

Research Article

Mobility-Centric Analysis of Communication Offloading for Heterogeneous Internet of Things Devices

Dmitry Kozyrev,^{1,2} Aleksandr Ometov^{1,2} ,³ Dmitri Moltchanov,³ Vladimir Rykov,^{1,4} Dmitry Efrosinin,^{1,2} Tatiana Milovanova,¹ Sergey Andreev,³ and Yevgeni Koucheryavy³

¹RUDN University, Moscow, Russia

²V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

³Tampere University of Technology, Tampere, Finland

⁴Gubkin Russian State University of Oil and Gas, Moscow, Russia

Correspondence should be addressed to Aleksandr Ometov; aleksandr.ometov@tut.fi

Received 6 April 2018; Revised 30 June 2018; Accepted 12 July 2018; Published 5 August 2018

Academic Editor: Sudarshan Guruacharya

Copyright © 2018 Dmitry Kozyrev et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Today, the number of interconnected Internet of Things (IoT) devices is growing tremendously followed by an increase in the density of cellular base stations. This trend has an adverse effect on the power efficiency of communication, since each new infrastructure node requires a significant amount of energy. Numerous enablers are already in place to offload the scarce cellular spectrum, thus allowing utilization of more energy-efficient short-range radio technologies for user content dissemination, such as moving relay stations and network-assisted direct connectivity. In this work, we contribute a new mathematical framework aimed at analyzing the impact of network offloading on the probabilistic characteristics related to the quality of service and thus helping relieve the energy burden on infrastructure network deployments.

1. Introduction and Motivation

The uncontrollable growth in the numbers of interconnected Internet of Things (IoT) devices has a tremendously adverse effect on the existing wireless networks [1]. It has been recently shown that there are more than four million base stations (BSs) each consuming an average of 2.3MWh per month [2], and the volume of newly deployed BSs is growing exponentially, especially in the developing countries. Such an extreme energy consumption burden has its influence on the energy costs, greenhouse effect, and global impact on climate in general [3].

One of the ways to control the increase in the numbers of newly deployed BSs is to efficiently utilize the existing system infrastructure [4] by, e.g., attempting to switch between the radio technologies instead of “densifying” the deployment area [5]. This paradigm is known as heterogeneous networking [6] as part of the next-generation (5G) systems and beyond. However, seamless implementation of this approach is associated with many practical difficulties, e.g., when

the existing infrastructure is owned by different service providers.

Another potential enabler of green communication is device-to-device (D2D) connectivity [7, 8]. This technique allows reducing power consumption and improving network capacity and throughput by utilizing, e.g., unlicensed wireless spectrum controlled by the operator [9]. Basically, a mobile device equipped with two or more radio interfaces has an opportunity to utilize short-range wireless links to communicate with its neighboring nodes by reaching common goals, like data caching, edge computing, coverage extension, etc. [10]. As cellular spectrum remains expensive in terms of capacity and energy, it could be freed with D2D offloading, thus enabling green communication [11]. In this regime, the cellular network is only responsible for controlling the connection between the user equipment (UE) devices.

One more way to improve the communication quality is based on the use of unmanned aerial vehicles (UAVs). Historically, UAVs have already demonstrated their applicability for weather monitoring, forest fire detection, traffic control, etc.

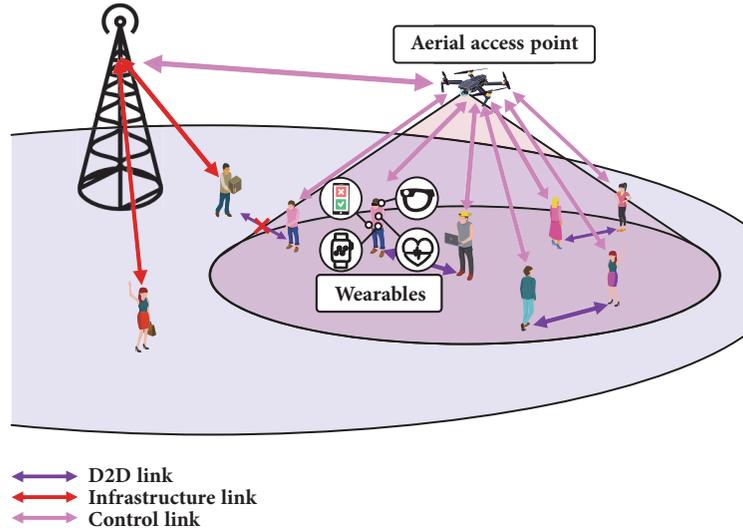


FIGURE 1: Joint operation of conventional and aerial cells serving D2D links.

[12, 13]. Among the many applications, the use of UAVs for achieving high data-rate wireless communication is expected to play an essential role in future wireless systems. In fact, UAV-aided wireless communication becomes a promising solution to provide wireless connectivity for devices without infrastructure coverage [14]. There has been recent interest in utilizing aerial access points (AAPs) mounted on drones, which could be utilized for offering on-demand connectivity in cases of imminent network overload. Further, AAPs may also be utilized to support the scenarios of Public Protection and Disaster Relief (PPDR) where the connectivity is unavailable due to an unpredictable situation [15].

In addition, a recent exciting trend that overtakes the market is a broad penetration of wearable devices, such as AR-glasses, smart watches, heart rate monitors, and many others. The functions of those are mostly related to human condition monitoring. Note that wearable devices typically have smaller form-factors and thus are limited in terms of computation and energy, which substantiates a natural step of employing one of the devices as a gateway for delivering wearables-generated data to the cloud. Most commonly, a smartphone is utilized as this more powerful node [16].

