Delay Analysis of Social Group Multicast-Aided Content Dissemination in Cellular System

Jie Hu, *Member, IEEE*, Lie-Liang Yang, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Based on the common interest of mobile users (MUs) in a social group, the dissemination of content across the social group is studied as a powerful supplement to conventional cellular communication with the goal of improving the delay performance of the content dissemination process. The content popularity is modeled by a Zipf distribution to characterize the MUs' different interests in different contents. The factor of altruism (FA) terminology is introduced for quantifying the willingness of content owners to share their content. We model the dissemination process of a specific packet by a pure-birth-based Markov chain and evaluate the statistical properties of both the network's dissemination delay as well as of the individual user-delay. Compared to the conventional base station (BS)-aided multicast, our scheme is capable of reducing the average dissemination delay by about 56.5%. Moreover, in contrast to the BS-aided multicast, increasing the number of MUs in the target social group is capable of reducing the average individual user-delay by 44.1% relying on our scheme. Furthermore, our scheme is more suitable for disseminating a popular piece of content.**

***Index Terms*—Content dissemination, content popularity, factor of altruism, pure-birth based Markov chain, delay analysis.**

I. Introduction

A. Background and Related Works

As a combination of social science and mobile networks, mobile social networks (MSNs) [1] are attracting an increasing attention across the research community. In the context of MSNs, mobile users (MUs) may form a social group in order to cooperatively disseminate the content of common interest. There are substantial contributions to the performance analysis of epidemic forwarding [2] in mobile ad hoc networks (MANETs). In the context of MANETs, a two-dimensional *continuous time Markov chain* (CTMC) was proposed in [3] for evaluating the performance of a heterogeneous MANETs. To a further advance, the authors of [4] derived a tight upper bound of the flooding time, which is defined as the number of time-steps required for broadcasting a message from a source node to all nodes. Furthermore, in [5] the end-to-end message delivery delay using an epidemic forwarding protocol was investigated theoretically in a composite twin-layer network, which includes a physical MANET and a virtual social network.

However, epidemic forwarding [6] is often criticised as being an end-to-end routing protocol, because it consumes substantial resources of the intermediate nodes, which might not be interested in the information to be relayed. However, if MUs can form a social group and request the content of common interest together, epidemic forwarding becomes an efficient way of cooperatively disseminating the content in the target social group[1]. Content dissemination in purely distributed opportunistic networks was investigated in [7] and [8]. Epidemic forwarding aided content dissemination was invoked in [7], where the users share any content updates with others that they meet in order to improve the coverage quality and to increase the capacity. A socially-aware content placement algorithm was proposed in [8] for enhancing the opportunity of MUs to gain access to their contents of interest.

Some researches focused on a hybrid content dissemination approach. In [9] and [10], the authors investigated how the content providers and network operators can interact for the sake of efficiently distributing the contents with the aid of a coalition game. At the time of writing, epidemic forwarding aided content dissemination is widely studied for the sake of offloading tele-traffic from cellular networks. In [11], the authors proposed a framework for initial content-receiver selection in order to disseminate the content of common interest to as many subscribers as possible before interest in the content subsides. In [12], where MUs were categorised into "helpers" and "subscribers", several algorithms were designed for solving the optimisation problem of offloading multiple types of contents from the cellular networks.

The above-mentioned contributions [2]–[12] focused their attention on user-encounter-based MANETs or 'large-scale MSNs', where the mobile nodes are sparsely distributed across a large area. Typically a rudimentary physical layer model is assumed, namely that if a pair of nodes enter each other's transmission range, the packet can be successfully delivered from the source to the target. Hence, the delivery delay is dominated by the *inter-contact duration*[2] of mobile nodes [15], rather than by the wireless signal propagation. Due to

Manuscript received April 12, 2015; revised October 18, 2015 and February 10, 2016; accepted February 22, 2016. This work was supported by the RC-UK's India-UK Advanced Technology Centre (IU-ATC), that of the EU's concerto project and of the China Scholarship Council (CSC). The work of L. Hanzo was supported by the European Research Council under its Advanced Fellow Award. The associate editor coordinating the review of this paper and approving it for publication was B. Liang.

J. Hu is with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: hujienavy@outlook.com).

L.-L. Yang and L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lly@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2016.2535330

[1]Other MUs that do not belong to the target social group are not relied upon for assisting the content dissemination process.

[2]In these treatises, the inter-contact duration of MUs is commonly assumed to obey an exponential distribution, which is demonstrated in [13] with the aid of artificial or synthetic mobility models and in [14] by realistic measured mobility traces.

the underlying long inter-contact duration of the MUs, this user-encounter-based content dissemination is only capable of delivering delay-tolerant services in a large-scale area. As a result, the contributions of [2]–[12] belong to the category of *delay-tolerant networks* (DTNs). However, typically idealised simplifying assumptions are used in the literature of the DTN paradigm:

- The commonly assumed simplified physical layer model ignores the impact of transmit power, of the path-loss and of the multipath fading, etc.
- The cooperative user-encounter based content dissemination in DTNs is not suitable for delivering delay-sensitive services.

B. Motivations and Contributions

The conventional method of disseminating the delay-sensitive content of common interest relies on BS-aided multicast, where the BS is the sole transmitter. Since the BS-aided multicast has to guarantee the quality of service (QoS) at every content requester, the capacity of multicast channels is predetermined by the worst channel amongst those connecting the BS to the content requesters. In this case, due to the time-variant nature of wireless channels, when the BS multicasts a packet, some MUs may receive it earlier than their less fortunate counterparts. Then, the successful receivers have to remain silent, because the BS would not multicast the second packet, before all the MUs successfully receive the current one.

In high-user-density scenarios, the MUs often share common interest in delay-sensitive content. For instance, the crowd participating in the inauguration of the new Pope share common interest in close-up video-clips of the Pope on the podium. Similarly, supporters in a football stadium share common interest in video-clips of a spectacular goal from different angles or in the score updates from another stadium, as exemplified by Fig. 1. However, the conventional BS-aided multicast is an inefficient technique of disseminating the delay-sensitive content of common interest in these typical densely populated scenarios. The reason for this is two-fold:

- As the content requesters' density increases, the worst channel amongst those connecting the BS and the content requesters becomes even worse, which results in excessive dissemination delay [16].
- Since the dissemination delay is increased, the BS is engaged in multicasting for a longer period, which further delays all other services.

If local MUs form a social group for requesting the content of common interest from the BS together, local communications amongst MUs can be exploited for cooperatively multicasting the packets from the packet owners to the hitherto unserved MUs in the target social group[3]. The potential performance gain of this social group multicast aided content dissemination over the conventional BS-aided multicast arises from the following two benefits:

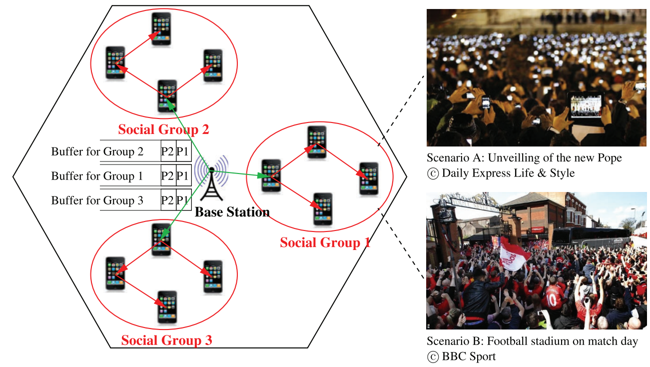


Fig. 1. Social group multicast aided content dissemination in cellular systems.

- Relying on the cooperative multicast of the multiple packet owners results in rich cooperative diversity gains, which in turn improves the packet delivery performance.
- Activating direct transmissions amongst the MUs is capable of reducing the distance between a transmitter and receiver pair, which in turn reduces the path-loss-induced channel attenuation between them.

Furthermore, since we offload the content dissemination task from the BS-aided multicast to the local communications amongst the social group members, the BS becomes capable of satisfying other communication demands, which consequently improves the efficiency of the BS's exploitation.

The size of the area covered by a social group should be carefully designed for different scenarios. If the area is as large as a macro-cell, cooperative user-encounter based communication amongst MUs is only suitable for disseminating delay-tolerant information, as we argued at the end of Section I-A. The best option for disseminating delay-sensitive information across a large area is that of classic BS-aided multicast. By contrast, if the area is relatively small, such as a circular area with a radius shorter than a hundred meters, which is comparable to the default transmission range of a MU[4], communication efficiency between a transmitter and receiver pair is dominated by the wireless signal propagation properties, rather than by their inter-contact duration. Hence, social group aided cooperative multicast is capable of significantly reducing the delay of the conventional BS-aided multicast, as we emphasized at the beginning of Section I-B. This scenario is termed as a "small-scale MSN" [15], where the channel attenuation factors dominate the associated delay characteristics [19]. Against this background, our novel contributions are as follows:

- A hybrid content dissemination approach is proposed, which relies both on BS-aided multicast [20] and on social group multicast aided content dissemination. This process is modelled by a *pure-birth based Markov chain (PBMC)*. Various factors that might affect the performance of the content dissemination are accounted for, including the path-loss-induced channel attenuation, the multipath fading and the users' altruistic versus selfish behaviours, which distinguishes our work from the existing literature of DTNs.

[3] A similar methodology of improving BS-aided multicast was also advocated in [17], which was mainly focused on the selection of the initial receivers. However, the authors of [17] have not analysed the content dissemination stage.

[4] New Wi-Fi protocols, such as 802.11n/ac [18], are capable of supporting a transmission range of hundreds of meters.

- We model the popularity of different pieces of contents by a Zipf distribution, which affects the specific formation of a social group and hence influences the dissemination process of the content of common interest across the target social group.
- Considering a specific packet of the content of common interest, we analyse the statistical properties of the dissemination delay, which is the time from the BS's instant of multicasting a packet until all the MUs in the target social group receive this packet. We also analyse the individual user-delay, which is the time spanning from the BS multicasting a packet until a specific MU receives this packet.
- The advantages of our social group multicast aided content dissemination scheme over the conventional BS-aided multicast are demonstrated by the mobility traces extracted from a realistic subway station scenario.

Note that improving the network infrastructure in high-user-density areas can certainly enhance the general communication experience of MUs, when supporting phone calls, texts, emails and basic data services. However, it may constitute an inefficient technique of disseminating the content of common interest. It may also be an unwise investment for the network operators, since people often temporarily get together for attending social events. Hence, improving the infrastructure capacity may be wasteful. By contrast, our social group multicast scheme constitutes a more economical and flexible solution for disseminating the content of common interest amongst the social group members, which is based on direct communications between the social group members. We will demonstrate that our social group multicast aided scheme outperforms the BS-aided multicast in terms of disseminating the popular content of common interest in high-user-density areas.

The rest of the paper is organized as follows. In Section II, our system model is introduced. In Section III, we analyse the delay metrics. Furthermore, the exact closed-form formulas are derived for two special cases in Section IV. Our numerical results are provided in Section V. Finally, we conclude in Section VI.

II. System Overview

Similar to the BS-controlled device-to-device communication services of the LTE network [21], our system operates by obeying a centralised-control regime combined with a decentralised-transmission paradigm[5], where the BS acts as a centralised controller in order to support the functions of synchronisation[6], of social group formation as well as of coordination and resource allocation for multiple content owners etc. By contrast, the information transmission is carried out by direct communications between a transmitter and receiver pair.

[5]This paradigm has been considered as a part of the forthcoming '5G' regime, known as the 'LTE-Assisted Wi-Fi Direct' technique [22], where the control signalling exchange is carried out by the LTE-based BS, while the information transmission is realised by the Wi-Fi-based direct communication between a transmitter and receiver pair.

[6]Since the MUs in the cellular system rely on regular control signalling exchange with associated BSs, they can readily synchronise with associated BSs and hence also with each other.

TABLE I
THE REQUEST PROBABILITIES OF $\mathcal{M} = 10$ RANKED POPULAR CONTENTS FOR BOTH $\alpha = 0.56$ [24] AND $\alpha = 1.0$ [23]

Content	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
$\alpha = 0.56$	21.4%	14.6%	11.6%	9.9%	8.7%
$\alpha = 1.0$	34.1%	17.1%	11.4%	8.5%	6.8%
Content	$\mathcal{C}_6$	$\mathcal{C}_7$	$\mathcal{C}_8$	$\mathcal{C}_9$	$\mathcal{C}_10$
$\alpha = 0.56$	7.9%	7.2%	6.7%	6.3%	6.0%
$\alpha = 1.0$	5.7%	4.9%	4.3%	3.8%	3.4%

A. Content Popularity and Social Group Formation

The interest of a MU in a specific piece of content $\mathcal{C}_i$ may be modelled by the probability $\Pr(\mathcal{C}_i)$ of this MU requesting $\mathcal{C}_i$ from the BS. Having a higher request probability $\Pr(\mathcal{C}_i)$ indicates that the MU is more interested in the content $\mathcal{C}_i$. The statistical analysis of the realistic video viewing behaviours exhibited by YouTube users revealed that a small fraction of popular contents attract the interest of a large fraction of users [23], [24]. Furthermore, the request probabilities of a set of ranked contents, say $\{\mathcal{C}_i | i = 1, \ldots, \mathcal{M}\}$, may be modelled by a Zipf distribution [25], [26]. Here $\mathcal{M}$ is the number of contents studied and the subscript i represents the particular position of $\mathcal{C}_i$ in the popularity list. A smaller integer subscript i indicates that the content is more popular and hence it is likely to be requested more frequently. Therefore, the probability of the piece of content $\mathcal{C}_i$ being requested is expressed as

$$\Pr(\mathcal{C}_i) = \frac{\frac{1}{i^\alpha}}{\sum_{j=1}^{\mathcal{M}} \frac{1}{j^\alpha}}, \tag{1}$$

where α is a predefined exponent. Having a higher value of α results in more intense interests in the top-ranked pieces of contents, as shown in TABLE I.

