packet dropping probability under PJQS and non-PJQS. Fig. 6a is PJQS scheme and Fig. 6b is non-PJQS scheme. The x-axis represents the average data arrival rate, the y-axis represents the proportion of the buffer allocated for the HP queue in the total buffer. It can be seen from Fig. 6a that when HP cache part is small, the increase in the average arrival rate has little effect on LP packet loss rate. The backlog of priority queues can maintain a very low dropping probability even if the arrival rate is large. When the proportion of HP buffers increases, especially when they account for more than 60% of the total buffer, the dropping probability increases sharply. It is because LP queue has insufficient buffer space, which causes overflow and packet loss. Therefore, the proportion of high-and LP buffers in the total buffer should be appropriately allocated to maximize the use of buffer capacity. In Fig. 6b, no matter how much the HP cache occupies, the increase in data arrival rate has a huge impact on dropping probability. When the arrival rate reaches 0.4, the dropping probability increases sharply. However, the buffer ratio has little effect on the change of the curve. This is because without the help of the jump queuing strategy, the LP queue is saturated when the arrival rate is 0.4, and the HP data is sent first and LP data cannot be sent. Comparing the two figures, we can see that the cache strategy reduces the maximum packet loss rate of LP queue from 0.25 to 0.05, which clearly shows the effectiveness of the jump queuing strategy.

Fig. 7 shows the delay of low priority queue *vs.* arrival rate.It can be seen that the numerical results coincide with the simulation results. When the data arrival rate is less than 1, on the premise of the same proportion, the delay of the priority jump queuing strategy is lower than that of non jump queuing strategy. As the data arrival rate is close to 1 and the capacity of both queues is close to full, the advantages of the jump strategy are gradually disappearing, and the curve of whether there is a jump strategy is close to the same. In the case of non jump queuing strategy, the delay curve of $\phi = 0.5$ is higher than $\phi = 0.8$, because the smaller the cache capacity allocated to low priority data, the shorter the queue length of low priority, and the less time required for transmission, the smaller the delay. There is the same trend in the curve with the priority jump queuing strategy.