In this paper, we assess the architecture of (beyond-) 5G systems where a mobile user is equipped with many wearable devices that are continuously exchanging data with the surrounding nodes as well as the cloud via a gateway node. For this purpose, the conventional cellular infrastructure is typically employed, augmented with a D2D-based caching-enabled relay or utilizing other heterogeneous networking solutions. In our scenario, an AAP is further used together with D2D communication to achieve the ultimate goal of offloading the existing cellular network as much as possible. Here, the AAP is expected to provide a control link to the D2D pairs located within the overloaded area of interest, thus relieving the limited cellular spectrum to some extent. This work proposes a new mathematical model to evaluate the probabilistic characteristics of the proposed system architecture.

The rest of the paper is organized as follows. Our system model of the considered scenario is introduced in Section 2. This model is solved for the performance metrics of interest in Section 3. An example illustrating the applicability of the proposed approach is offered in Section 4. The last section concludes the paper.

2. System Model

2.1. Deployment Model. Consider a city square with a mass event (e.g., a concert) in progress, which is shown in Figure 1. The area of interest has a certain number of heterogeneous UEs acting as gateways for their proximate wearable devices and is partially covered by the AAP's connectivity providing aid in communication offloading via D2D link management. New mobile UE can arrive in the cell according to a stochastic process and leave it after a certain random time due to user mobility. All objects in the cell can disseminate their content of interest by utilizing the D2D links.

Since the AAP coverage within the area of interest is limited and potentially provided by the event holders, wireless connectivity outside of this zone is only available through the conventional infrastructure links. The primary goal of the network in the considered scenario is to offload as many communication sessions as possible onto the direct links, thus enabling more power efficient content dissemination. Note that the D2D capacity regarding the quality of service is constrained due to a limited number of channels in one collision domain; i.e., one can estimate the number of D2D pairs that can operate simultaneously under the AAP coverage for a given application.

Since the conventional metrics in such a scenario are well studied in numerous works [17, 18], the primary focus of this paper is set on less popular performance indicators. First, we consider *connection unavailability*, i.e., a situation where one of UE nodes intends to connect to another UE over a D2D link but that second UE is already occupied or a certain UE attempts to reach another device outside of the cell but there

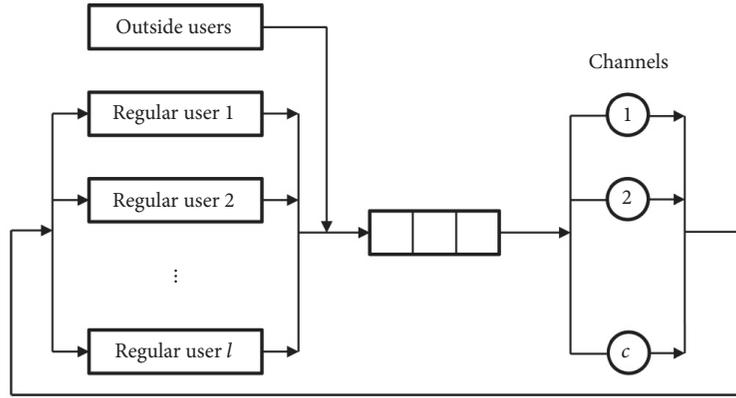


FIGURE 2: Proposed queuing model.

are no more available D2D links. Another metric of interest is the probability of a connection to be discarded, i.e., when one of the UE nodes leaves the AAP coverage while still having an active connection with another UE inside the same cell.

2.2. Mobility Model. We assume that the initial UE positions follow a Poisson point process (PPP) in \mathfrak{R}^2 with the spatial density of λ_U . Let $\lambda_A = \lim_{t \rightarrow 0} p_A(0, t)/t$, where $p_A(t)$ is the probability of a session initiation over the interval $(0, t)$. It is the temporal intensity of the arrival process from a single UE. Using the superposition property of the point processes, we note that the session arrival process is Poisson with the intensity of $\lambda_A \lambda_U S_A$ sessions per second [19], where S_A is the area of the zone of interest. The choice of the UE that initiates a session is random. Hence, the geometric locations of users associated with a session are distributed uniformly within the AAP's coverage area [20].

To capture the mobility behavior, we assume that humans move according to a random direction model (RDM, [21]), since it characterizes the essentials of random movement while still allowing for a tractable analysis. According to the RDM formulation, the UE first chooses its direction of movement uniformly within $(0, 2\pi)$ and then travels in the selected direction for an exponential amount of time with the parameter ν at the constant speed of ν .

2.3. Model Formalization. We assess the AAP service process by utilizing a queuing system model illustrated in Figure 2. We assume that l sessions exist in the system, while there is a certain external flow of impatient heterogeneous sessions, which arrive from the outside of the system according to a Poisson process. Consider that the regular users can connect directly over D2D, while the external users may employ both the infrastructure and the D2D connections.

We also assume that the AAP-served UE arrivals follow a marked point process:

$$Z = \{Z_n = (\xi_n, M_n), n = 1, 2, \dots\}, \quad (1)$$

where ξ_n is the arrival time of n^{th} user and M_n is the UE type, $M_n = i \in \{1, 2, \dots, l\}$. The service times of type- i UE, i.e.,

the duration of an active session initiated by the i -th UE, and B_i are all random variables with the cumulative distribution function (CDF):

$$B_i(t) = \mathbf{P}\{B_i \leq t\} \quad (i = 1, 2, \dots, l). \quad (2)$$

Each external session requires exactly one server. The duration of time spent by the i -th user in the D2D service zone is a random variable G_i with the CDF:

$$G_i(t) = \mathbf{P}\{G_i \leq t\} \quad (i, j = 1, 2, \dots, l). \quad (3)$$

A session is lost during service if the corresponding UE leaves the cell before this session is completed. An arriving session is lost when there are no vacant servers (channels) at the time of its arrival. Finally, a session can be lost if the distance between the communicating users becomes higher than a specified value. We remind that, for the introduced model, we are interested in the performance metrics that characterize the reliability of the system. Primarily, we derive the following parameters: session loss probability, system unavailability, and reliability of a connection.