Assuming that we have $\mathcal{N}$ MUs within the area studied, these MUs independently request one piece of contents from the set $\{\mathcal{C}_i | i = 1, \ldots, \mathcal{M}\}$ with the corresponding probability defined in (1). The MUs requesting the same content $\mathcal{C}_i$ form a social group $\mathcal{G}_i$ in order to cooperatively disseminate the content of common interest across the social group. Hence, the size of the social group $\mathcal{G}_i$ requesting the same content $\mathcal{C}_i$ obeys a Binomial distribution, which is denoted as $|\mathcal{G}_i| \sim B[\mathcal{N}, \Pr(\mathcal{C}_i)]$. In order to exclude the case of $|\mathcal{G}| = 0$, we adjust the probability mass function (pmf)[7] of $|\mathcal{G}_i|$, which is expressed as

$$\Pr(|\mathcal{G}_i| = N) = \frac{\binom{\mathcal{N}}{N} \left[\Pr(\mathcal{C}_i)\right]^N \left[1 - \Pr(\mathcal{C}_i)\right]^{\mathcal{N}-N}}{1 - \left[1 - \Pr(\mathcal{C}_i)\right]^{\mathcal{N}}}. \tag{2}$$

where N is the specific size of the social group $\mathcal{G}_i$. As a result, the average $\overline{\mathcal{P}}(\mathcal{C}_i)$ of a specific delay metric associated with disseminating the content $\mathcal{C}_i$ across the social group $\mathcal{G}_i$, whose size is an adjusted-Binomially distributed random variable, can be expressed as

$$\overline{\mathcal{P}}(\mathcal{C}_i) = \sum_{N=1}^{\mathcal{N}} \mathcal{P}(|\mathcal{G}_i| = N) \cdot \Pr(|\mathcal{G}_i| = N), \tag{3}$$

[7]If no MUs requests the content, we do not have to study the content dissemination performance.

where $\mathcal{P}(|\mathcal{G}_i| = N)$ is a delay metric, which is a function of the deterministic social group size $|\mathcal{G}_i| = N$. Given the social group size N, in Section III, we will derive various delay metrics that can replace $\mathcal{P}(|\mathcal{G}_i| = N)$ in (3) in order to evaluate the impact of content popularity on the content dissemination performance.

To sum up, we assume that N MUs form a social group in order to request the content of common interest from a BS, as shown in Fig. 1. The formation of a social group depends on the following conditions:

- MUs share the same interest in a given piece of content;
- The content of common interest is of delay-sensitive nature;
- MUs roam in a bounded area having a relatively small size and they are geographically close to each other.

B. Network Layer

In order to disseminate the content of common interest across a social group, the BS creates a specific queue for buffering all the packets of the requested content and prepares for disseminating these packets one by one, as described below.

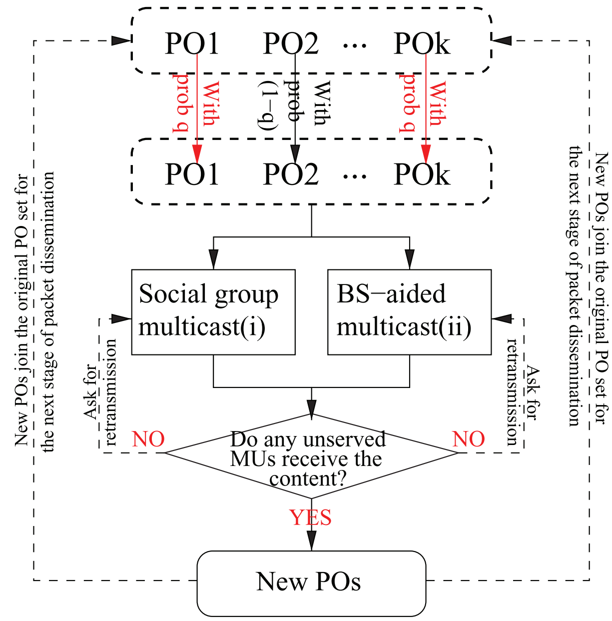
Firstly, the BSs are employed for repeatedly multicasting the packet currently at the head of the buffer, until at least one of the MUs in the target social group successfully receives it. Then, this packet is cooperatively disseminated across the social group using multicast techniques.

During the dissemination process, after successfully receiving the packet, the packet owners (POs) may make their decisions independently as to whether they would or would not forward the packet during the following stage of the dissemination, as shown in Fig. 2. Once some POs decided to further forward the packet, they would repeatedly multicast it until at least one unserved MU in the target social group successfully receives it. Afterwards, the new POs join the original PO set. Both the new POs and the original POs make new packet forwarding decisions again for the subsequent stage of dissemination. The probability of a PO willing to forward the packet is denoted as q $(0 \leq q \leq 1)$, which is termed as the Factor of Altruism (FA). At a given instant, there might not be any POs willing to further forward the packet. As a result, the unserved MUs in the target social group have to receive the packet directly from the BS. Similarly, the BS repeatedly multicasts the packet until at least one unserved MU in the target social group receives it.

During the content dissemination process, similar to the conventional BS-aided multicast, the BS keeps a specific packet at the head of the buffer, until all the MUs in the target social group successfully receive it. Then the packet is dropped from the buffer and the BS is ready to disseminate the subsequent one.

C. Physical (PHY) Layer

In the PHY layer, the radio propagation between any pair of transmitter and receiver is assumed to experience uncorrelated stationary Rayleigh flat-fading. Hence, the square of the fading amplitudes $|h_l(t)|^2$ during the t^{th} time slot (TS) obeys an *exponential distribution* having a unity mean, whose tail distribution



(i): Some POs are willing to further forward the content of common interest.
(ii): None POs are willing to further forward the content of common interest.

Fig. 2. Actions of POs during the spontaneous content dissemination.

function (tdf) is $\Pr[|h_l(t)|^2 > x] = \mathrm{e}^{-x}$. Given an arbitrary distance y_l in meters, the path loss (PL) Ω_l is expressed as [27]:

$$\Omega_l(y_l) = \begin{cases} 1, & y_l < d_0, \\ \left(\dfrac{4\pi f_c}{c}\right)^{\kappa} y_l^{\kappa}, & y_l \geq d_0, \end{cases} \tag{4}$$

where c is the speed of light and f_c is the carrier frequency, whereas κ is the PL exponent and d_0 is the distance from the transmitter to the 'near-field' edge.

The random distance Y_l is determined by the mobility pattern of the MUs in the target social group. The following mobility model is invoked for our performance analysis:

Definition 1 (Uniform mobility model): The position of the i^{th} MU during the t^{th} time interval is denoted by $\mathbf{P}_i(t)$, which obeys a stationary and ergodic process having a uniform distribution in the area considered. Moreover, the positions of different MUs are independently and identically distributed (i.i.d.).

This mobility model has been widely adopted for the performance analysis of MANETs [28], [29]. Let the probability density function (pdf) of the random distance Y_l between any two MUs be denoted by $f_{Y_l}(y_l)$. Our forthcoming performance analysis is applicable not only to the uniform mobility model, but to any arbitrary mobility model.

Note that, the index l in the formulas is a generic subscript, which represents 'b' when the BS is the transmitter, while it represents 's' when a MU is the transmitter. In the rest of the paper, 'l', 'b' and 's' hold the same meaning.

D. Medium-Access-Control (MAC) Layer

During a TS, a packet of the content is assumed to be successfully received by a MU, provided that the instantaneous received signal-to-noise-ratio (SNR) is higher than a pre-defined threshold γ [30]. In order to avoid collisions amongst multiple transmitters, orthogonal-frequency-division-multiple-access (OFDMA) or code-division-multiple-access (CDMA) may be invoked for allocating each transmitter an orthogonal channel. We denote the successful packet reception probability (SPRP) of a link as $\mu_l(y_l)$. By jointly considering the PHY layer model, the SPRP is derived as

$$\mu_l(y_l) = \Pr\left(\frac{\Pr_l^{tx}|h_l(t)|^2}{\Omega_l(y_l)N_0W_l} > \gamma\right) = \begin{cases} \mathrm{e}^{-\frac{\gamma N_0 W_l}{\Pr_l^{tx}}}, & y_l < d_0, \\ \mathrm{e}^{-\frac{\gamma N_0 W_l}{\Pr_l^{tx}}\left(\frac{4\pi f_c}{c}\right)^{\kappa} y_l^{\kappa}}, & y_l \geq d_0, \end{cases} \tag{5}$$

where P_l^{tx} is the corresponding transmit power and N_0W_l is the noise power in a communication bandwidth W_l. Given the pdf $f_{Y_l}(y_l)$ of the random distance Y_l, the average SPRP $\overline{\mu}_l$ of a link is derived as

$$\overline{\mu}_l = \int_0^{d_0} \mathrm{e}^{-\frac{\gamma N_0 W_l}{P_l^{tx}}} f_{Y_l}(y_l)dy_l + \int_{y_l \geq d_0} \mathrm{e}^{-\frac{\gamma N_0 W_l}{P_l^{tx}}\left(\frac{4\pi f_c}{c}\right)^{\kappa} y^{\kappa}} f_{Y_l}(y_l)dy_l. \tag{6}$$

Substituting the corresponding parameters and the pdf of the random distance into (6), we can obtain the average SPRP $\overline{\mu}_s$ between a pair of MUs and $\overline{\mu}_b$ between the BS and a MU. Moreover, the following lemma is proposed for our further analysis:

Lemma 1: Given the average SPRP $\overline{\mu}_l$ of a link during a TS, the average SPRP during a sufficiently short time interval Δt ($\Delta t \ll 1$ TS) is approximately $\overline{\mu}_l \Delta t$.

Proof: The proof can be found in Appendix A. ∎

Note that the SPRP also represents the *normalized throughput*, whose unit is packet/TS [30]. In more details, $\overline{\mu}_l$ indicates that $\overline{\mu}_l$ packets in average can be successfully received during a TS. Therefore, during Δt (≤ 1) TS, only $\overline{\mu}_l \Delta t$ packets in average can be successfully received.

III. Delay Analysis of the Packet Dissemination

In this section, various delay metrics of the packet dissemination process are derived with respect to a specific group size N. These metrics may replace the performance function $\mathcal{P}(|\mathcal{G}_i| = N)$ in (3) in order to characterize the average performance as a function of the content popularity.

A. Pure Birth Markov Chain (PBMC)

Let us assume that there are N MUs in a considered social group. During the process of packet dissemination across the target social group, the number of POs steadily increases until all the N social group members successfully receive the packet of common interest. Hence, the packet dissemination process can be modelled by a discrete-time PBMC having $(N+1)$ states, as shown in Fig. 3. In this PBMC, the states represent the corresponding numbers of POs having received the packet. State transition only occurs from a lower-indexed state to a higher-indexed one. Specifically, the state transition emerges from state 0, which represents the initial stage of the BS-aided multicast, and terminates in state N, which indicates that all the N MUs in the target social group have received the desired packet.

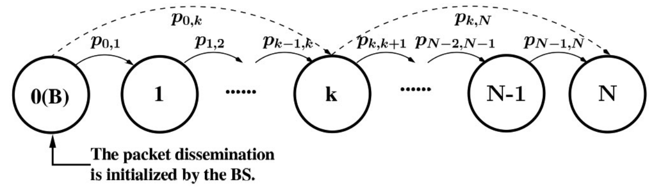


Fig. 3. A pure-birth Markov chain having an absorption state.

Let us first consider the general transition probability from state k to state $(k+m)$, where we have $1 \leq k \leq (N-1)$ and $0 \leq m \leq (N-k)$. In the light of the selfish user-behaviour considered, we assume that only n_k, $1 \leq n_k \leq k$, POs are willing to further disseminate the packet at the current stage. Therefore, any unserved MU out of the $(N-k)$ unserved ones is connected to the n_k POs by n_k wireless links, and any of these links has the probability of $\overline{\mu}_s \Delta t$ to successfully deliver the packet during the time interval Δt according to *Lemma 1*. As a result, given that n_k POs independently deliver their packets to the same target, the SPRP of an unserved MU is expressed as $[1-(1-\overline{\mu}_s\Delta t)^{n_k}]$. Furthermore, the state transition probability $p_{k,k+m|n_k \neq 0}$, which is also the probability of m out of the $(N-k)$ unserved MUs successfully receiving the packet during the current time interval Δt, can be expressed as

$$\begin{aligned} p_{k,k+m|n_k\neq 0} &= \binom{N-k}{m}\left[1-(1-\overline{\mu}_s\Delta t)^{n_k}\right]^m \\ &\quad .(1-\overline{\mu}_s\Delta t)^{n_k(N-k-m)} \\ &= \binom{N-k}{m}\left[1-\sum_{i=0}^{n_k}\binom{n_k}{i}(-\overline{\mu}_s\Delta t)^i\right]^m \\ &\quad .(1-\overline{\mu}_s\Delta t)^{n_k(N-k-m)} \\ &= \binom{N-k}{m}\left[\sum_{i=1}^{n_k}\binom{n_k}{i}(-1)^{i+1}(\overline{\mu}_s\Delta t)^i\right]^m \\ &\quad .(1-\overline{\mu}_s\Delta t)^{n_k(N-k-m)}. \end{aligned} \tag{7}$$

According to (7), the state transition probability $p_{k,k+m|n_k\neq 0}$ has the same growth rate as $\overline{\mu}_s^m \Delta t^m$. Hence, the adjacent-state transition probability $p_{k,k+1|n_k\neq 0}$ of traversing from state k to state $(k+1)$ has the same growth rate as $\overline{\mu}_s \Delta t$. Substituting $m=1$ into (7), $p_{k,k+1|n_k\neq 0}$ can be expressed as

$$\begin{aligned} p_{k,k+1|n_k\neq 0} &= (N-k)n_k\overline{\mu}_s\Delta t \\ &\quad + (N-k)\left[\sum_{i=2}^{n_k(N-k-1)}\binom{n_k(N-k-1)}{i}(-\overline{\mu}_s\Delta t)^i \right. \\ &\quad \left. - \sum_{j=2}^{n_k(N-k)}\binom{n_k(N-k)}{j}(-\overline{\mu}_s\Delta t)^j\right]. \end{aligned} \tag{8}$$

TABLE II
PARAMETERS OF THE PHY LAYER

	BS to MUs	MUs to MUs
Transmit Power	$P_b^{tx} = 31$ dBm	$P_s^{tx} = 0 \sim 10$ dBm
Carrier Freq	$f_{c,b} = 1.8$ GHz	$f_{c,s} = 2.4$ GHz
Bandwidth	$W_b = 10$ MHz	$W_s = 10$ MHz
Noise PSD	$N_0 = -174$ dBm/Hz ($20^o C$)	
SNR Threshold	$\gamma = 10$ dB	
PL Parameters	Exponent: $\kappa = 3$; Ref Distance: $d_0 = 1$ m	

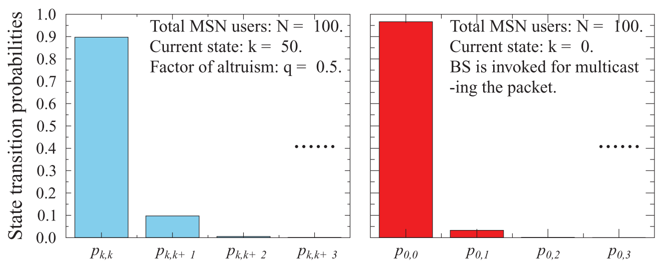


Fig. 4. State transition probabilities when $\Delta t = 0.001$ TS.