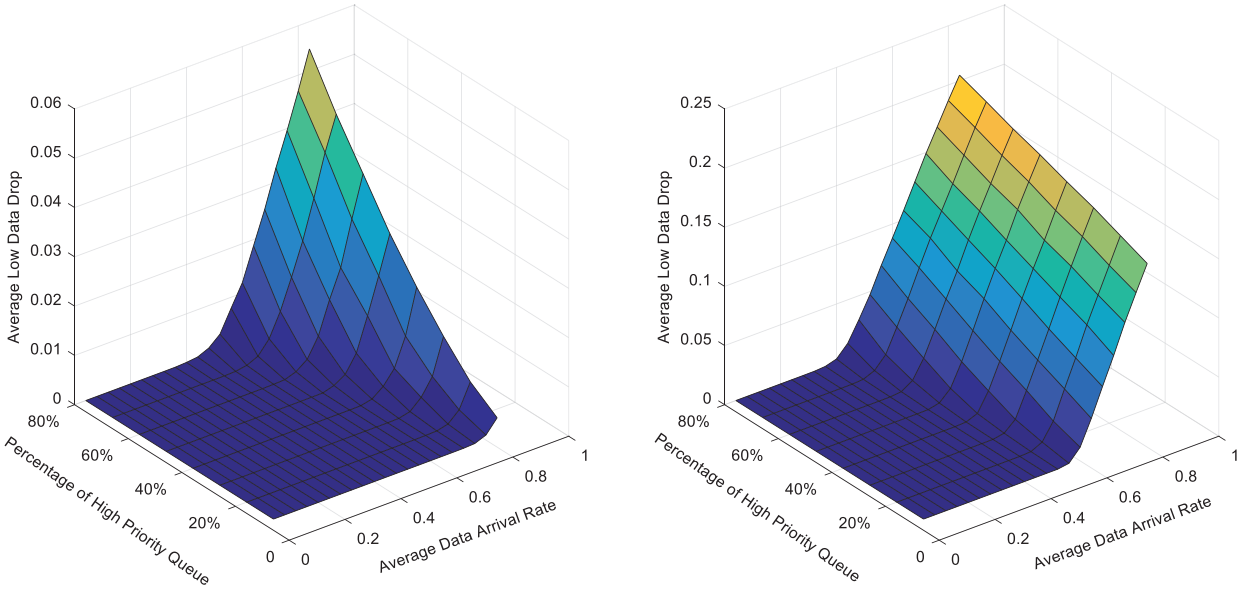

Figure 6: The LP data drop of PJQS and non-PJQS

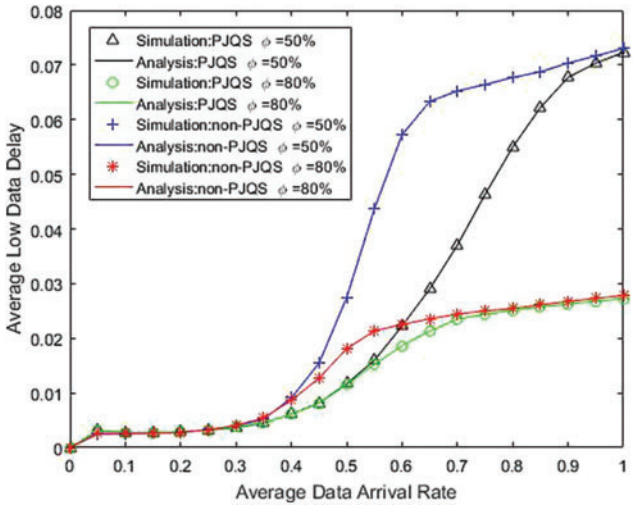

Figure 7: The LP data delay with data arrival rate

Fig. 8 is a three-dimensional graph of the low-priority data delay as a function of the average arrival rate and the proportion of high-priority data caches in the total cache under PJQS scheme. When the high-priority data cache occupies a small amount of total cache space, it is allocated to the low-priority data cache. The high-level data space is large, and more low-priority data can be accommodated at this time. When the arrival rate is high, the longer the queue length of the low-priority queue is, the longer the sending delay is, and when the high-priority data cache occupies a large proportion, with the increase of the average arrival rate, the increase of the delay slows down. This is because the limited low-priority data buffer space prevents the arriving data from entering the queue.The queue length is shorter, the delay is lower.