3. Performance Analysis

In this section, we analyze the proposed system. First, we follow a stochastic geometry approach to characterize the input parameters. Then, we proceed by assessing the specified queuing system for the performance metrics of interest.

3.1. System Parametrization. To completely parametrize our model, we need to provide the following input parameters: (i) distribution of the time interval between UE entries into the D2D service zone, (ii) distribution of the time spent in the D2D service area, (iii) probability that UE having an active D2D session is going to leave the D2D zone before this session ends.

Following [22], the time distribution between the cases when the UE moving according to the RDM model and being initially distributed uniformly in a certain area of interest hits a certain subspace follows an exponential distribution. Note that this result holds when the mean run length in the RDM

model, v/v , is at least $1/4$ of the zone of interest; see [22] for more details. Recall that the stationary distribution of UEs moving in a certain zone is uniform [21].

To determine the parameter of an exponential distribution, we consider an area increment of the D2D service zone ΔS that corresponds to the unit speed v . Assuming a circularly shaped D2D service zone and recalling that the direction of the UE movement is distributed uniformly in $(0, 2\pi)$, we have the following approximation for the temporal intensity of users entering the D2D service zone:

$$\lambda = \frac{1}{2}\lambda_U(\pi(r+v)^2 - \pi r^2) = \frac{1}{2}\lambda_U\pi v(2r+v), \quad (4)$$

where λ_U is the spatial intensity of UE in the area of interest and r is the radius of the D2D service zone. Observe that the factor $1/2$ accounts for a fraction of UEs moving towards the D2D service zone.

The exact distribution of time spent by the UE moving according to the RDM in a certain enclosed subspace is unknown even for simple configurations of the subspace [23]. In this case, the common procedure is to use a diffusion approximation with a suitable diffusion coefficient [24]. However, taking into account our assumptions regarding the mean run length and the diameter of the D2D service zone, we observe that in most cases the UE crosses the D2D service zone without changes in its direction. The length of a random chord, L , which specifies the distance traveled within the D2D service zone, is given by [25]

$$f_L(x) = \frac{2}{\pi\sqrt{4r^2 - x^2}}, \quad 0 < x < 2r, \quad (5)$$

with the mean value of $E[L] = 4r/\pi$, where r is the radius of the D2D service zone. The distribution of time spent in the D2D service zone is obtained by scaling the pdf $f_L(x)$ by the movement speed v [26], which leads to

$$f_T(x) = \frac{2v}{\pi\sqrt{4r^2 - (xv)^2}}, \quad 0 < x < \frac{2r}{v}, \quad (6)$$

where T is the time spent in the D2D service zone.

Finally, we require the probability that the UE leaves the D2D service zone before its active session expires. We have two cases to consider: (i) the session has been initiated before entering the D2D service area and (ii) the session has been initiated when the UE is already in the service area. In the former case, owing to the memoryless property, we do not need to track the backward recurrence time of the session and the sought probability is $p_1 = \mathbb{P}\{T - D > 0\}$, where D is the service duration. Since T and D are independent, by employing a convolution we obtain

$$\begin{aligned} p_1 &= \int_0^\infty \int_0^t f_T(t) f_D(t-x) dt dx \\ &= \int_0^{2r/v} \int_0^t \frac{2v}{\pi\sqrt{4r^2 - (xv)^2}} \lambda e^{-\lambda(t-x)} dt dx, \end{aligned} \quad (7)$$

which can be produced with a numerical integration.

Let us now determine the probability that the UE leaves the D2D service zone before its session is completed given that it was initiated after entering the zone, p_2 . Observe that since a session is assumed to be initiated randomly and should progress over the time the UE spends in the D2D service zone, it is distributed uniformly over T . Then, the forward recurrent time that the UE spends in the zone once its session was initiated at the time instant x is [26]

$$f_T(t+x|x) = \frac{f_T(t+x)}{1-F_T(x)}, \quad (8)$$

where $F_T(x)$ is the CDF of time spent by the UE in the D2D service zone, which is given by

$$F_T(x) = \frac{1}{\pi} 2v \tan^{-1} \left(\frac{x}{\sqrt{4r^2 v^2 - x^2}} \right), \quad 0 < x < 2rv. \quad (9)$$

Substituting (9) into (8), we obtain

$$f_T(t+x|x) = \frac{2(4r^2 - (t+x)^2/v^2)^{-1/2}}{2v \tan^{-1}(x/v\sqrt{\pi - 4r^2 - x^2/v^2})}. \quad (10)$$

The unconditional pdf of time that the UE spends in the D2D coverage zone after initiating its session is given by

$$f_{T_R}(t) = \int_0^\infty f_T(t+x|x) f_T(x) dx. \quad (11)$$

Further, the sought probability is given by

$$p_2 = \int_0^\infty \int_0^t f_{T_R}(t) f_D(t-x) dt dx, \quad (12)$$

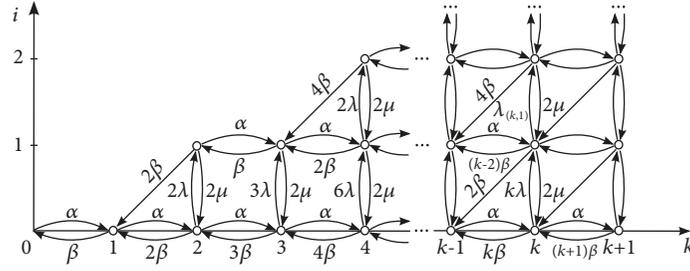
which can be produced with a numerical integration.