The terms in the square brackets of (8) have the same growth rate as $\overline{\mu}_s^2 \Delta t^2$. Compared to the first term $(N-k)n_k\overline{\mu}_s\Delta t$ in (8), the terms in the square brackets are negligibly low, when $\overline{\mu}_s\Delta t$ is close to zero. Hence, in this case, we can approximate $p_{k,k+1|n_k \neq 0}$ as $p_{k,k+1|n_k \neq 0} \approx (N-k)n_k\overline{\mu}_s\Delta t$. Similarly, when $\overline{\mu}_s\Delta t$ is close to zero, $p_{k,k+m|n_k \neq 0}$ associated with $m \geq 2$ in (7) can be approximated as $p_{k,k+m|n_k \neq 0} \approx 0$. Moreover, substituting $m = 0$ into (7), we obtain the probability of the PBMC sojourning in the current state k after the time interval Δt, which is $p_{k,k|n_k \neq 0} = (1 - \overline{\mu}_s\Delta t)^{n_k(N-k)}$. Again, when $\overline{\mu}_s\Delta t$ is very close to zero, $p_{k,k|n_k \neq 0}$ can be approximated as $p_{k,k|n_k \neq 0} \approx 1 - n_k(N-k)\overline{\mu}_s\Delta t$.

Another scenario is that no POs are willing to forward the packet, corresponding to the case $n_k = 0$. Then the $(N-k)$ unserved MUs have to receive the packet directly from the BS. Similarly, we can also demonstrate that $p_{k,k+1|n_k=0} \approx (N-k)\overline{\mu}_b\Delta t$ and $p_{k,k|n_k=0} \approx 1-(N-k)\overline{\mu}_b\Delta t$, while $p_{k,k+m|n_k=0} \approx 0$ for $m \geq 2$, provided that $\overline{\mu}_b\Delta t$ is sufficiently small. Furthermore, it can be shown that $p_{0,1} \approx N\overline{\mu}_b\Delta t$, $p_{0,0} \approx 1 - N\overline{\mu}_b\Delta t$ and $p_{0,m} \approx 0$ for $m \geq 2$, provided that $\overline{\mu}_b\Delta t$ is sufficiently small.

According to the PHY layer parameters in TABLE II, we plot the state transition probabilities for state $k = 50$ and for state $k = 0$, respectively, in Fig. 4. We observe from Fig. 4 that the state transition probabilities of $p_{k,k+m}$ and $p_{0,m}$ for $m \geq 2$ are negligibly low, which demonstrates the high accuracy of the above approximations involved.

Therefore, assuming a sufficiently short time interval Δt, only adjacent-state transitions occur during the process modelled by the discrete-time PBMC, as shown in Fig. 3.

B. Delay of State Transition

In order to study the delay statistics of disseminating a specific packet, we need to know the specific delay that the PBMC spends in a particular state, which is termed as the *state transition delay*. As a result, the following lemma may be formulated:

Lemma 2: Given the state transition probability $\widetilde{\mu}_k\Delta t$ from the current state k to state $(k+1)$, the transition delay from state k to state $(k+1)$ obeys the exponential distribution with a mean of $1/\widetilde{\mu}_k$ TS, provided that Δt is sufficiently small. Here, $\widetilde{\mu}_k$ is termed as the transition rate.

Proof: The proof can be found in Appendix B. ■

Based on *Lemma* 2, the discrete-time PBMC seen in Fig. 3 can be further simplified to a continuous-time PBMC, which only has adjacent-state transitions. The transition rate of this continuous-time PBMC can be shown to be $p_{k,k+1}/\Delta t$, where $p_{k,k+1}$ is the adjacent-state transition probability derived in Section III-A.

Let us first consider the delay T_k of the transition from state k to $(k+1)$, when $k \geq 1$. Since each PO has a probability q of forwarding the packet, in the current state k, the number n_k $(0 \leq n_k \leq k)$ of POs willing to forward the packet obeys a Binomial distribution having a pair of parameters k and q, whose pmf is given by [31]

$$p(n_k) = \binom{k}{n_k} q^{n_k}(1-q)^{k-n_k}, \quad n_k = 0, 1, \ldots, k. \tag{9}$$

For the case of $n_k \neq 0$, we have $p_{k,k+1|n_k \neq 0} \approx n_k(N-k)\overline{\mu}_s\Delta t$. According to *Lemma* 2, the delay T_k of the transition from state k to state $(k+1)$ obeys an exponential distribution having a rate of $n_k(N-k)\overline{\mu}_s = n_k\mu_{s,k}$, where $\mu_{s,k} = (N-k)\overline{\mu}_s$. Hence, when $n_k \neq 0$, the conditional pdf, the mean and the second moment of T_k may be formulated as

$$f_{T_k|n_k}(t_k) = n_k\mu_{s,k} \cdot \mathrm{e}^{-n_k\mu_{s,k}t_k}, \ t_k \geq 0 \tag{10}$$

$$\mathcal{E}\left[T_k|\, n_k\right] = \int_0^\infty t_k f_{T_k|n_k}(t_k)dt_k = \frac{1}{n_k\mu_{s,k}}, \tag{11}$$

$$\mathcal{E}\left[T_k^2\middle|\, n_k\right] = \int_0^\infty t_k^2 f_{T_k|n_k}(t_k)dt_k = \frac{2}{(n_k\mu_{s,k})^2}. \tag{12}$$

For the case of $n_k = 0$, we have $p_{k,k+1|n_k=0} \approx (N-k)\overline{\mu}_b\Delta t$, as the MUs in the target social group have to receive the packet from the BS. According to *Lemma* 2, the delay T_k of the transition from state k to $(k+1)$ obeys an exponential distribution having a rate of $\mu_{b,k} = (N-k)\overline{\mu}_b$. Hence, given $n_k = 0$, the conditional pdf, the mean and the second moment of T_k are derived as

$$f_{T_k|n_k=0}(t_k) = \mu_{b,k} \cdot \mathrm{e}^{-\mu_{b,k}t_k}, \ t_k \geq 0 \tag{13}$$

$$\mathcal{E}\left[T_k|\, n_k = 0\right] = \int_0^\infty t_k f_{T_k|n_k=0}(t_k)dt_k = \frac{1}{\mu_{b,k}}, \tag{14}$$

$$\mathcal{E}\left[T_k^2\middle|\, n_k = 0\right] = \int_0^\infty t_k^2 f_{T_k|n_k=0}(t_k)dt_k = \frac{2}{\mu_{b,k}^2}. \tag{15}$$

According to the classic Bayesian principle [31], the pdf of T_k may be expressed as

$$\begin{aligned} f_{T_k}(t_k) &= \sum_{n_k=1}^{k} f_{T_k|n_k}(t_k) \cdot p(n_k) + f_{T_k|n_k=0}(t_k) \cdot p(n_k = 0) \\ &= \sum_{n_k=1}^{k} \binom{k}{n_k} q^{n_k}(1-q)^{k-n_k} \cdot n_k \mu_{s,k} \mathrm{e}^{-n_k \mu_{s,k} t_k} \\ &\quad + (1-q)^k \mu_{b,k} \mathrm{e}^{-\mu_{b,k} t_k}. \end{aligned} \tag{16}$$

Moreover, the mean of T_k is formulated as

$$\begin{aligned} \mathcal{E}[T_k] &= \mathcal{E}[T_k | n_k = 0] \, p(n_k = 0) + \sum_{n_k=1}^{k} \mathcal{E}[T_k | n_k] \, p(n_k) \\ &= \underbrace{\frac{(1-q)^k}{\mu_{b,k}}}_{\mathcal{E}[T_{k,b}]} + \underbrace{\sum_{n_k=1}^{k} \binom{k}{n_k} \frac{q^{n_k}(1-q)^{k-n_k}}{n_k \mu_{s,k}}}_{\mathcal{E}[T_{k,s}]}, \end{aligned} \tag{17}$$

where $\mathcal{E}[T_{k,b}]$ represents the average duration of the BS-aided multicasting invoked during the transition from state k to state $(k+1)$, where $\mathcal{E}[T_{k,s}]$ is the average duration of the social group multicasting during this state transition. Furthermore, the second moment of T_k is formulated as

$$\begin{aligned} \mathcal{E}\left[T_k^2\right] &= \mathcal{E}\left[T_k^2 \middle| n_k = 0\right] p(n_k = 0) + \sum_{n_k=1}^{k} \mathcal{E}\left[T_k^2 \middle| n_k\right] p(n_k) \\ &= \frac{2(1-q)^k}{\mu_{b,k}^2} + \sum_{n_k=1}^{k} \binom{k}{n_k} \frac{2q^{n_k}(1-q)^{k-n_k}}{(n_k \mu_{s,k})^2}. \end{aligned} \tag{18}$$

From (17) and (18), we can also derive the variance of T_k by using the formula of $Var[T_k] = \mathcal{E}[T_k^2] - \{\mathcal{E}[T_k]\}^2$. Furthermore, we may simply derive the pdf, the mean and the second moment of the transition delay T_0 from state 0 to state 1 by substituting $k = 0$ in (13), (14), and (15), respectively.

C. Dissemination Delay

Since the delay of the transition from a state to its successor is independent of any other state transition's delay, and given that the dissemination delay across the target social group is defined as $T_D = \sum_{k=0}^{N-1} T_k$, the mean of T_D can be expressed as

$$\mathcal{E}[T_D] = \sum_{k=0}^{N-1} \frac{(1-q)^k}{\mu_{b,k}} + \sum_{k=1}^{N-1} \sum_{n_k=1}^{k} \binom{k}{n_k} \frac{q^{n_k}(1-q)^{k-n_k}}{n_k \mu_{s,k}}, \tag{19}$$

while the variance of T_D can be formulated as $Var[T_D] = \sum_{k=0}^{N-1} Var[T_k]$.

There is no exact closed-form tdf for the dissemination delay T_D in this general case. However, given its mean and variance, we may approximate it as a random variable obeying the *Gamma distribution*, which is usually more accurate than its Gaussian counterpart, when non-negative random variables are concerned [32]. According to the theory of the Gamma distribution [33], it is uniquely and unambiguously described by its *shape parameter* $m = \{E[T_D]\}^2 / Var[T_D]$ and *scale parameter* $\Theta = Var[T_D]/E[T_D]$. Then, given a delay threshold D_{th}, we may derive the approximate probability of the dissemination delay T_D exceeding D_{th} as

$$\Pr(T_D > D_{th}) \approx \frac{\Gamma\left(m, \dfrac{D_{th}}{\Theta}\right)}{\Gamma(m)} = \frac{\Gamma\left(\dfrac{\{E[T_D]\}^2}{Var[T_D]}, \dfrac{D_{th} E[T_D]}{Var[T_D]}\right)}{\Gamma\left(\dfrac{\{E[T_D]\}^2}{Var[T_D]}\right)}. \tag{20}$$

The accuracy of (20) will be verified by the Monte-Carlo simulation in Section V.

D. Individual User-Delay

A specific MU $\mathcal{A}$ in the target social group may receive the packet at any state spanning from 1 to N during the process of state transitions. When considering the transition from state $(k-1)$ to k $(1 \le k \le N)$, any of the $(N-k+1)$ unserved MUs may successfully receive the packet with a probability of $1/(N-k+1)$, and may not receive it with a probability of $(N-k)/(N-k+1)$. Specifically, the probability of $\mathcal{A}$ receiving the packet in state k, which naturally implies that $\mathcal{A}$ has not received the packets at any of the previous states, may be expressed as

$$p_k = \frac{1}{N-k+1} \cdot \prod_{i=1}^{k-1} \frac{N-i}{N-i+1} = \frac{1}{N}, \quad 1 \le k \le N. \tag{21}$$

Hence, given that $\mathcal{A}$ receives the packet in state k, the individual user-delay of $\mathcal{A}$ is expressed as $T_{\mathcal{A}|k} = \sum_{j=0}^{k-1} T_k$ and the conditional pdf of $T_{\mathcal{A}|k}$ is expressed as $f_{T_{\mathcal{A}|k}}(t_{\mathcal{A}}) = f_{T_0+\cdots+T_{k-1}}(t_{\mathcal{A}})$. According to the Bayesian principle [31], the pdf of the individual user-delay $T_{\mathcal{A}}$ can be expressed as:

$$f_{T_{\mathcal{A}}}(t_{\mathcal{A}}) = \sum_{k=1}^{N} f_{T_{\mathcal{A}|k}}(t_{\mathcal{A}}) \cdot p_k = \sum_{k=1}^{N} \frac{f_{T_0+\cdots+T_{k-1}}(t_{\mathcal{A}})}{N}. \tag{22}$$

Furthermore, owing to the fact that $\{T_0, T_1, \ldots, T_{k-1}\}$ are independent of each other, the average of $T_{\mathcal{A}}$ can be obtained as

$$\begin{aligned} \mathcal{E}[T_{\mathcal{A}}] &= \int_0^\infty t_{\mathcal{A}} \sum_{k=1}^{N} \frac{f_{T_0+\cdots+T_{k-1}}(t_{\mathcal{A}})}{N} dt_{\mathcal{A}} = \sum_{k=1}^{N} \frac{1}{N} \cdot \sum_{i=0}^{k-1} \mathcal{E}[T_i] \\ &= \sum_{k=1}^{N} \frac{N-k+1}{N} \mathcal{E}\left[T_{k-1}\right], \end{aligned} \tag{23}$$

where $\mathcal{E}\left[T_{k-1}\right]$ is given by (17). Furthermore, the second moment of $T_{\mathcal{A}}$ is given by

$$\begin{aligned} \mathcal{E}[T_{\mathcal{A}}^2] &= \int_0^\infty \sum_{k=1}^{N} \frac{t_{\mathcal{A}}^2 f_{T_0+\cdots+T_{k-1}}(t_{\mathcal{A}})}{N} dt_{\mathcal{A}} \\ &= \sum_{k=1}^{N} \frac{\mathcal{E}\left[(T_0 + T_1 + \cdots + T_{k-1})^2\right]}{N} = \sum_{k=1}^{N} \sum_{i,j=0}^{k-1} \frac{\mathcal{E}[T_i T_j]}{N} \\ &= \sum_{k=1}^{N} \frac{N-k+1}{N} \mathcal{E}[T_{k-1}^2] + \sum_{k=1}^{N} \frac{\boldsymbol{\xi}_k^T \left[\mathbf{H}_k - \mathbf{I}_k\right] \boldsymbol{\xi}_k}{N}, \end{aligned} \tag{24}$$

where $\boldsymbol{\xi}_k = (\mathcal{E}[T_0], \mathcal{E}[T_1], \ldots, \mathcal{E}[T_{k-1}])^T$, $\mathbf{H}_k$ is a $k \times k$ matrix, whose elements are all ones, and $\mathbf{I}_k$ is a $k \times k$ identity matrix. Consequently, the variance of $T_\mathcal{A}$ can be expressed as $Var(T_\mathcal{A}) = \mathcal{E}[T_\mathcal{A}^2] - \{\mathcal{E}[T_\mathcal{A}]\}^2$. Hence, by substituting $\mathcal{E}[T_\mathcal{A}]$ and $Var[T_\mathcal{A}]$ into (20), we may obtain the approximate probability of $T_\mathcal{A}$ exceeding threshold D_{th}.