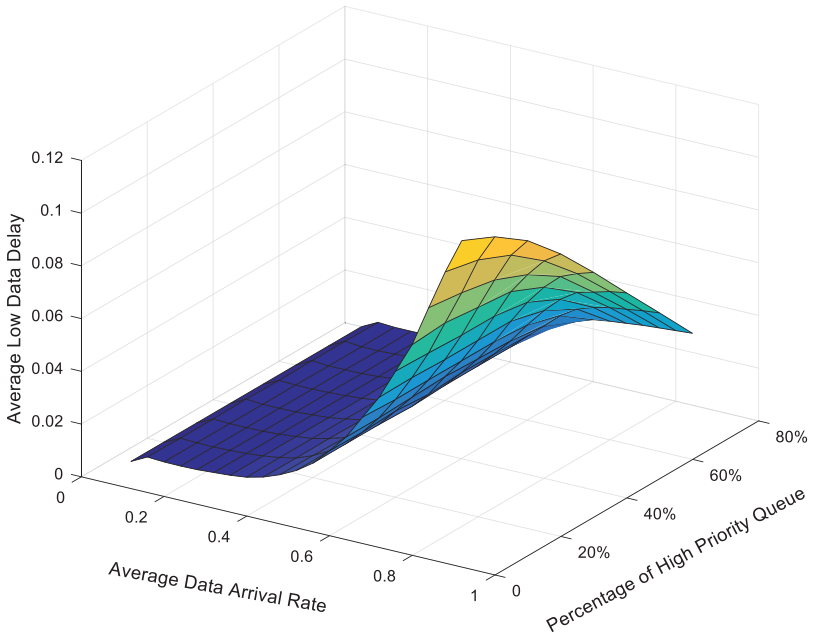

Figure 8: The LP data delay with PJQS scheme

Fig. 9 shows the average throughput of HP and LP *vs.* arrival rate. It can be seen that the numerical results coincide with the simulation results. Since data arrival rate of HP is much smaller than its maximum transmission capacity, HP data throughput increases linearly. As the data arrival rate increases, the throughput of LP first rises rapidly and then decreases. This is because the two queues store less data. The HP queue sends some LP data with probability $\gamma_i(j)$ to increase the LP throughput. As the both queue lengths increase, the HP queue needs to send arriving HP data. When the transmission capacity is not enough to continue sending data packets from the LP queue, and the LP throughput will decrease. The numerical results are basically coincide with the simulation results.

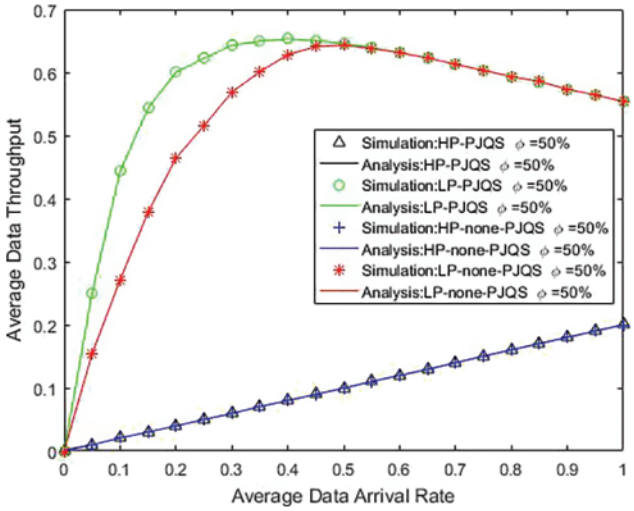

Figure 9: The average throughput with data arrival rate

6 Conclusions

A spatiotemporal mathematical model in the D2D underlying cellular network was proposed. Different from the existing results, we focus on the multi-priority business mode. Stochastic geometry and priority queuing theory are adapted in this network. We set a priority jump strategy to provide transmission opportunities for LP data and reduce the packet dropping probability of the LP queue. In addition, considering the transmission mode of potential D2D users in the cellular network, we adopt a distance-based D2D mode selection scheme. To show the relationship between the D2D user queues, we model the positions of D2D users with Non-empty queue buffers using a thinned Poisson point process to derive the CCDF of the D2D receiving SINR. Secondly, we use a two-dimensional Geo/G/1 Markov chain to describe the queue model with priority jumps for D2D users, and the QBD method evaluates the queue state transition process. Finally, a series of expressions of performance parameters are derived, such as average queue length, average throughput, average delay and packet dropping probability. Simulation analysis confirmed the correctness of the numerical analysis. In addition, by comparing the packet loss rate of priority queues with and without jump strategy, the rationality of the model proposed in this paper is explained. In future work, we can apply the queuing model and analytical approach presented in this paper to conducting performance studies on D2D-assisted wireless caching networks. Specially, we can discuss the service transmission model of D2D heterogeneous cellular network in full duplex mode, and further analyze various types of jump strategies. We can also consider D2D equipment powered by pre-charged batteries from the perspective of green communication.The harvested energy can signifificantly improve the lifetime of the device and the network. we will concentrate on using more realistic assumptions to build a dynamic model in further research. We hope our work can provide a more accurate theoretical basis for the design of the actual system.