3.2. Problem Setting. To solve the problem formulated above, we introduce the system state space E . Let us denote the system state by the following triplet:

$$x = (k, i, \vec{d}), \quad (13)$$

where $k = 0, 1, 2, \dots$ is the number of infrastructure controlled UEs, i ($0 \leq i \leq \min\{c, k + \lfloor l/2 \rfloor\}$) is the number of occupied channels, and $\vec{d} = (d_1, \dots, d_l)$ is the vector of states of the UE. The value of d_j is set according to the following rules: (i) $d_j = 0$ when the j -th user does not have any ongoing sessions; (ii) $d_j = i$ when the j -th UE is currently connected to the i -th UE directly; and (iii) $d_j = *$ when the j -th UE utilizes one of the infrastructure links.

The overall state space E is then defined as a set of triplets: $E = \{x = (k, i, \vec{d}) : k \in \mathbb{Z}, i \in \mathbb{Z}, \vec{d} = (d_1, \dots, d_l)\}$. We introduce a random process $X(t) = \{K(t)I(t), \vec{D}(t), t \in \mathbb{R}_+\}$ evolving over E , which enters the state $x \in E$ if the system at time t is in the state x . We also introduce the state probabilities of the system $\pi(x, t) = \mathbb{P}\{X(t) = x\}$ and the


 FIGURE 3: Transition graph of the process X .

steady-state probabilities $\lim_{t \rightarrow \infty} \pi(x, t) = \pi(x)$. The steady-state distribution of this process $\vec{\pi}$ is characterized by linear equations in the following form:

$$\begin{aligned} \vec{Q}\vec{\pi} &= \vec{0}^T, \\ \vec{\pi}^T \vec{e} &= 1, \end{aligned} \quad (14)$$

where \vec{e} is the vector of ones of the appropriate size and Q is the infinitesimal generator of the Markov process. One may solve it by using the conventional techniques; see, e.g., [27].

4. A Numerical Example

To demonstrate the applicability of the proposed model, consider the simplest case without the pure infrastructure connected UEs, when

- (i) the users arrive into the system according to a Poisson flow of intensity α ;
- (ii) the distribution of the sojourn time of any user in the cell is exponential with the parameter β :

$$B(t) = \beta e^{-\beta t}; \quad (15)$$

- (iii) the process of contacts between the users is also Poisson with the intensity of λ ;
- (iv) the duration of the user sessions G has an exponential distribution with the parameter μ :

$$G(t) = \mu e^{-\mu t}; \quad (16)$$

- (v) only $c = 3$ infrastructure links are available for the connections.

We also assume full control of the D2D links as managed by the network operator. To characterize the system behavior, consider a two-dimensional stochastic process $X = \{X(t) = (K(t), I(t)), t \in R_+\}$ with the state space:

$$E = \bigcup_{k \geq 0} E_k, \quad \text{with } E_k = \{x = (k, i) : i = \overline{0, 3}\}, \quad (17)$$

where k is the number of users in the system and i is the number of occupied D2D links. The state space of the process

X is demonstrated in Figure 3 together with its transition graph.

The subsets $E_0 = \{(0, 0)\}$ and $E_1 = \{(1, 0)\}$ represent a single state, the subsets $E_2 = \{(2, 0), (2, 1)\}$ and $E_3 = \{(3, 0), (3, 1)\}$ contain two states, the subsets $E_4 = \{(4, 0), (4, 1), (4, 2)\}$ and $E_5 = \{(5, 0), (5, 1), (5, 2)\}$ contain three states, and only for $k \geq 6$ the number of states is equal to four. Note that in case where the servers and/or users are heterogeneous, to describe the system behavior with a Markov process, the subsets of states E_k with a fixed number k of users in the system should be detailed in order to show the configuration of the busy servers. With the help of the constructed stochastic process $X(t)$, we will study the quality-of-service (QoS) metrics in the considered system, i.e., the probability of a connection loss (when a user with no ongoing session leaves the service area), the probability of connection unavailability (when there are no vacant channels in the system), and the average number of busy channels.

Under the given assumptions, the process $X(t)$ may be considered as Markov, and its transition intensities have the following forms:

- (1) α_k is the transition intensity from the state (k, i) to the state $(k + 1, i)$, which occurs when a user enters the system.
- (2) $\beta_{(k,i)}$ is the transition intensity from the state (k, i) to the state $(k - 1, i)$, which occurs when a user with no ongoing session leaves the system.
- (3) $\beta_{(k,i)}^-$ is the transition intensity from the state (k, i) to the state $(k - 1, i - 1)$, which occurs when a user with an ongoing session leaves the system.
- (4) $\lambda_{(k,i)}$ is the transition intensity from the state (k, i) to the state $(k, i + 1)$, which occurs when a user with no ongoing session initiates a session.
- (5) $\mu_{(k,i)}$ is the transition intensity from the state (k, i) to the state $(k, i - 1)$, which occurs when a user with an ongoing session ends this session.

Consider the state (k, i) . Since i channels are occupied by exactly $2i$ users (2 users are connected via a single D2D link), the number of users with no ongoing session is $k - 2i$, the number of users with an ongoing session is $2i$, and the number of possible ways to choose a pair of users out of $k - 2i$ nonbusy users to initiate a session is $\binom{k-2i}{2}$. Therefore,

we arrive at the following expressions for the transition intensities:

$$\begin{aligned}
\alpha_k &= \alpha, \\
\beta_{(k,i)} &= (k - 2i) \beta, \\
\beta_{(k,i)}^- &= 2i\beta, \\
\lambda_{(k,i)} &= \binom{k - 2i}{2} \lambda, \\
\mu_{(k,i)} &= 2i\mu,
\end{aligned} \tag{18}$$

where $i = 0$ for $k = 0, 1$, $i = 0, 1$ for $k = 2, 3$, $i = 0, 1, 2$ for $k = 4, 5$, and $i = 0, 1, 2, 3$ for $k \geq 6$.