IV. Delay Metrics for Special Cases

A. Case 1: Conventional BS-Aided Multicast (q = 0)

In this pessimistic case, all the MUs in the target social group are selfish during the packet dissemination process. Hence, the BS has to disseminate the packet to all the MUs in the target social group.

1) Dissemination Delay: When FA is $q = 0$, according to Eqs.(13)∼(15) in Section III-B, the state transition delays $\{T_k, k = 0, 1, \ldots, (N-1)\}$ are the independent exponentially distributed variables associated with the rates of $\{\widetilde{\mu}_k = (N-k)\overline{\mu}_b, k = 0, 1, \ldots, (N-1)\}$. Since the dissemination delay is defined as $T_D = \sum_{k=0}^{N-1} T_k$, T_D obeys the *hypoexponential distribution* [34]. Furthermore, since the rates of $\{T_k, k = 0, 1, \ldots, (N-1)\}$ are different from each other, the pdf of T_D can be expressed as

$$f_{T_D|q=0}(t_D) = \sum_{k=0}^{N-1} \prod_{j=0, j\neq k}^{N-1} \frac{N-j}{k-j}(N-k)\overline{\mu}_b \mathrm{e}^{-(N-k)\overline{\mu}_b t_D}. \tag{25}$$

In order to derive the probability of T_D exceeding a given threshold D_{th}, we integrate the above pdf $f_{T_D|q=0}(t_D)$ over the region $[D_{th}, \infty)$, which is expressed as

$$\begin{aligned}\Pr(T_D > D_{th}|q=0) &= \int_{D_{th}}^{\infty} f_{T_D|q=0}(t_D) dt_D \\ &= \sum_{k=0}^{N-1} \prod_{j=0, j\neq k}^{N-1} \frac{N-j}{k-j} \mathrm{e}^{-(N-k)\overline{\mu}_b D_{th}}.\end{aligned} \tag{26}$$

2) Individual User-Delay: When the FA is $q = 0$, the individual user-delay is solely determined by the quality of the wireless link connecting the MU $\mathcal{A}$ to the BS. As a result, according to *Lemma 2*, the individual user-delay $T_\mathcal{A}$ obeys an exponential distribution having a mean of $1/\overline{\mu}_b$. Furthermore, the probability of $T_\mathcal{A}$ exceeding a given threshold D_{th} is derived as $\Pr(T_\mathcal{A} > D_{th}|q=0) = \exp(-\overline{\mu}_b D_{th})$.

B. Case 2: Fully Altruistic Behaviours (q = 1)

In this optimistic scenario, all the MUs in the target social group are completely altruistic. Since there are always some POs willing to forward the packet during the dissemination process, the BS is not invoked for multicasting the packet any more, once some of the MUs have initially received it from the BS.

1) Dissemination Delay: When the FA is $q = 1$, by substituting $n_k = k$ into Eqs.(10)∼(12) in Section III-B, we know that the state transition delays $\{T_k, k = 1, \ldots, (N-1)\}$ are independent exponentially distributed variables associated with the rates of $\{\widetilde{\mu}_k = k(N-k)\overline{\mu}_s, k = 1, 2, \ldots, (N-1)\}$. Furthermore, by substituting $k = 0$ into Eqs.(13)∼(15) in Section III-B, the initial state transition delay T_0 is also an exponentially distributed variable associated with a rate of $\widetilde{\mu}_0 = N\overline{\mu}_b$. Note furthermore that T_0 is also independent of $\{T_k, k = 1, \ldots, (N-1)\}$. Since the dissemination delay is defined as $T_D = \sum_{k=0}^{N-1} T_k$, T_D obeys the hypoexponential distribution.

However, the rates of $\{\widetilde{\mu}_k = k(N-k)\overline{\mu}_s, k = 1, 2, \ldots, (N-1)\}$ associated with $\{T_k, k = 1, \ldots, (N-1)\}$ exhibit a symmetric structure. For example, the rates of T_k and T_{N-k} share the same value of $k(N-k)\overline{\mu}_s$. Hence, the closed-form equation for the tdf of T_D may only be expressed in the form of a *continuous phase-type distribution* [35]. As a result, when $q = 1$, the transition rate matrix of the PBMC is expressed as

$$\mathbf{P} = \begin{pmatrix} -\widetilde{\mu}_0 & \widetilde{\mu}_0 & 0 & \cdots & 0 & 0 \\ 0 & -\widetilde{\mu}_1 & \widetilde{\mu}_1 & \ddots & 0 & 0 \\ \vdots & \ddots & -\widetilde{\mu}_k & \widetilde{\mu}_k & \ddots & \vdots \\ 0 & 0 & \ddots & -\widetilde{\mu}_{N-2} & \widetilde{\mu}_{N-2} & 0 \\ 0 & 0 & \cdots & 0 & -\widetilde{\mu}_{N-1} & \widetilde{\mu}_{N-1} \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{Q} & \mathbf{Q}_0 \\ \mathbf{0} & 0 \end{pmatrix}, \tag{27}$$

where $\mathbf{Q}$ is a $(N \times N)$-element matrix containing all the transition rates between transient states, $\mathbf{Q}_0$ is a $(N \times 1)$ column vector containing all the transition rates from transient states to the absorbing state N, whose last entry is $\widetilde{\mu}_{N-1}$ and finally, the remaining entries are all zeros. As shown in Fig.3, the packet dissemination process starts from the initial state 0. Thus, the probability of T_D exceeding a given threshold D_{th} is expressed as

$$\Pr(T_D \geq D_{th}|q=1) = \boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q}) \times \mathbf{1}_N. \tag{28}$$

Note that in (28), the $(N \times 1)$ column vector $\boldsymbol{\tau}_{k+1}$ $(0 \leq k \leq N-1)$, whose $(k+1)$th entry is one but all the others are zeros, indicates that the PBMC starts at state k, while the $(N \times 1)$ column vector $\mathbf{1}_{k+1}$, whose first $(k+1)$ entries are ones and the remaining entries are zeros, indicates that the PBMC process is absorbed at state $(k+1)$. The proof of (28) can be found in [36].

2) Individual User-Delay: Given an event that the MU $\mathcal{A}$ successfully receives the packet at state $(k+1)$ $(0 \leq k \leq N-1)$, the PBMC used for modelling the packet dissemination in Fig.3 is considered to be terminated at state $(k+1)$. According to the physical meaning of both $\boldsymbol{\tau}_{k+1}$ and $\mathbf{1}_{k+1}$, similar to (28), the probability of $T_\mathcal{A}$ exceeding the threshold D_{th}, given that $\mathcal{A}$ receives the desired packet at state $(k+1)$ for $(0 \leq k \leq N-1)$, is expressed as

$$\Pr(T_\mathcal{A} \geq D_{th}|q=1, k+1) = \boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q}) \times \mathbf{1}_{k+1}. \tag{29}$$

Since we have already derived the probability of $p_{k+1} = 1/N$ that $\mathcal{A}$ receives the packet at state $(k+1)$ in (21), according to the Bayesian principle [31], the probability of $T_\mathcal{A}$ exceeding the threshold D_{th} is derived as

$$\begin{aligned}\Pr(T_{\mathcal{A}} \geq D_{th}|q=1) &= \sum_{k=0}^{N-1} \Pr(T_{\mathcal{A}} \geq D_{th}|q=1, k+1) \cdot p_{k+1} \\ &= \sum_{k=0}^{N-1} \frac{\boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q}) \times \mathbf{1}_{k+1}}{N} = \frac{\boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q})}{N} \times \sum_{k=0}^{N-1} \mathbf{1}_{k+1} \\ &= \frac{\boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q}) \times \boldsymbol{\eta}}{N}, \end{aligned} \tag{30}$$

where $\boldsymbol{\eta} = (N, N-1, \ldots, 1)^T$ is a $(N \times 1)$ column vector.

C. Case 3: Moderately Altruistic Behaviours (q = 0.5)

Unfortunately, we are unable to derive the exact tdf for the scenario, when the FA is set to $q = 0.5$. However, we are still able to offer some interesting insights concerning the delay metrics of this specific case. Substituting $q = 0.5$ into the second term of (17), the average duration of the social group multicast process during the transition from state k to $(k+1)$ for $k \geq 1$ can be given by

$$\mathcal{E}[T_{k,s}|q=0.5] = \frac{1}{2^k \cdot \mu_{s,k}} \sum_{n_k=1}^{k} \binom{k}{n_k} \frac{1}{n_k}. \tag{31}$$

According to Eq.(68.1) of [33], we arrive at the following lower bound for $\mathcal{E}[T_{k,s}|q=0.5]$, which is expressed as:

$$\begin{aligned}\mathcal{E}[T_{k,s}|q=0.5] &> \frac{1}{2^k \cdot \mu_{s,k}} \left[\sum_{n_k=0}^{k} \binom{k}{n_k} \frac{1}{n_k+1} - 1 \right] \\ &= \frac{1}{2^k \cdot \mu_{s,k}} \frac{2^{k+1}-k+2}{k+1}. \end{aligned} \tag{32}$$

Similarly, substituting $q = 1.0$ into the second term of (17), the corresponding formula of $\mathcal{E}[T_{k,s}|q=1.0]$ for this fully altruistic behaviour may be expressed as $\mathcal{E}[T_{k,s}|q=1.0] = 1/(k\mu_{s,k})$. As a result, the ratio $\mathcal{R}_{k,s}$ of these two expressions can be formulated as

$$\mathcal{R}_{k,s} = \frac{\mathcal{E}[T_{k,s}|q=0.5]}{\mathcal{E}[T_{k,s}|q=1.0]} > \frac{(2^{k+1}-k+2)k}{2^k(k+1)}. \tag{33}$$

In the ideal scenario, when k tends to infinity, this ratio can be expressed as $\lim_{k\to\infty} R_{k,s} > 2$. Since the lower bound derived in (32) is very tight[8], we can summarise that by assuming moderately altruistic behaviours, the average duration of the social group multicasting during the transition from state k to $(k+1)$ is twice that of the fully altruistic scenario, provided that k is sufficiently high.

Let us now demonstrate the tightness of the lower bound (32) in terms of the average dissemination delay. Substituting (32) into (19), the lower bound of the average dissemination delay $\mathcal{E}[T_D|q=0.5]$ can be formulated as

$$\begin{aligned}\mathcal{E}[T_D|q=0.5] &= \sum_{k=0}^{N-1} \frac{1}{2^k \mu_{b,k}} + \sum_{k=1}^{N-1} \frac{1}{2^k \mu_{s,k}} \sum_{n_k=1}^{k} \binom{k}{n_k} \frac{1}{n_k} \\ &> \sum_{k=0}^{N-1} \frac{1}{2^k \mu_{b,k}} + \sum_{k=1}^{N-1} \frac{1}{2^k \cdot \mu_{s,k}} \frac{2^{k+1}-k+2}{k+1}. \end{aligned} \tag{34}$$

When we compute the exact result of $\mathcal{E}[T_D|q=0.5]$, which is represented by the first line of (34), and its lower bound, which is quantified by the second line of (34), then for a large social group size N, such as $N = 50\sim200$, using a set of other related parameters in line with those of Fig. 6, the root-mean-square-deviation (RMSD) of these two sets of results can be shown to be 0.094 TS. Hence, we can claim that for a large social group size, which represents our densely populated scenario, the lower bound expressed in (34) can be regarded as an approximate result of $\mathcal{E}[T_D|q=0.5]$. Furthermore, the tightness of the lower bound derived in (33) can also be readily demonstrated.

Similarly, with the aid of (32), we can also obtain the lower bound for the average individual user-delay $\mathcal{E}(T_{\mathcal{A}}|q=0.5)$.

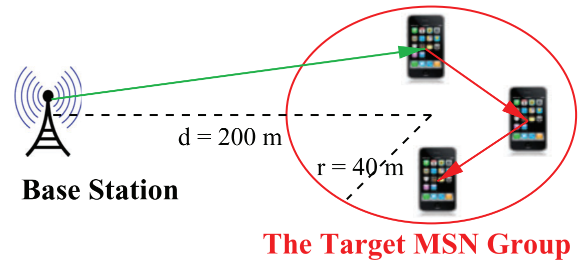


Fig. 5. Geographic features for obtaining numerical results.

V. NUMERICAL RESULTS

The parameters of the PHY layer are presented in TABLE II. The specific parameters used for transmissions from the BS to the MUs are in line with FDD-LTE standard[9], while the transmission parameters between the MUs are in line with the commonly used 802.11 protocol [18].

As shown in Fig. 5, we assume that all MUs in the target social group roam in a circular area having a radius of $r = 40$ m by obeying the *uniform mobility model*. The BS is $d = 200$ m away from the centre of the circular area. In this scenario, the pdf $f_{Y_s}(y_s)$ of the distance between a pair of MUs is given by Eq. (23) of [38], and $f_{Y_b}(y_b)$ between the BS and a MU can be found in our technical report [39]. Substituting $f_{Y_s}(y_s)$ and $f_{Y_b}(y_b)$ into (6), alongside the parameters offered in TABLE II, we may obtain the average SPRP $\overline{\mu}_s$ and $\overline{\mu}_b$, which further lead us to the analytical (ana) results for the various metrics. If we let $q = 0$ in our model, the corresponding analytical results are derived for conventional BS-aided multicast.