Acknowledgement: Guangjun Liang and Jianfang Xin conceived and designed the experiments; Guangjun Liang and Lingling Xia performed the experiments; Xueli Ni and Yi Cao analyzed the data; Guangjun Liang wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding Statement: The work is supported by 2020 Major Natural Science Research Project of Jiangsu Province Colleges and Universities: Research on Forensic Modeling and Analysis of the Internet of Things (20KJA520004), 2020 Open Project of National and Local Joint Engineering Laboratory of Radio Frequency Integration and Micro-assembly Technology: Research on the Security Performance of Radio Frequency Energy Collection Cooperative Communication Network (KFJJ20200201), 2021 Jiangsu Police Officer Academy Scientific Research Project: Research on D2D Cache Network Resource Optimization Based on Edge Computing Technology (2021SJYZK01), High-level Introduction of Talent Scientific Research Start-up Fund of Jiangsu Police Institute (JSPI19GKZL407).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] T. Ahmad, I. Khan, A. Irshad, S. Ahmad, A. T. Soliman *et al.,* "Spark spectrum allocation for D2D communication in cellular networks," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 6381–6394, 2022.

[2] X. R. Zhang, W. F. Zhang, W. Sun, X. M. Sun and S. K. Jha, "A robust 3-D medical watermarking based on wavelet transform for data protection," *Computer Systems Science & Engineering*, vol. 41, no. 3, pp. 1043–1056, 2022.
[3] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally and M. A. Javed, "A survey of device-to-device communications: Research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 3, no. 3, pp. 2133–2168, 2018.
[4] X. R. Zhang, X. Sun, X. M. Sun, W. Sun and S. K. Jha, "Robust reversible audio watermarking scheme for telemedicine and privacy protection," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.
[5] Y. Zhong, M. Haenggi, T. Q. S. Quek and W. Zhang, "On the stability of static poisson networks under random access," *IEEE Transactions on Communications*, vol. 64, no. 7, pp. 2985–2998, 2016.
[6] Y. Zhong, T. Q. S. Quek and X. Ge, "Heterogeneous cellular networks with spatio-temporal traffic: Delay analysis and scheduling," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1373–1386, 2017.
[7] H. H. Yang, Y. Wang and T. Q. S. Quek, "Delay analysis of random scheduling and round robin in small cell networks," *IEEE Wireless Communications Letters*, vol. 7, no. 6, pp. 978–981, 2018.
[8] I. A. M. Balapuwaduge, L. Jiao, V. Pla and F. Y. Li, "Channel assembling with priority-based queues in cognitive radio networks: Strategies and performance evaluation," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 630–645, 2014.
[9] J. Walraevens, B. Steyaert and H. Bruneel, "Performance analysis of a single-server ATM queue with a priority scheduling," *Computers & Operations Research*, vol. 30, no. 12, pp. 1807–1829, 2003.
[10] Z. Zhao, C. Yi, J. Cai and H. Cao, "Queueing analysis for medical data transmissions with delay-dependent packet priorities in WBANs," in *Int. Conf. on Wireless Communications & Signal Processing*, Yangzhou, pp. 24–37, 2016.
[11] J. F. Xin, Q. Zhu and G. J. Liang, "Performance analysis of D2D underlying cellular networks based on dynamic priority queuing model," *IEEE Access*, vol. 7, no. 1, pp. 27479–27489, 2020.
[12] A. K. Hamid, F. N. Al-Wesabi, N. Nemri, A. Zahary and I. Khan, "An optimized algorithm for resource allocation for D2D in heterogeneous networks," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 2923–2936, 2022.
[13] R. Z. Ahamad, A. R. Javed, S. Mehmood, M. Z. Khan, A. Noorwali *et al.,* "Interference mitigation in D2D communication underlying cellular networks: Towards green energy," *Computers, Materials & Continua*, vol. 68, no. 1, pp. 45–58, 2021.
[14] S. Huang, B. Liang and J. Li, "Distributed interference and delay aware design for D2D communication in large wireless networks with adaptive interference estimation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3924–3939, 2017.
[15] H. Elsawy and E. Hossain, "Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4147–4161, 2014.