The process transition intensity matrix has the following blockwise structure

$$Q = \begin{pmatrix} B_0 & C_0 & 0 & 0 & 0 & \dots \\ A_1 & B_1 & C_1 & 0 & 0 & \dots \\ 0 & A_2 & B_2 & C_2 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{19}$$

where the k -th blockwise row corresponds to the subset E_k of states for which k users are present in the system, A_k describes the output from this subset to the subset with $k - 1$ users in the system, B_k describes the transitions inside the subset with k users in the system, and C_k contains the output intensities from the subset E_k to the subset E_{k+1} . Transition matrices for the considered example are as follows:

$$B_0 = -\alpha, \tag{20}$$

$$C_0 = \alpha.$$

$$A_1 = \beta,$$

$$B_1 = -(\alpha + \beta), \tag{21}$$

$$C_1 = (\alpha \ 0).$$

$$A_2 = \begin{pmatrix} 2\beta \\ 2\beta \end{pmatrix}, \tag{22}$$

$$C_2 = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}.$$

$$B_2 = \begin{pmatrix} -(\alpha + 2\beta + \lambda) & \lambda \\ 2\mu & -(\alpha + 2\beta + 2\mu) \end{pmatrix}. \tag{23}$$

$$A_3 = \begin{pmatrix} 3\beta & 0 \\ 2\beta & \beta \end{pmatrix}, \tag{24}$$

$$C_3 = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \end{pmatrix}.$$

$$B_3 = \begin{pmatrix} -(\alpha + 3\beta + 3\lambda) & 3\lambda \\ 2\mu & -(\alpha + 3\beta + 2\mu) \end{pmatrix}. \tag{25}$$

$$A_4 = \begin{pmatrix} 4\beta & 0 \\ 2\beta & 2\beta \\ 0 & 4\beta \end{pmatrix}, \tag{26}$$

$$C_4 = \begin{pmatrix} \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \alpha & 0 \end{pmatrix}.$$

$$B_4 = \begin{pmatrix} -(\alpha + 4\beta + 6\lambda) & 6\lambda & 0 \\ 2\mu & -(\alpha + 4\beta + 2\lambda + 2\mu) & 2\lambda \\ 0 & 2\mu & -(\alpha + 2\mu + 4\beta) \end{pmatrix}. \tag{27}$$

$$A_5 = \begin{pmatrix} 5\beta & 0 & 0 \\ 2\beta & 3\beta & 0 \\ 0 & 4\beta & \beta \end{pmatrix}, \quad (28)$$

$$C_5 = \begin{pmatrix} \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \alpha & 0 \end{pmatrix}.$$

$$B_5 = \begin{pmatrix} -(\alpha + 5\beta + 10\lambda) & 10\lambda & 0 \\ 2\mu & -(\alpha + 5\beta + 3\lambda + 2\mu) & 3\lambda \\ 0 & 4\mu & -(\alpha + 4\mu + 5\beta) \end{pmatrix}. \quad (29)$$

Further, for $k \geq 6$, the matrices have the form

$$A_k = \begin{pmatrix} k\beta & 0 & 0 & 0 \\ 2\beta & (k-2)\beta & 0 & 0 \\ 0 & 4\beta & (k-4)\beta & 0 \\ 0 & 0 & 6\beta & (k-6)\beta \end{pmatrix}, \quad (30)$$

$$C_k = \begin{pmatrix} \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & \alpha \end{pmatrix}.$$

$$B_k = \begin{pmatrix} -b_{(k,0)} \binom{k}{2} \lambda & 0 & 0 \\ 2\mu & -b_{(k,1)} \binom{k-2}{2} \lambda & 0 \\ 0 & 4\mu & -b_{(k,2)} \binom{k-4}{2} \lambda \\ 0 & 0 & 6\mu & -b_{(k,3)} \end{pmatrix}, \quad (31)$$

where the diagonal elements are

$$b_{k,0} = \left(\alpha + k\beta + \binom{k}{2} \lambda \right), \quad (32)$$

$$b_{k,1} = \left(\alpha + k\beta + \binom{k-2}{2} \lambda + 2\mu \right),$$

$$b_{k,2} = \left(\alpha + \binom{k-4}{2} \lambda + 4\mu + k\beta \right), \quad (33)$$

$$b_{k,3} = (\alpha + 6\mu + k\beta).$$

The blockwise form of the matrix Q represents the steady-state probability (s.s.p.) vector in the blockwise form of the appropriate dimensions that provides $\vec{\pi}^\top =$

$(\pi_0, \pi_1, \vec{\pi}_2, \dots, \vec{\pi}_k, \dots)$. Hence, the balance equations may be represented as

$$\begin{aligned} 0 &= -\alpha\pi_0 + \beta\pi_1, \\ 0 &= \alpha\pi_0 - (\alpha + \beta)\pi_1 + 2\beta(\pi(2,0) + \pi(2,1)), \\ 0 &= \pi_1 C_1 + \vec{\pi}_2^\top B_2 + \vec{\pi}_3^\top A_3, \\ 0 &= \vec{\pi}_{k-1}^\top C_{k-1} + \vec{\pi}_k^\top B_k + \vec{\pi}_{k+1}^\top A_{k+1}, \quad k \geq 3 \end{aligned} \quad (34)$$

with the normalizing condition

$$\pi_0 + \pi_1 + \sum_{k=2}^{\infty} \vec{\pi}_k^\top \vec{1} = 1, \quad (35)$$

where $\vec{1}$ is a column-vector, while its components are equal to 1.