In order to obtain a reliable statistical characterization of the simulation performance (sim), we repeatedly run Monte-Carlo

[8] The tightness of this lower bound will be demonstrated in the following paragraph in terms of the average dissemination delay.

[9] We assume a 1.8 GHz carrier frequency in line with the LTE networks operated by the British company EE [37].

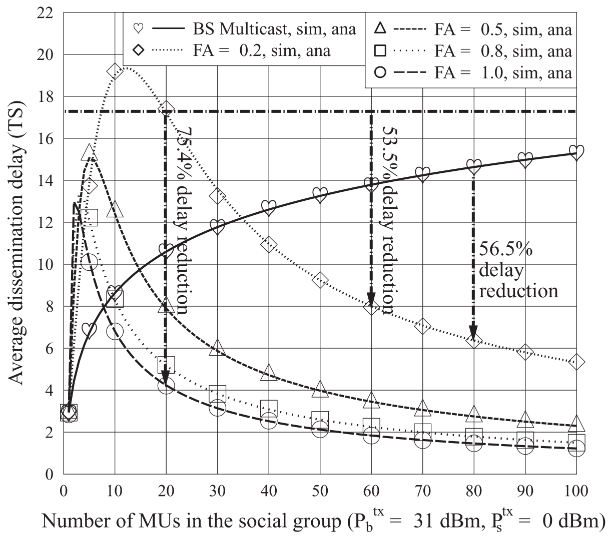


Fig. 6. Average dissemination delay affected by the number of MUs in the target social group, which is parameterized by the FA. The analytical results were evaluated from Eq. (19).

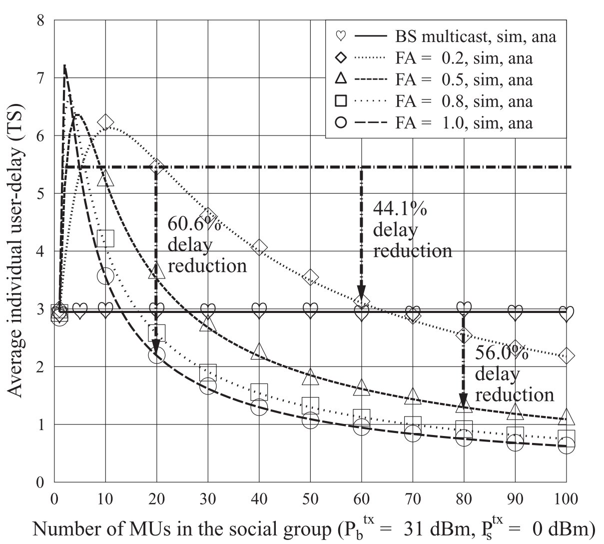


Fig. 7. Average individual user-delay as a function of the number of MUs in the target social group, which is parameterized by the FA. The analytical results were evaluated from (23).

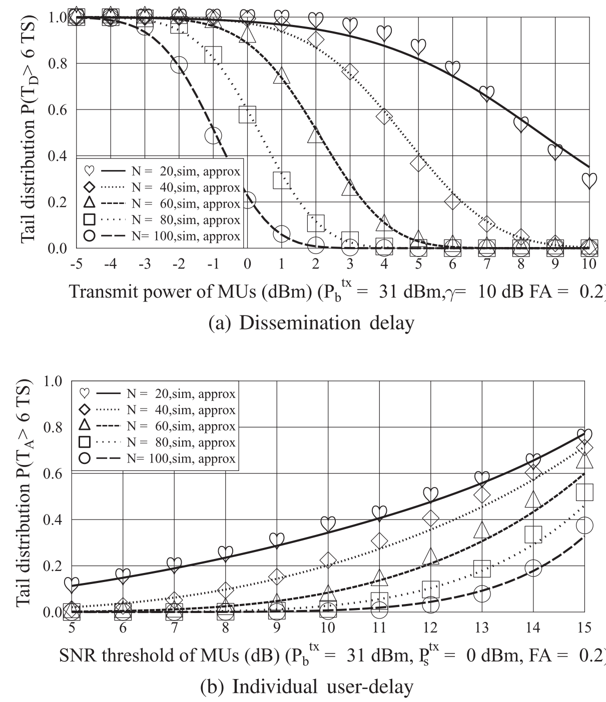


Fig. 8. The tail distribution of the delay versus (a) the transmit power and (b) the SNR threshold for successful reception, which is parameterized by the number of MUs in the target social group. The analytical results were either directly or indirectly derived from Eq.(20).

simulations 10 000 times and set the time-interval of our system to be $\Delta t = 0.001$ TS, where a TS can be considered as a packet duration. All the delay related metrics are evaluated by the number of TSs. In the numerical results of Figs. 6–8, we study the impact of the social group size N on the delay metrics of the packet dissemination process without considering any specific content popularity.

A. Delay Metrics for Uniform Mobility Model

As shown in Fig. 6, when FA $\neq 0$, the average dissemination delay firstly increases, as the number of MUs is increased. When only a few MUs are in the target social group, a longer period is required for disseminating the packet to all of the group members due to the increasing content demand of the unserved MUs. However, by further increasing the number of MUs, the diversity gain incurred by the cooperation of the multiple multicasters becomes sufficiently high to mitigate the adverse effect of the increasing content demand. As a result, we observe that the average dissemination delay decays after reaching its peak, as the number of MUs is further increased. For example, for FA $= 0.2$, the delay is reduced by 53.5%, as the number of MUs is increased from $N = 20$ to 60. Furthermore, a higher FA incurs a lower delay, since more POs are willing to forward the packet after they successfully receive it. For example, for $N = 20$, the average dissemination delay is reduced by 75.4%, as the FA is increased from 0.2 to 1. By contrast, when FA $= 0$, the conventional BS-aided multicast technique is invoked. However, as the number of the MUs increases, the average dissemination delay also increases. We observe from Fig. 6 that our approach is capable of reducing the average dissemination delay of the conventional BS-aided multicast by 56.5% for $N = 80$, when a small FA value of 0.2 is assumed.

As shown in Fig. 7, when only a few MUs are in the target social group and the FA is non-zero, due to the users' selfishness, fewer than two POs are willing to forward the packet during the dissemination process. Therefore, we observe from Fig. 7 that the average individual user-delay initially increases, because it does not benefit from any diversity gain. However, as we further increase the number of MUs, an increasing number of POs become willing to forward the packet, which substantially reduces the average individual user-delay, as observed from Fig. 7. For example, for FA $= 0.2$, the average individual user-delay is reduced by 44.1%, as the number of MUs is increased from $N = 20$ to 60. Nevertheless, when the conventional BS-aided multicast is invoked, the average individual user-delay, which only relies on the link connecting this

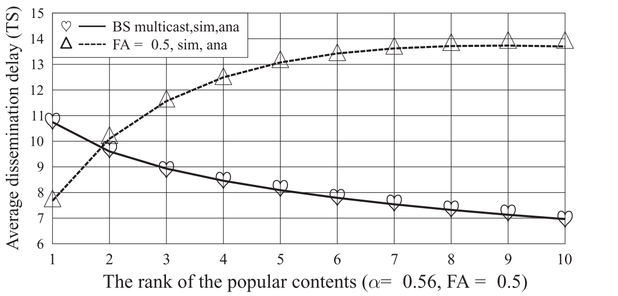


Fig. 9. Average dissemination delay as a function of the rank of the popular content. The transmit power of the BS is $P_b^{tx} = 31$ dBm and the transmit power of a MU is $P_t^{tx} = 0$ dBm. $\mathcal{N} = 100$ MUs independently request $\mathcal{M} = 10$ ranked-popularity pieces of contents according to the request probabilities listed in TABLE I when $\alpha = 0.56$. The analytical results were evaluated from Eq.(3).

specific MU to the BS, remains near-constant at 2.95 TS, as the number of MUs increases. Furthermore, the average individual user-delay is improved, when we increase the value of the FA. For example, given $N = 20$ MUs in the target social group, the average individual user-delay is reduced by 60.6%, as the FA is increased from 0.2 to 1.0. Additionally, given $N = 80$ MUs in the target social group, the average individual user-delay drops from 2.95 TS to 1.3 TS, comparing the conventional BS-aided multicast to our approach associated with FA = 0.5.

Observe in Fig. 8(a) that the probability of the dissemination delay exceeding a threshold of $D_{th} = 6$ TS reduces upon increasing the transmit power of each MU. By contrast, as portrayed in Fig. 8(b), the probability of the individual user-delay exceeding the same threshold increases upon increasing the SNR threshold to be exceeded for ensuring successful packet reception. Our Gamma-distribution-based approximations match the simulation results.

Then, we study the average dissemination delay as a function of the specific popularity of the pieces of contents in Fig. 9. Observe from Fig. 9 that as a piece of contents becomes less popular, the average dissemination delay of our scheme increases, when we have a moderate degree of altruism associated with FA = 0.5. When a piece of content is less popular, fewer MUs may request this content, hence the resultant smaller social group fails to provide sufficient cooperative multicast opportunities for rapidly disseminating the packet across the social group. By contrast, since a less popular piece of contents results in a lower content demand, the average dissemination delay of the BS-aided multicast reduces, as the content becomes less popular. Furthermore, as shown in Fig. 9, our scheme associated with FA = 0.5 outperforms the conventional BS-aided multicast in terms of its delay of disseminating the most popular content. Nevertheless, the BS-adied multicast is more suitable for disseminating the less popular pieces of contents.

B. Investigations Using Real Mobility Traces

Let us now study the content dissemination performance in a densely-populated subway station scenario [40]. The mobility traces for this scenario can be downloaded from the CRAWDAD database[10]. The active area in this scenario is

[10]http://crawdad.cs.dartmouth.edu/kth/walkers/

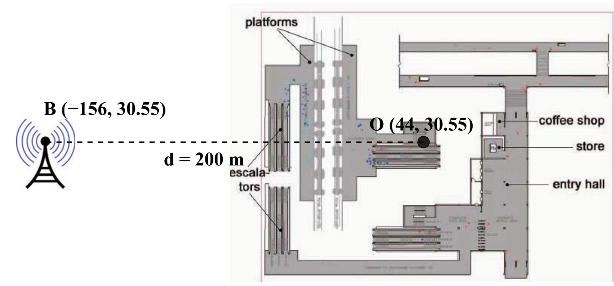


Fig. 10. A densely popluated subway station.

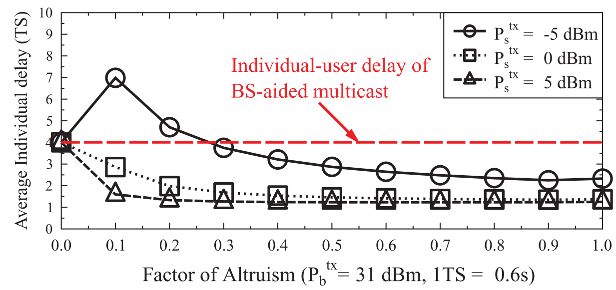


Fig. 11. Average individual user-delay in a subway station when all the MUs in the subway station form a grand social group for downloading a content of common interest.

1921 m^2. After analysing the mobility traces, the centre O of the active area is found to be at the coordinates of (44, 30.55) m, as shown in Fig. 10. In our simulations, we placed the BS at the point (−156, 30.55) m, which is 200 m away from the centre of the subway station. Since the MUs arrive/depart either through the entrances or during the arrival/departure of trains, the number of MUs is dynamic during the simulation time. As a result, we cannot readily obtain the dissemination delay in this scenario. However, we are still able to evaluate the individual user-delay, when our content dissemination scheme and conventional BS-aided multicast scheme are invoked. Again, the physical layer parameters are summarised in TABLE II. Since the positions of the MUs are captured every 0.6 s in this mobility trace, in our simulations we set the basic time interval of $\Delta t = 0.6$ s as a single TS, which can be considered as a packet's duration. Then the delay was evaluated in terms of the number of TSs.

We first assume that all the MUs in the subway station form a large social group in order to download the train schedule of common interest. Observe from Fig. 11 that for the cases of $P_s^{tx} = 0$ dBm and $P_s^{tx} = 5$ dBm, the average individual user-delay is reduced, as we increase the FA from 0.0 to 1.0. For $P_s^{tx} = -5$ dBm, when FA is increased from 0.0 to 0.1, we observe an increasing average individual user-delay. This is because the SPRP between the MUs is low and also, because fewer POs are willing to forward the packet. As FA becomes higher, more POs may join to assist the packet dissemination process, which significantly reduces the average individual user-delay. Specifically, when FA = 0, conventional BS-aided multicast is invoked for disseminating the packets. For $P_s^{tx} = 0$ or 5 dBm, if the MUs become only modestly altruistic, say we have FA = 0.1, our content dissemination scheme outperforms

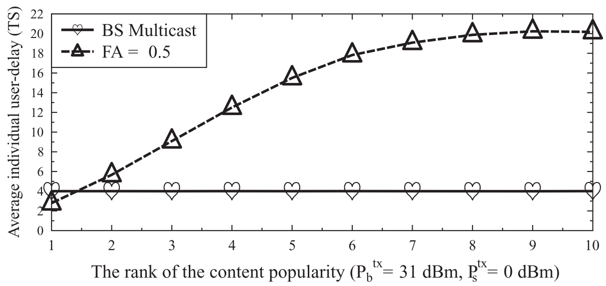


Fig. 12. Average individual user-delay in a subway station when the MUs in the subway station independently request $\mathcal{M} = 10$ ranked-popularity pieces of contents according to the probabilities listed in TABLE I when $\alpha = 1$.

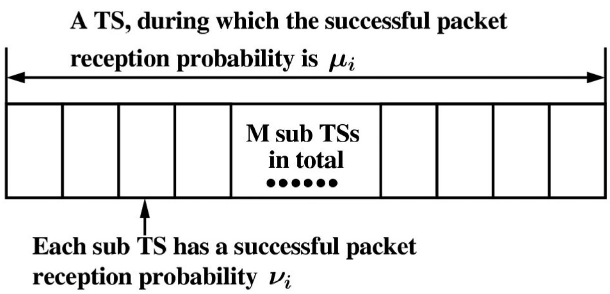


Fig. 13. The structure of a TS.

the conventional BS-aided multicast. For $P_s^{tx} = -5$ dBm, our scheme starts to outperform the classic BS-aided multicast, provided that FA is higher than 0.4.

We then study the impact of the specific content popularity on the average individual user-delay in a subway station. Observe from Fig. 12 that when disseminating the most popular content in the subway station, our dissemination scheme associated with FA $= 0.5$ outperforms the conventional BS-aided multicast. However, the BS-aided multicast is more suitable for disseminating less popular content in this scenario. The reason behind this trend is the same as that associated with Fig. 9.

VI. Conclusions

In this paper, we proposed a social group multicast aided content dissemination scheme as a supplement to the conventional cellular system. The content popularity is modelled by a Zipf distribution and the concept of FA was introduced for the sake of quantifying the probability of a PO forwarding a packet of the content of common interest. In our scheme, the BSs are invoked for multicasting the packet at the initial stage, as well as when no POs are willing to share the packet with others. By modelling the packet dissemination process as a PBMC, closed-form expressions were derived for the statistical properties of the various delay metrics. We demonstrated that our approach outperforms the conventional BS-adied multicast in terms of both the dissemination delay and the individual-user delay, especially when the density of MUs in a target group is high. Furthermore, we found that our approach is more suitable for disseminating a more popular content. By contrast, the conventional BS-aided multicast performs better for disseminating a less popular content.

Appendix A
The Proof of Lemma 1

As shown in Fig. 13, a TS is divided into M sub-TSs, each of which has a duration of $\Delta t = 1/M$ TS. We assume that the SPRP in a sub-TS is ν_i. As a result, given the SPRP $\overline{\mu}_i$ in a TS, we may derive the relation between $\overline{\mu}_i$ and ν_i, which is expressed as

$$\overline{\mu}_i = \sum_{j=1}^{M} (1-\nu_i)^{j-1}\nu_i = 1-(1-\nu_i)^M. \tag{35}$$

Rewriting the above expression, we obtain

$$\nu_i = 1-(1-\overline{\mu}_i)^{1/M} = 1-(1-\overline{\mu}_i)^{\Delta t}, \tag{36}$$

where the second equality is derived according to $\Delta t = 1/M$ TS. If we expand $(1-\overline{\mu}_i)^{\Delta t}$ according to the Taylor series, we have

$$(1-\overline{\mu}_i)^{\Delta t} = \sum_{n=0}^{\infty} \binom{\Delta t}{n} (-\overline{\mu}_i)^n = 1-\overline{\mu}_i\Delta t + O(\overline{\mu}_i^2), \tag{37}$$

where $O(\overline{\mu}_i^2)$ is the infinitesimal by small quantity on the same order as $\overline{\mu}_i^2$. Substituting the above equation into (36), we have

$$\nu_i = \overline{\mu}_i\Delta t + O(\overline{\mu}_i^2) \approx \overline{\mu}_i\Delta t. \tag{38}$$

According to our experiments, if we vary $\overline{\mu}_i$ from 0 to 0.8, the *root-mean-square-deviation* (RMSD) between the exact ν_i given by (36) and the approximated ν_i given by (38) is 9.45×10^{-4}. As a result, it is reasonable to claim that $\nu_i \approx \overline{\mu}_i\Delta t$.

Appendix B
The Proof of Lemma 2

During a time interval Δt, the PBMC may transit from state k to $(k+1)$ with a probability of $\widetilde{\mu}_k\Delta t$. Naturally, the successful state transition first occurring during the $(M_k = m_k)$-th Δt interval obeys a *geometric distribution*. According to the PMF of a geometric distribution having a parameter of $\widetilde{\mu}_k\Delta t$, we arrive at:

$$\Pr(M_k\Delta t \leq m_k\Delta t) = \sum_{m=1}^{m_k} (1-\widetilde{\mu}_k\Delta t)^{m-1}\widetilde{\mu}_k\Delta t, \tag{39}$$

$$\Pr(M_k\Delta t \leq (m_k+1)\Delta t) = \sum_{m=1}^{m_k+1} (1-\widetilde{\mu}_k\Delta t)^{m-1}\widetilde{\mu}_k\Delta t. \tag{40}$$

The continuous-valued delay of the adjacent-state transition is denoted as $T_k = M_k\Delta t$, which is associated with a specific value of $t_k = m_k\Delta t$. Hence, we may derive the pdf of T_k as:

$$\begin{aligned} f_{T_k}(t_k) &= \lim_{\Delta t\to 0} \frac{\Pr(T_k \leq t_k+\Delta t) - \Pr(T_k \leq t_k)}{\Delta t} \\ &= \lim_{\Delta t\to 0} \frac{\Pr(M_k\Delta t \leq (m_k+1)\Delta t) - \Pr(M_k\Delta t \leq m_k\Delta t)}{\Delta t} \\ &= \lim_{\Delta t\to 0} \frac{(1-\widetilde{\mu}_k\Delta t)^{m_k}\widetilde{\mu}_k\Delta t}{\Delta t} = \lim_{\Delta t\to 0} \widetilde{\mu}_k \mathrm{e}^{-m_k\widetilde{\mu}_k\Delta t} \\ &= \widetilde{\mu}_k \cdot \mathrm{e}^{-\widetilde{\mu}_k t_k}, \end{aligned} \tag{41}$$

where the last two lines are derived based on $\lim_{\Delta t\to 0}\widetilde{\mu}_k\Delta t = 1-\mathrm{e}^{-\widetilde{\mu}_k\Delta t}$ and $m_k = t_k/\Delta t$, respectively.

References

[1] N. Kayastha, D. Niyato, P. Wang, and E. Hossain, "Applications, architectures, and protocol design issues for mobile social networks: A survey," *Proc. IEEE*, vol. 99, no. 12, pp. 2130–2158, Dec. 2011.

[2] A. Vahdat and D. Becker, "Epidemic routing for partially-connected ad hoc networks," Master thesis, Dept. Comput. Sci., Duke Univ., Durham, NC 27708 USA, Tech. Rep., 2000.

[3] Y.-K. Ip, W.-C. Lau, and O.-C. Yue, "Performance modeling of epidemic routing with heterogeneous node types," in *Proc. IEEE Int. Conf. Commun. (ICCŠ08)*, May 2008, pp. 219–224.

[4] A. Clementi, F. Pasquale, and R. Silvestri, "Opportunistic manets: Mobility can make up for low transmission power," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 610–620, Feb. 2013.

[5] H. Sun and C. Wu, "Epidemic forwarding in mobile social networks," in *Proc. IEEE Int. Conf. Commun. (ICC'12)*, Jun. 2012, pp. 1421–1425.

[6] J. Whitbeck, V. Conan, and M. D. de Amorim, "Performance of opportunistic epidemic routing on edge-markovian dynamic graphs," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1259–1263, May 2011.

[7] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. IEEE INFOCOM*, 2009, pp. 1422–1430.

[8] C. Boldrini, M. Conti, and A. Passarella, "Contentplace: Social-aware data dissemination in opportunistic networks," in *Proc. 11th Int. Symp. Model. Anal. Simul. Wireless Mobile Syst. (MSWiM'08)*, 2008, pp. 203–210.

[9] D. Niyato, P. Wang, W. Saad, and A. Hjϕandrungnes, "Controlled coalitional games for cooperative mobile social networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1812–1824, May 2011.

[10] K. Akkarajitsakul, E. Hossain, and D. Niyato, "Cooperative packet delivery in hybrid wireless mobile networks: A coalitional game approach," *IEEE Trans. Mobile Comput.*, vol. 12, no. 5, pp. 1–15, May 2013.

[11] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Trans. Mobile Comput.*, vol. 11, no. 5, pp. 821–834, May 2012.

[12] Y. Li, M. Qian, D. Jin, P. Hui, Z. Wang, and S. Chen, "Multiple mobile data offloading through disruption tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 7, pp. 1579–1596, Jul. 2014.

[13] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Perform. Eval.*, vol. 62, nos. 1–4, pp. 210–228, Oct. 2005.

[14] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, "Power law and exponential decay of intercontact times between mobile devices," *IEEE Trans. Mobile Comput.*, vol. 9, no. 10, pp. 1377–1390, Oct. 2010.

[15] J. Hu, L.-L. Yang, and L. Hanzo, "Mobile social networking aided content dissemination in heterogeneous networks," *China Commun.*, vol. 10, no. 6, p. 1, 2013.

[16] J. Wang, S. Park, D. Love, and M. Zoltowski, "Throughput delay tradeoff for wireless multicast using hybrid-ARQ protocols," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2741–2751, Sep. 2010.

[17] J. Seo, T. Kwon, and V. Leung, "Social groupcasting algorithm for wireless cellular multicast services," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 47–50, Jan. 2013.

[18] Information Technology–Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks–Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, ISO/IEC/IEEE Standard 8802–11:2012(E), Nov. 2012, pp. 1–2798.

[19] Z. Gong and M. Haenggi, "Interference and outage in mobile random networks: Expectation, distribution, and correlation," *IEEE Trans. Mobile Comput.*, vol. 13, no. 2, pp. 337–349, Feb. 2014.

[20] H. Kwon and B. G. Lee, "Cooperative power allocation for broadcast/multicast services in cellular OFDM systems," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3092–3102, Oct. 2009.

[21] D. Feng, L. Lu, Y. Yuan-Wu, G. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3541–3551, Aug. 2013.

[22] R. E. Hattachi and J. Erfanian, "NGMN 5G White Paper version 1.0," *Next Generation Mobile Networks (NGMN)*, NGMN Alliance, Frankfurt, Germany, 2015.

[23] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas. (IMC'07)*, Oct. 2007, p. 1.

[24] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Characteristics of YouTube network traffic at a campus network—Measurements, models, and implications," *Comput. Netw.*, vol. 53, no. 4, pp. 501–514, Mar. 2009.

[25] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[26] M. Taghizadeh, K. Micinski, S. Biswas, C. Ofria, and E. Torng, "Distributed cooperative caching in social wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 6, pp. 1037–1053, Jun. 2013.

[27] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.

[28] H. Zhang, Z. Zhang, and H. Dai, "Gossip-based information spreading in mobile networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5918–5928, Nov. 2013.

[29] X. Wang, Q. Peng, and Y. Li, "Cooperation achieves optimal multicast capacity-delay scaling in MANET," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3023–3031, Oct. 2012.

[30] J. Hu, L.-L. Yang, and L. Hanzo, "Maximum average service rate and optimal queue scheduling of delay-constrained hybrid cognitive radio in Nakagami fading channels," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2220–2229, Jun. 2013.

[31] P. V. Mieghem, *Performance Analysis of Communications Networks and Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[32] M. Derakhshani and T. Le-Ngoc, "Aggregate interference and capacity-outage analysis in a cognitive radio network," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 196–207, Jan. 2012.

[33] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. New York, NY, USA: Elsevier/Academic, 2007.

[34] W. Gao, Q. Li, B. Zhao, and G. Cao, "Social-aware multicast in disruption-tolerant networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1553–1566, Oct. 2012.

[35] M. W. Fackrell, "Characterization of matrix-exponential distributions," Ph.D. dissertation, Faculty of Eng., Comput. Math. Sci., School of Appl. Math., Univ. Adelaide, Adelaide, South Australia, 5005, Australia, 2003.

[36] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Philadelphia, PA, USA: SIAM, 1999.

[37] Office of Communications. "Notice of proposed variation of Everything Everywhere's 1800 MHz spectrum licences to allow use of LTE and WiMAX technologies," Ofcom, Mar. 2012 [Online]. Available: http://stakeholders.ofcom.org.uk/consultations/variation-1800mhz-lte-wimax/summary.

[38] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, "Stochastic properties of the random waypoint mobility model," *Wireless Netw.*, vol. 10, no. 5, pp. 555–567, 2004.

[39] J. Hu, L.-L. Yang, and L. Hanzo, "Stochastic geometry in the cellular networks," Univ. Southampton, SO17 1BJ, U.K., Tech. Rep., 2015 [Online]. Available: http://eprints.soton.ac.uk/id/eprint/374932.

[40] O. Helgason, S. T. Kouyoumdjieva, and G. Karlsson, "Opportunistic communication and human mobility," *IEEE Trans. Mobile Comput.*, vol. 13, no. 7, pp. 1597–1610, Jul. 2014.

Jie Hu (S'11–M'16) received the B.Eng. degree in communication engineering and the M.Eng. degree in communication and information system from the School of Communication and Information Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2011, respectively, and the Ph.D. degree from the Southampton Wireless Group, University of Southampton, Southampton, U.K., in 2015.

Since March 2016, he has been working with the School of Communication and Information Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, as a Lecturer. He has a broad range of interests in wireless communication and networking, such as cognitive radio and cognitive networks, queuing analysis, resource allocation and scheduling, ad hoc wireless networks, and mobile social networks.

Lie-Liang Yang (M'98–SM'02–F'16) received the B.Eng. degree in communications engineering from Shanghai TieDao University, Shanghai, China, in 1988, and the M.Eng. and Ph.D. degrees in communications and electronics from Beijing (Northern) Jiaotong University, Beijing, China, in 1991 and 1997, respectively. From June 1997 to December 1997, he was a Visiting Scientist at the Institute of Radio Engineering and Electronics, Academy of Sciences of the Czech Republic, Prague, Czech Republic. Since December 1997, he has been with the University of Southampton, Southampton, U.K., where he is currently a Professor of Wireless Communications with the School of Electronics and Computer Science. His research interests include wireless communications, wireless networking, and signal processing for wireless communications. He has authored more than 300 research papers in journals and conference proceedings, authored or coauthored three books, and also published several book chapters.

Prof. Yang is a Fellow of The Institution of Engineering and Technology, U.K. He served as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the *Journal of Communications and Networks*, and is currently an Associate Editor of the IEEE ACCESS and the *Security and Communication Networks Journal*.

Lajos Hanzo (M'91–SM'92–F'03) received the degree in electronics, in 1976, the Doctorate degree, in 1983, and the D.Sc. degree. In 2009, he received the Honorary Doctorate "Doctor Honoris Causa" from the Technical University of Budapest, Budapest, Hungary. During his 35-year career in telecommunications, he has held various research and academic posts in Hungary, Germany, and the U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K., where he holds the Chair in telecommunications. He has successfully supervised 80 Ph.D. students, coauthored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10,000 pages, published over 1500 research entries at IEEE Xplore, acted both as TPC and General Chair of the IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently, he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) U.K., the European IST Programme and the Mobile Virtual Centre of Excellence (VCE), U.K. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. From 2008 to 2012, he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing, China. His research is funded by the European Research Council's Senior Research Fellow Grant. He is the Fellow of the REng, IET, and EURASIP.