By using A_k, B_k, C_k for $k \geq 3$, we arrive at

$$\begin{aligned} A_{k-1} \vec{1} &= (k-1) \beta \vec{1}, \\ B_k \vec{1} &= -(\alpha + k\beta) \vec{1}, \\ C_{k+1} \vec{1} &= \alpha \vec{1}. \end{aligned} \quad (36)$$

Further, we multiply (34) by $\vec{1}$ as

$$\pi_k = \vec{\pi}_k^\top \vec{1}, \quad (37)$$

with the following birth-and-death s.s.p. equation:

$$\alpha\pi_{k-1} - (\alpha + k\beta)\pi_k + (k+1)\beta\pi_{k+1} = 0. \quad (38)$$

For $k \geq 3$, this gives us

$$\alpha\pi_{k-1} k\beta\pi_k = \alpha\pi_k (k+1)\beta\pi_{k+1}. \quad (39)$$

However, the same equalities can be obtained from (34), thus confirming it for $k \leq 3$:

$$0 = \alpha\pi_0\beta\pi_1 = \alpha\pi_1 2\beta\pi_2 = \alpha\pi_k (k+1)\beta\pi_{k+1}. \quad (40)$$

The solution to this system of equations for all $k \geq 0$ with the normalizing condition (35) and with $\rho = \alpha/\beta$ is

$$\pi_k = \frac{\alpha}{k\beta} \pi_{k-1} = \frac{\rho^k}{k!} \pi_0 = \frac{\rho^k}{k!} e^{-\rho}, \quad (41)$$

which is a Poisson distribution of the number of users in the system.

In order to calculate the probabilities of interest, there is a need to evaluate $\pi_{(k,i)}$, which requires a more detailed analysis. To achieve this goal, the process X on the separate subset of states should be considered as $E_k = \{(k, 0); (k, 1); (k, 2); (k, 3)\}$. Taking into account X and its transition graph (see Figure 3), it is clear that each of the subsets E_0, E_1 contains only one state; hence, the corresponding probabilities are given with (41) as

$$\begin{aligned} \pi_0 &= e^{-\rho}, \\ \pi_1 &= \rho e^{-\rho}. \end{aligned} \quad (42)$$

For the subsets E_k , the process X is a birth-death process with the birth $\lambda_{(k,i)} = \binom{k-2i}{2} \lambda$ and death $\mu_{(k,i)} = 2i\mu$ (see Figure 3) intensities. Therefore, the detailed balance equations may be utilized to establish the results for $k = 2, 3$:

$$\binom{k}{2} \lambda \pi_{(k,0)} = 2\mu \pi_{(k,1)}, \quad (43)$$

additionally for $k = 4, 5$

$$\binom{k-2}{2} \lambda \pi_{(k,1)} = 4\mu \pi_{(k,2)}, \quad (44)$$

at least for all $k \geq 6$, and

$$\binom{k-4}{2} \lambda \pi_{(k,2)} = 6\mu \pi_{(k,3)}. \quad (45)$$

For the probabilities of the subsets E_k , by taking into account (41) and using the notation $\gamma = \lambda/\mu$, the following expressions for the probabilities $\pi_{(k,i)}$ for $k \geq 2$ can be derived:

$$\pi_{(2,0)} = \frac{1}{1 + \gamma/2} \frac{\rho^2}{2!} e^{-\rho}, \quad (46)$$

$$\pi_{(2,1)} = \frac{\gamma/2}{1 + \gamma/2} \frac{\rho^2}{2!} e^{-\rho},$$

$$\pi_{(3,0)} = \frac{1}{1 + (3/2)\gamma} \frac{\rho^3}{3!} e^{-\rho}, \quad (47)$$

$$\pi_{(3,1)} = \frac{(3/2)\gamma}{1 + (3/2)\gamma} \frac{\rho^3}{3!} e^{-\rho},$$

$$\pi_{(4,0)} = \frac{1}{1 + 3\gamma(1 + (1/4)\gamma)} \frac{\rho^4}{4!} e^{-\rho},$$

$$\pi_{(4,1)} = \frac{3\gamma}{1 + 3\gamma(1 + (1/4)\gamma)} \frac{\rho^4}{4!} e^{-\rho}, \quad (48)$$

$$\pi_{(4,2)} = \frac{(3/4)\gamma^2}{1 + 3\gamma(1 + (1/4)\gamma)} \frac{\rho^4}{4!} e^{-\rho},$$

$$\pi_{(5,0)} = \frac{1}{1 + 5\gamma(1 + (3/4)\gamma)} \frac{\rho^5}{5!} e^{-\rho},$$

$$\pi_{(5,1)} = \frac{5\gamma}{1 + 5\gamma(1 + (3/4)\gamma)} \frac{\rho^5}{5!} e^{-\rho}, \quad (49)$$

$$\pi_{(5,2)} = \frac{(15/4)\gamma^2}{1 + 5\gamma(1 + (3/4)\gamma)} \frac{\rho^5}{5!} e^{-\rho}.$$

For $k \geq 6$, the following holds:

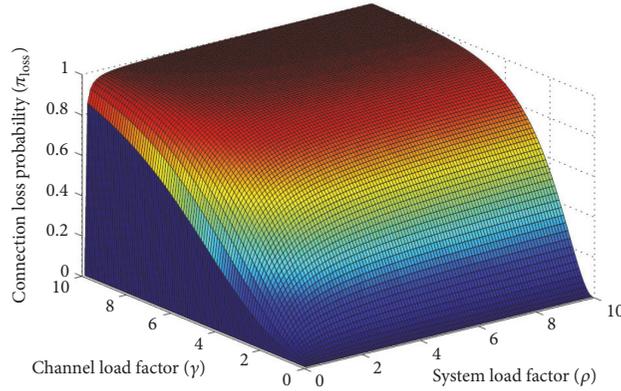
$$\begin{aligned} \pi_{(k,0)} &= c_k \frac{\rho^k}{k!} e^{-\rho}, \\ \pi_{(k,1)} &= c_k \frac{k(k-1)}{4} \gamma \frac{\rho^k}{k!} e^{-\rho}, \\ \pi_{(k,2)} &= c_k \frac{k(k-1)}{4} \gamma \left(1 + \frac{(k-2)(k-3)}{8} \gamma \right) \frac{\rho^k}{k!} e^{-\rho}, \\ \pi_{(k,3)} &= c_k \frac{k(k-1)}{4} \\ &\quad \cdot \gamma \left(1 + \frac{(k-2)(k-3)}{8} \gamma \left(1 + \frac{(k-4)(k-5)}{12} \gamma \right) \right) \\ &\quad \cdot \frac{\rho^k}{k!} e^{-\rho}, \end{aligned} \quad (50)$$

where

$$\begin{aligned} c_k &= \left[1 + \left(\frac{k(k-1)}{4} \gamma \left(3 \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{(k-2)(k-3)}{8} \gamma \left(2 + \frac{(k-4)(k-5)}{12} \gamma \right) \right) \right) \right]^{-1}. \end{aligned} \quad (51)$$

Based on the above set of probabilities, it is possible to calculate the sought QoS characteristics, such as the following.