Delay Analysis of Social Group Multicast-Aided Content Dissemination in Cellular System

Jie Hu, *Member, IEEE*, Lie-Liang Yang, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Based on the common interest of mobile users (MUs) in a social group, the dissemination of content across the social group is studied as a powerful supplement to conventional cellular communication with the goal of improving the delay performance of the content dissemination process. The content popularity is modeled by a Zipf distribution to characterize the MUs' different interests in different contents. The factor of altruism (FA) terminology is introduced for quantifying the willingness of content owners to share their content. We model the dissemination process of a specific packet by a pure-birth-based Markov chain and evaluate the statistical properties of both the network's dissemination delay as well as of the individual user-delay. Compared to the conventional base station (BS)-aided multicast, our scheme is capable of reducing the average dissemination delay by about 56.5%. Moreover, in contrast to the BS-aided multicast, increasing the number of MUs in the target social group is capable of reducing the average individual user-delay by 44.1% relying on our scheme. Furthermore, our scheme is more suitable for disseminating a popular piece of content.**

***Index Terms*—Content dissemination, content popularity, factor of altruism, pure-birth based Markov chain, delay analysis.**

I. Introduction

A. Background and Related Works

As a combination of social science and mobile networks, mobile social networks (MSNs) [1] are attracting an increasing attention across the research community. In the context of MSNs, mobile users (MUs) may form a social group in order to cooperatively disseminate the content of common interest. There are substantial contributions to the performance analysis of epidemic forwarding [2] in mobile ad hoc networks (MANETs). In the context of MANETs, a two-dimensional *continuous time Markov chain* (CTMC) was proposed in [3] for evaluating the performance of a heterogeneous MANETs. To a further advance, the authors of [4] derived a tight upper bound of the flooding time, which is defined as the number of time-steps required for broadcasting a message from a source node to all nodes. Furthermore, in [5] the end-to-end message delivery delay using an epidemic forwarding protocol was investigated theoretically in a composite twin-layer network, which includes a physical MANET and a virtual social network.

However, epidemic forwarding [6] is often criticised as being an end-to-end routing protocol, because it consumes substantial resources of the intermediate nodes, which might not be interested in the information to be relayed. However, if MUs can form a social group and request the content of common interest together, epidemic forwarding becomes an efficient way of cooperatively disseminating the content in the target social group[1]. Content dissemination in purely distributed opportunistic networks was investigated in [7] and [8]. Epidemic forwarding aided content dissemination was invoked in [7], where the users share any content updates with others that they meet in order to improve the coverage quality and to increase the capacity. A socially-aware content placement algorithm was proposed in [8] for enhancing the opportunity of MUs to gain access to their contents of interest.

Some researches focused on a hybrid content dissemination approach. In [9] and [10], the authors investigated how the content providers and network operators can interact for the sake of efficiently distributing the contents with the aid of a coalition game. At the time of writing, epidemic forwarding aided content dissemination is widely studied for the sake of offloading tele-traffic from cellular networks. In [11], the authors proposed a framework for initial content-receiver selection in order to disseminate the content of common interest to as many subscribers as possible before interest in the content subsides. In [12], where MUs were categorised into "helpers" and "subscribers", several algorithms were designed for solving the optimisation problem of offloading multiple types of contents from the cellular networks.

The above-mentioned contributions [2]–[12] focused their attention on user-encounter-based MANETs or 'large-scale MSNs', where the mobile nodes are sparsely distributed across a large area. Typically a rudimentary physical layer model is assumed, namely that if a pair of nodes enter each other's transmission range, the packet can be successfully delivered from the source to the target. Hence, the delivery delay is dominated by the *inter-contact duration*[2] of mobile nodes [15], rather than by the wireless signal propagation. Due to

Manuscript received April 12, 2015; revised October 18, 2015 and February 10, 2016; accepted February 22, 2016. This work was supported by the RC-UK's India-UK Advanced Technology Centre (IU-ATC), that of the EU's concerto project and of the China Scholarship Council (CSC). The work of L. Hanzo was supported by the European Research Council under its Advanced Fellow Award. The associate editor coordinating the review of this paper and approving it for publication was B. Liang.

J. Hu is with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: hujienavy@outlook.com).

L.-L. Yang and L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lly@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2016.2535330

[1] Other MUs that do not belong to the target social group are not relied upon for assisting the content dissemination process.

[2] In these treatises, the inter-contact duration of MUs is commonly assumed to obey an exponential distribution, which is demonstrated in [13] with the aid of artificial or synthetic mobility models and in [14] by realistic measured mobility traces.

the underlying long inter-contact duration of the MUs, this user-encounter-based content dissemination is only capable of delivering delay-tolerant services in a large-scale area. As a result, the contributions of [2]–[12] belong to the category of *delay-tolerant networks* (DTNs). However, typically idealised simplifying assumptions are used in the literature of the DTN paradigm:

- The commonly assumed simplified physical layer model ignores the impact of transmit power, of the path-loss and of the multipath fading, etc.
- The cooperative user-encounter based content dissemination in DTNs is not suitable for delivering delay-sensitive services.

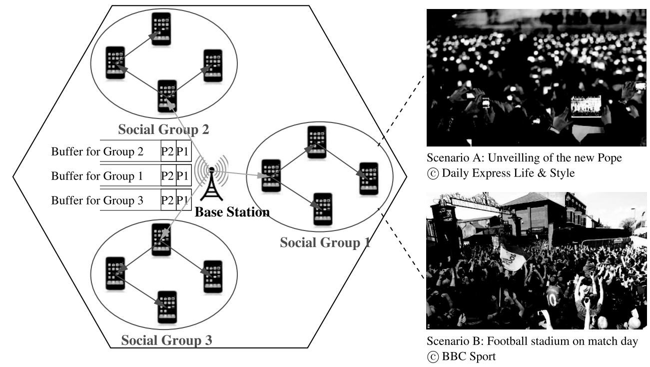


Fig. 1. Social group multicast aided content dissemination in cellular systems.

B. Motivations and Contributions

The conventional method of disseminating the delay-sensitive content of common interest relies on BS-aided multicast, where the BS is the sole transmitter. Since the BS-aided multicast has to guarantee the quality of service (QoS) at every content requester, the capacity of multicast channels is predetermined by the worst channel amongst those connecting the BS to the content requesters. In this case, due to the time-variant nature of wireless channels, when the BS multicasts a packet, some MUs may receive it earlier than their less fortunate counterparts. Then, the successful receivers have to remain silent, because the BS would not multicast the second packet, before all the MUs successfully receive the current one.

In high-user-density scenarios, the MUs often share common interest in delay-sensitive content. For instance, the crowd participating in the inauguration of the new Pope share common interest in close-up video-clips of the Pope on the podium. Similarly, supporters in a football stadium share common interest in video-clips of a spectacular goal from different angles or in the score updates from another stadium, as exemplified by Fig. 1. However, the conventional BS-aided multicast is an inefficient technique of disseminating the delay-sensitive content of common interest in these typical densely populated scenarios. The reason for this is two-fold:

- As the content requesters' density increases, the worst channel amongst those connecting the BS and the content requesters becomes even worse, which results in excessive dissemination delay [16].
- Since the dissemination delay is increased, the BS is engaged in multicasting for a longer period, which further delays all other services.

If local MUs form a social group for requesting the content of common interest from the BS together, local communications amongst MUs can be exploited for cooperatively multicasting the packets from the packet owners to the hitherto unserved MUs in the target social group[3]. The potential performance gain of this social group multicast aided content dissemination over the conventional BS-aided multicast arises from the following two benefits:

- Relying on the cooperative multicast of the multiple packet owners results in rich cooperative diversity gains, which in turn improves the packet delivery performance.
- Activating direct transmissions amongst the MUs is capable of reducing the distance between a transmitter and receiver pair, which in turn reduces the path-loss-induced channel attenuation between them.

Furthermore, since we offload the content dissemination task from the BS-aided multicast to the local communications amongst the social group members, the BS becomes capable of satisfying other communication demands, which consequently improves the efficiency of the BS's exploitation.

The size of the area covered by a social group should be carefully designed for different scenarios. If the area is as large as a macro-cell, cooperative user-encounter based communication amongst MUs is only suitable for disseminating delay-tolerant information, as we argued at the end of Section I-A. The best option for disseminating delay-sensitive information across a large area is that of classic BS-aided multicast. By contrast, if the area is relatively small, such as a circular area with a radius shorter than a hundred meters, which is comparable to the default transmission range of a MU[4], communication efficiency between a transmitter and receiver pair is dominated by the wireless signal propagation properties, rather than by their inter-contact duration. Hence, social group aided cooperative multicast is capable of significantly reducing the delay of the conventional BS-aided multicast, as we emphasized at the beginning of Section I-B. This scenario is termed as a "small-scale MSN" [15], where the channel attenuation factors dominate the associated delay characteristics [19]. Against this background, our novel contributions are as follows:

- A hybrid content dissemination approach is proposed, which relies both on BS-aided multicast [20] and on social group multicast aided content dissemination. This process is modelled by a *pure-birth based Markov chain (PBMC)*. Various factors that might affect the performance of the content dissemination are accounted for, including the path-loss-induced channel attenuation, the multipath fading and the users' altruistic versus selfish behaviours, which distinguishes our work from the existing literature of DTNs.

[3]A similar methodology of improving BS-aided multicast was also advocated in [17], which was mainly focused on the selection of the initial receivers. However, the authors of [17] have not analysed the content dissemination stage.

[4]New Wi-Fi protocols, such as 802.11n/ac [18], are capable of supporting a transmission range of hundreds of meters.

- We model the popularity of different pieces of contents by a Zipf distribution, which affects the specific formation of a social group and hence influences the dissemination process of the content of common interest across the target social group.
- Considering a specific packet of the content of common interest, we analyse the statistical properties of the dissemination delay, which is the time from the BS's instant of multicasting a packet until all the MUs in the target social group receive this packet. We also analyse the individual user-delay, which is the time spanning from the BS multicasting a packet until a specific MU receives this packet.
- The advantages of our social group multicast aided content dissemination scheme over the conventional BS-aided multicast are demonstrated by the mobility traces extracted from a realistic subway station scenario.

Note that improving the network infrastructure in high-user-density areas can certainly enhance the general communication experience of MUs, when supporting phone calls, texts, emails and basic data services. However, it may constitute an inefficient technique of disseminating the content of common interest. It may also be an unwise investment for the network operators, since people often temporarily get together for attending social events. Hence, improving the infrastructure capacity may be wasteful. By contrast, our social group multicast scheme constitutes a more economical and flexible solution for disseminating the content of common interest amongst the social group members, which is based on direct communications between the social group members. We will demonstrate that our social group multicast aided scheme outperforms the BS-aided multicast in terms of disseminating the popular content of common interest in high-user-density areas.

The rest of the paper is organized as follows. In Section II, our system model is introduced. In Section III, we analyse the delay metrics. Furthermore, the exact closed-form formulas are derived for two special cases in Section IV. Our numerical results are provided in Section V. Finally, we conclude in Section VI.

II. System Overview

Similar to the BS-controlled device-to-device communication services of the LTE network [21], our system operates by obeying a centralised-control regime combined with a decentralised-transmission paradigm[5], where the BS acts as a centralised controller in order to support the functions of synchronisation[6], of social group formation as well as of coordination and resource allocation for multiple content owners etc. By contrast, the information transmission is carried out by direct communications between a transmitter and receiver pair.

[5]This paradigm has been considered as a part of the forthcoming '5G' regime, known as the 'LTE-Assisted Wi-Fi Direct' technique [22], where the control signalling exchange is carried out by the LTE-based BS, while the information transmission is realised by the Wi-Fi-based direct communication between a transmitter and receiver pair.

[6]Since the MUs in the cellular system rely on regular control signalling exchange with associated BSs, they can readily synchronise with associated BSs and hence also with each other.

TABLE I
THE REQUEST PROBABILITIES OF $\mathcal{M} = 10$ RANKED POPULAR CONTENTS FOR BOTH $\alpha = 0.56$ [24] AND $\alpha = 1.0$ [23]

Content	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
$\alpha = 0.56$	21.4%	14.6%	11.6%	9.9%	8.7%
$\alpha = 1.0$	34.1%	17.1%	11.4%	8.5%	6.8%
Content	$\mathcal{C}_6$	$\mathcal{C}_7$	$\mathcal{C}_8$	$\mathcal{C}_9$	$\mathcal{C}_10$
$\alpha = 0.56$	7.9%	7.2%	6.7%	6.3%	6.0%
$\alpha = 1.0$	5.7%	4.9%	4.3%	3.8%	3.4%

A. Content Popularity and Social Group Formation

The interest of a MU in a specific piece of content $\mathcal{C}_i$ may be modelled by the probability $\Pr(\mathcal{C}_i)$ of this MU requesting $\mathcal{C}_i$ from the BS. Having a higher request probability $\Pr(\mathcal{C}_i)$ indicates that the MU is more interested in the content $\mathcal{C}_i$. The statistical analysis of the realistic video viewing behaviours exhibited by YouTube users revealed that a small fraction of popular contents attract the interest of a large fraction of users [23], [24]. Furthermore, the request probabilities of a set of ranked contents, say $\{\mathcal{C}_i | i = 1, \ldots, \mathcal{M}\}$, may be modelled by a Zipf distribution [25], [26]. Here $\mathcal{M}$ is the number of contents studied and the subscript i represents the particular position of $\mathcal{C}_i$ in the popularity list. A smaller integer subscript i indicates that the content is more popular and hence it is likely to be requested more frequently. Therefore, the probability of the piece of content $\mathcal{C}_i$ being requested is expressed as

$$\Pr(\mathcal{C}_i) = \frac{\frac{1}{i^\alpha}}{\sum_{j=1}^{\mathcal{M}} \frac{1}{j^\alpha}}, \tag{1}$$

where α is a predefined exponent. Having a higher value of α results in more intense interests in the top-ranked pieces of contents, as shown in TABLE I.