- (1) Probability of a connection loss π_{loss} due to leaving the AAP coverage area by one of the users: Here, a loss of connection occurs in case where the process transitions from the state (k, i) to the state $(k-1, i-1)$, which is only possible for the states (k, i) with $i > 0$

FIGURE 4: Probability π_{loss} versus different values of γ and ρ .

and $k \geq 2$. Hence, the connection loss may be derived as

$$\begin{aligned}
 \pi_{\text{loss}} = & \pi_{(2,1)} \frac{2\beta}{\alpha + 2\beta + 2\mu} + \pi_{(3,1)} \frac{2\beta}{\alpha + \beta + 2\mu + 2\beta} \\
 & + \pi_{(4,1)} \frac{2\beta}{\alpha + 4\beta + \lambda + 2\mu} + \pi_{(4,2)} \frac{4\beta}{\alpha + 4\beta + 2\mu} \\
 & + \pi_{(5,1)} \frac{2\beta}{\alpha + 5\beta + 3\lambda + 2\mu} + \pi_{(5,2)} \frac{4\beta}{\alpha + 5\beta + 4\mu} \\
 & + \sum_{k \geq 6} \left(\pi_{(k,1)} \frac{2\beta}{\alpha + k\beta + \binom{k-2}{2} \lambda + 2\mu} \right. \\
 & + \pi_{(k,2)} \frac{4\beta}{\alpha + k\beta + \binom{k-4}{2} \lambda + 4\mu} \\
 & \left. + \pi_{(k,3)} \frac{6\beta}{\alpha + k\beta + 6\mu} \right). \tag{52}
 \end{aligned}$$

The plots of the loss probabilities π_{loss} versus the parameters ρ and γ are given in Figure 4. Since the parameter γ is defined as a ratio between the session initiation intensity λ and the intensity of the session termination flow μ , the physical meaning of this parameter is the channel load. The parameter $\rho = \alpha/\beta$ is the system load factor (or traffic load), which can be interpreted as a relative rate of populating the cell by the outside users.

Then, it is obvious that, at the zero level of ρ , i.e., when no users are entering the cell, the connection loss probability π_{loss} remains zero for any value of the parameter γ . However, as the system load factor ρ increases, the probability π_{loss} also grows, as it can be observed in Figure 4: the higher the value of the parameter γ is, the steeper the curve of π_{loss} on γ becomes. This behavior is easily explained by the fact that both parameters affect the occupation intensity for the free channels, which are limited in their number, by the users entering the cell. Since

the users leave the cell regardless of whether they occupy the server (channel) or not, the probability of a connection loss increases with the growth of either parameter.

- (2) Probability that i D2D channels are currently occupied is

$$\pi_{i,\text{busy}} = \sum_{k \geq 2} \pi_{(k,i)}, \tag{53}$$

while the mean number of busy channels is

$$\pi_{\text{busy-mean}} = \sum_{1 \leq i \leq 3} i \pi_{i,\text{busy}}, \tag{54}$$

and the connection unavailability intensity is $\beta \pi_{\text{busy-mean}}$.

- (3) Probability of a connection unavailability state is possible to derive only if the number of users is not less than 6. Hence, the probability of a connection unavailability becomes

$$\pi_{\text{unav}} = \sum_{x:i=3} \pi(x) = \sum_{k \geq 6} \pi_{(k,3)}, \tag{55}$$

while the intensity of unavailable connections is $\lambda \pi_{\text{unav}}$.

The plots for the unavailability probabilities π_{unav} versus the parameters ρ and γ are displayed in Figure 5. The behavior of the curves of the connection unavailability probability π_{unav} is similar in many ways to that of π_{loss} , as π_{unav} directly correlates with the parameters γ and ρ . If the value of any of the associated parameters is zero, the probability of a connection unavailability is also zero, since either the system lacks users (parameter ρ), or there are no users willing to initiate a connection session, i.e., zero loads on the channels (parameter γ). As each of these parameters grows, the probability to have an available unoccupied channel for establishing another connection between the activating users decreases.

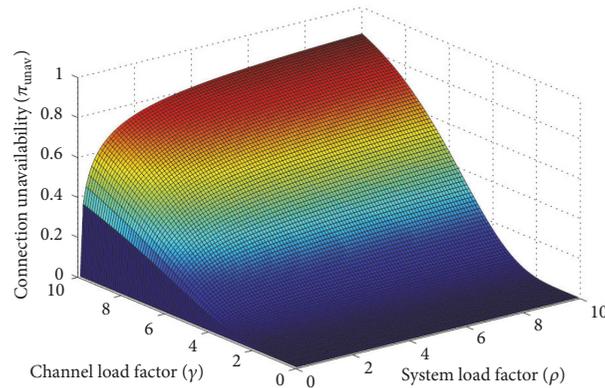


FIGURE 5: Probability π_{unav} versus different values of γ and ρ .

The considered model admits useful generalizations that may include different types of heterogeneous users, different types of input processes for the outside users, and different mechanisms of their service.

5. Conclusion

In today's wireless networks, multiple enablers are available to offload the expensive cellular spectrum, thus allowing utilization of more efficient short-range radio technologies for user content dissemination. This work proposed a novel mathematical framework that enables assessing the impact of network offloading on the probabilistic characteristics related to the quality of service. We demonstrated that our developed model may be employed when an aerial access point conducts D2D link management based on the channel and system loads. The results reported useful knowledge of connection unavailability and connection loss probabilities.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The publication was supported by the Ministry of Education and Science of the Russian Federation (Project no. 2.882.2017/4.6).

References

- [1] M. Z. Shakir, K. A. Qaraqe, H. Tabassum, M.-S. Alouini, E. Serpedin, and M. Imran, "Green heterogeneous small-cell networks: Toward reducing the CO₂ emissions of mobile communications industry using uplink power adaptation," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 52–61, 2013.
- [2] C. Gao, X. Sheng, J. Tang, W. Zhang, S. Zou, and M. Guizani, "Joint mode selection, channel allocation and power assignment for green device-to-device communications," in *Proceedings of IEEE International Conference on Communications (ICC)*, pp. 178–183, Sydney, NSW, June 2014.
- [3] B. Lee, "Energy Efficiency Gain of Cellular Base Stations with Large-Scale Antenna Systems for Green Information and Communication Technology," *Sustainability*, vol. 9, no. 7, p. 1123, 2017.
- [4] T. Zhou, N. Jiang, D. Qin, Z. Liu, and C. Li, "Joint Cell Selection and Activation for Green Communications in Ultra-Dense Heterogeneous Networks," in *Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, pp. 32–37, Guangzhou, China, July 2017.
- [5] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: challenges and solution approaches," *IEEE Wireless Communications Magazine*, vol. 22, no. 5, pp. 22–31, 2015.
- [6] V. Jungnickel, K. Manolakis, W. Zirwas et al., "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44–51, 2014.
- [7] J. An, K. Yang, J. Wu, N. Ye, S. Guo, and Z. Liao, "Achieving Sustainable Ultra-Dense Heterogeneous Networks for 5G," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 84–90, 2017.
- [8] A. H. Sakr, H. Tabassum, E. Hossain, and D. I. Kim, "Cognitive spectrum access in device-to-device-enabled cellular networks," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 126–133, 2015.
- [9] S. Andreev, J. Hosek, T. Olsson et al., "A unifying perspective on proximity-based cellular-assisted mobile social networking," *IEEE Communications Magazine*, vol. 54, no. 4, pp. 108–116, 2016.
- [10] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 40–48, 2014.
- [11] E. Yaacoub, H. Ghazza, and M.-S. Alouini, "A game theoretic framework for green hetnets using D2D traffic offload and renewable energy powered base stations," *Game Theory Framework Applied to Wireless Communication Networks*, pp. 333–367, 2015.

- [12] G. Pajares, "Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs)," *Photogrammetric Engineering and Remote Sensing*, vol. 81, no. 4, pp. 281–329, 2015.
- [13] L. Wang, B. Hu, and S. Chen, "Energy Efficient Placement of a Drone Base Station for Minimum Required Transmit Power," *IEEE Wireless Communications Letters*, pp. 1-1, 2018.
- [14] K. Gomez, S. Kandeepan, M. M. Vidal et al., "Aerial base stations with opportunistic links for next generation emergency communications," *IEEE Communications Magazine*, vol. 54, no. 4, pp. 31–39, 2016.
- [15] T. J. Tanzi, M. Chandra, J. Isnard, D. Camara, O. Sebastien, and F. Harivelo, "Towards "drone-based" disaster management: future application scenarios," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-8, pp. 181–189, 2016.
- [16] G. Aloï, G. Caliciuri, G. Fortino et al., "Enabling IoT interoperability through opportunistic smartphone-based mobile gateways," *Journal of Network and Computer Applications*, vol. 81, pp. 74–84, 2017.
- [17] K. Gomez, A. Hourani, L. Goratti, R. Riggio, S. Kandeepan, and I. Bucaille, "Capacity evaluation of Aerial LTE base-stations for public safety communications," in *Proceedings of the European Conference on Networks and Communications, EuCNC*, pp. 133–138, France, July 2015.
- [18] S. Chandrasekharan, S. Kandeepan, R. J. Evans et al., "Clustering approach for aerial base-station access with terrestrial cooperation," in *Proceedings of the IEEE Globecom Workshops, GC Wkshps 2013*, pp. 1397–1402, USA, December 2013.
- [19] J. F. Kingman, *Poisson processes*, vol. 3 of *Oxford Studies in Probability*, The Clarendon Press, Oxford University Press, New York, 1993.
- [20] D. Moltchanov, "Distance distributions in random networks," *Ad Hoc Networks*, vol. 10, no. 6, pp. 1146–1166, 2012.
- [21] P. Nain, D. Towsley, B. Liu, and Z. Liu, "Properties of random direction models," in *Proceedings of the Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, pp. 1897–1907, Miami, FL, USA.
- [22] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, no. 1-4, pp. 210–228, 2005.
- [23] G. H. Weiss and R. J. Rubin, "Random Walks: Theory and Selected Applications," in *Advances in Chemical Physics*, Advances in Chemical Physics, pp. 363–505, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1982.
- [24] S. Redner, *A Guide to First-Passage Processes*, Cambridge University Press, Cambridge, UK, 2001.
- [25] A. M. Mathai, *An introduction to geometrical probability*, vol. 1 of *Statistical Distributions and Models with Applications*, Gordon and Breach Science Publishers, Amsterdam, 1999.
- [26] S. M. Ross, *Introduction to probability models*, Elsevier/Academic Press, Amsterdam, Eleventh edition, 2014.
- [27] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, *Queueing networks and Markov chains*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., 1998.