Assuming that we have $\mathcal{N}$ MUs within the area studied, these MUs independently request one piece of contents from the set $\{\mathcal{C}_i | i = 1, \ldots, \mathcal{M}\}$ with the corresponding probability defined in (1). The MUs requesting the same content $\mathcal{C}_i$ form a social group $\mathcal{G}_i$ in order to cooperatively disseminate the content of common interest across the social group. Hence, the size of the social group $\mathcal{G}_i$ requesting the same content $\mathcal{C}_i$ obeys a Binomial distribution, which is denoted as $|\mathcal{G}_i| \sim B[\mathcal{N}, \Pr(\mathcal{C}_i)]$. In order to exclude the case of $|\mathcal{G}| = 0$, we adjust the probability mass function (pmf)[7] of $|\mathcal{G}_i|$, which is expressed as

$$\Pr(|\mathcal{G}_i| = N) = \frac{\binom{\mathcal{N}}{N} \left[\Pr(\mathcal{C}_i)\right]^N \left[1 - \Pr(\mathcal{C}_i)\right]^{\mathcal{N}-N}}{1 - \left[1 - \Pr(\mathcal{C}_i)\right]^{\mathcal{N}}}. \tag{2}$$

where N is the specific size of the social group $\mathcal{G}_i$. As a result, the average $\overline{\mathcal{P}}(\mathcal{C}_i)$ of a specific delay metric associated with disseminating the content $\mathcal{C}_i$ across the social group $\mathcal{G}_i$, whose size is an adjusted-Binomially distributed random variable, can be expressed as

$$\overline{\mathcal{P}}(\mathcal{C}_i) = \sum_{N=1}^{\mathcal{N}} \mathcal{P}(|\mathcal{G}_i| = N) \cdot \Pr(|\mathcal{G}_i| = N), \tag{3}$$

[7]If no MUs requests the content, we do not have to study the content dissemination performance.

where $\mathcal{P}(|\mathcal{G}_i| = N)$ is a delay metric, which is a function of the deterministic social group size $|\mathcal{G}_i| = N$. Given the social group size N, in Section III, we will derive various delay metrics that can replace $\mathcal{P}(|\mathcal{G}_i| = N)$ in (3) in order to evaluate the impact of content popularity on the content dissemination performance.

To sum up, we assume that N MUs form a social group in order to request the content of common interest from a BS, as shown in Fig. 1. The formation of a social group depends on the following conditions:

- MUs share the same interest in a given piece of content;
- The content of common interest is of delay-sensitive nature;
- MUs roam in a bounded area having a relatively small size and they are geographically close to each other.

B. Network Layer

In order to disseminate the content of common interest across a social group, the BS creates a specific queue for buffering all the packets of the requested content and prepares for disseminating these packets one by one, as described below.

Firstly, the BSs are employed for repeatedly multicasting the packet currently at the head of the buffer, until at least one of the MUs in the target social group successfully receives it. Then, this packet is cooperatively disseminated across the social group using multicast techniques.

During the dissemination process, after successfully receiving the packet, the packet owners (POs) may make their decisions independently as to whether they would or would not forward the packet during the following stage of the dissemination, as shown in Fig. 2. Once some POs decided to further forward the packet, they would repeatedly multicast it until at least one unserved MU in the target social group successfully receives it. Afterwards, the new POs join the original PO set. Both the new POs and the original POs make new packet forwarding decisions again for the subsequent stage of dissemination. The probability of a PO willing to forward the packet is denoted as q $(0 \leq q \leq 1)$, which is termed as the Factor of Altruism (FA). At a given instant, there might not be any POs willing to further forward the packet. As a result, the unserved MUs in the target social group have to receive the packet directly from the BS. Similarly, the BS repeatedly multicasts the packet until at least one unserved MU in the target social group receives it.

During the content dissemination process, similar to the conventional BS-aided multicast, the BS keeps a specific packet at the head of the buffer, until all the MUs in the target social group successfully receive it. Then the packet is dropped from the buffer and the BS is ready to disseminate the subsequent one.

C. Physical (PHY) Layer

In the PHY layer, the radio propagation between any pair of transmitter and receiver is assumed to experience uncorrelated stationary Rayleigh flat-fading. Hence, the square of the fading amplitudes $|h_l(t)|^2$ during the t^{th} time slot (TS) obeys an *exponential distribution* having a unity mean, whose tail distribution

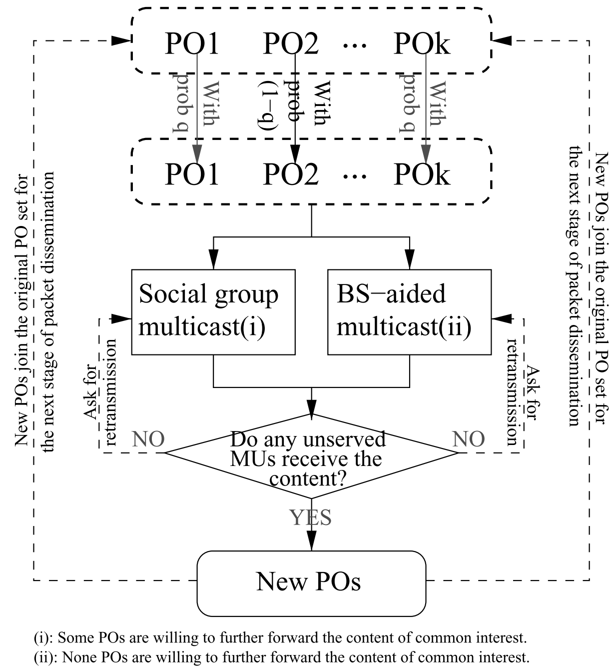


Fig. 2. Actions of POs during the spontaneous content dissemination.

function (tdf) is $\Pr[|h_l(t)|^2 > x] = \mathrm{e}^{-x}$. Given an arbitrary distance y_l in meters, the path loss (PL) Ω_l is expressed as [27]:

$$\Omega_l(y_l) = \begin{cases} 1, & y_l < d_0, \\ \left(\dfrac{4\pi f_c}{c}\right)^{\kappa} y_l^{\kappa}, & y_l \geq d_0, \end{cases} \tag{4}$$

where c is the speed of light and f_c is the carrier frequency, whereas κ is the PL exponent and d_0 is the distance from the transmitter to the 'near-field' edge.

The random distance Y_l is determined by the mobility pattern of the MUs in the target social group. The following mobility model is invoked for our performance analysis:

Definition 1 (Uniform mobility model): The position of the i^{th} MU during the t^{th} time interval is denoted by $\mathbf{P}_i(t)$, which obeys a stationary and ergodic process having a uniform distribution in the area considered. Moreover, the positions of different MUs are independently and identically distributed (i.i.d.).

This mobility model has been widely adopted for the performance analysis of MANETs [28], [29]. Let the probability density function (pdf) of the random distance Y_l between any two MUs be denoted by $f_{Y_l}(y_l)$. Our forthcoming performance analysis is applicable not only to the uniform mobility model, but to any arbitrary mobility model.

Note that, the index l in the formulas is a generic subscript, which represents 'b' when the BS is the transmitter, while it represents 's' when a MU is the transmitter. In the rest of the paper, 'l', 'b' and 's' hold the same meaning.

D. Medium-Access-Control (MAC) Layer

During a TS, a packet of the content is assumed to be successfully received by a MU, provided that the instantaneous received signal-to-noise-ratio (SNR) is higher than a predefined threshold γ [30]. In order to avoid collisions amongst multiple transmitters, orthogonal-frequency-division-multiple-access (OFDMA) or code-division-multiple-access (CDMA) may be invoked for allocating each transmitter an orthogonal channel. We denote the successful packet reception probability (SPRP) of a link as $\mu_l(y_l)$. By jointly considering the PHY layer model, the SPRP is derived as

$$\mu_l(y_l) = \Pr\left(\frac{\Pr_l^{tx}|h_l(t)|^2}{\Omega_l(y_l)N_0W_l} > \gamma\right) = \begin{cases} e^{-\frac{\gamma N_0 W_l}{\Pr_l^{tx}}}, & y_l < d_0, \\ e^{-\frac{\gamma N_0 W_l}{\Pr_l^{tx}}\left(\frac{4\pi f_c}{c}\right)^{\kappa} y_l^{\kappa}}, & y_l \geq d_0, \end{cases} \tag{5}$$

where P_l^{tx} is the corresponding transmit power and N_0W_l is the noise power in a communication bandwidth W_l. Given the pdf $f_{Y_l}(y_l)$ of the random distance Y_l, the average SPRP $\overline{\mu}_l$ of a link is derived as

$$\overline{\mu}_l = \int_0^{d_0} e^{-\frac{\gamma N_0 W_l}{P_l^{tx}}} f_{Y_l}(y_l)dy_l + \int_{y_l \geq d_0} e^{-\frac{\gamma N_0 W_l}{P_l^{tx}}\left(\frac{4\pi f_c}{c}\right)^{\kappa} y^{\kappa}} f_{Y_l}(y_l)dy_l. \tag{6}$$

Substituting the corresponding parameters and the pdf of the random distance into (6), we can obtain the average SPRP $\overline{\mu}_s$ between a pair of MUs and $\overline{\mu}_b$ between the BS and a MU. Moreover, the following lemma is proposed for our further analysis:

Lemma 1: Given the average SPRP $\overline{\mu}_l$ of a link during a TS, the average SPRP during a sufficiently short time interval Δt ($\Delta t \ll 1$ TS) is approximately $\overline{\mu}_l \Delta t$.

Proof: The proof can be found in Appendix A. ■

Note that the SPRP also represents the *normalized throughput*, whose unit is packet/TS [30]. In more details, $\overline{\mu}_l$ indicates that $\overline{\mu}_l$ packets in average can be successfully received during a TS. Therefore, during Δt (≤ 1) TS, only $\overline{\mu}_l \Delta t$ packets in average can be successfully received.

III. Delay Analysis of the Packet Dissemination

In this section, various delay metrics of the packet dissemination process are derived with respect to a specific group size N. These metrics may replace the performance function $\mathcal{P}(|\mathcal{G}_i| = N)$ in (3) in order to characterize the average performance as a function of the content popularity.

A. Pure Birth Markov Chain (PBMC)

Let us assume that there are N MUs in a considered social group. During the process of packet dissemination across the target social group, the number of POs steadily increases until all the N social group members successfully receive the packet of common interest. Hence, the packet dissemination process can be modelled by a discrete-time PBMC having $(N+1)$ states, as shown in Fig. 3. In this PBMC, the states represent the corresponding numbers of POs having received the packet. State transition only occurs from a lower-indexed state to a higher-indexed one. Specifically, the state transition emerges from state 0, which represents the initial stage of the BS-aided multicast, and terminates in state N, which indicates that all the N MUs in the target social group have received the desired packet.

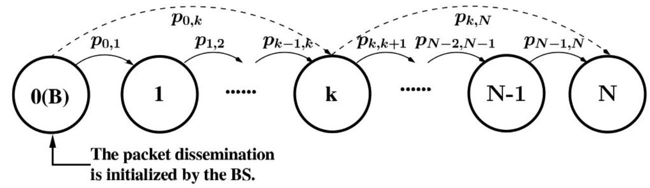


Fig. 3. A pure-birth Markov chain having an absorption state.

Let us first consider the general transition probability from state k to state $(k+m)$, where we have $1 \leq k \leq (N-1)$ and $0 \leq m \leq (N-k)$. In the light of the selfish user-behaviour considered, we assume that only n_k, $1 \leq n_k \leq k$, POs are willing to further disseminate the packet at the current stage. Therefore, any unserved MU out of the $(N-k)$ unserved ones is connected to the n_k POs by n_k wireless links, and any of these links has the probability of $\overline{\mu}_s \Delta t$ to successfully deliver the packet during the time interval Δt according to *Lemma 1*. As a result, given that n_k POs independently deliver their packets to the same target, the SPRP of an unserved MU is expressed as $[1-(1-\overline{\mu}_s\Delta t)^{n_k}]$. Furthermore, the state transition probability $p_{k,k+m|n_k\neq 0}$, which is also the probability of m out of the $(N-k)$ unserved MUs successfully receiving the packet during the current time interval Δt, can be expressed as

$$\begin{aligned} p_{k,k+m|n_k\neq 0} &= \binom{N-k}{m}\left[1-(1-\overline{\mu}_s\Delta t)^{n_k}\right]^m \\ &\quad .(1-\overline{\mu}_s\Delta t)^{n_k(N-k-m)} \\ &= \binom{N-k}{m}\left[1-\sum_{i=0}^{n_k}\binom{n_k}{i}(-\overline{\mu}_s\Delta t)^i\right]^m \\ &\quad .(1-\overline{\mu}_s\Delta t)^{n_k(N-k-m)} \\ &= \binom{N-k}{m}\left[\sum_{i=1}^{n_k}\binom{n_k}{i}(-1)^{i+1}(\overline{\mu}_s\Delta t)^i\right]^m \\ &\quad .(1-\overline{\mu}_s\Delta t)^{n_k(N-k-m)}. \end{aligned} \tag{7}$$

According to (7), the state transition probability $p_{k,k+m|n_k\neq 0}$ has the same growth rate as $\overline{\mu}_s^m \Delta t^m$. Hence, the adjacent-state transition probability $p_{k,k+1|n_k\neq 0}$ of traversing from state k to state $(k+1)$ has the same growth rate as $\overline{\mu}_s\Delta t$. Substituting $m=1$ into (7), $p_{k,k+1|n_k\neq 0}$ can be expressed as

$$\begin{aligned} p_{k,k+1|n_k\neq 0} &= (N-k)n_k\overline{\mu}_s\Delta t \\ &\quad + (N-k)\left[\sum_{i=2}^{n_k(N-k-1)}\binom{n_k(N-k-1)}{i}(-\overline{\mu}_s\Delta t)^i\right. \\ &\quad \left. - \sum_{j=2}^{n_k(N-k)}\binom{n_k(N-k)}{j}(-\overline{\mu}_s\Delta t)^j\right]. \end{aligned} \tag{8}$$

TABLE II
PARAMETERS OF THE PHY LAYER

	BS to MUs	MUs to MUs
Transmit Power	$P_b^{tx} = 31$ dBm	$P_s^{tx} = 0 \sim 10$ dBm
Carrier Freq	$f_{c,b} = 1.8$ GHz	$f_{c,s} = 2.4$ GHz
Bandwidth	$W_b = 10$ MHz	$W_s = 10$ MHz
Noise PSD	$N_0 = -174$ dBm/Hz ($20^o C$)	
SNR Threshold	$\gamma = 10$ dB	
PL Parameters	Exponent: $\kappa = 3$; Ref Distance: $d_0 = 1$ m	

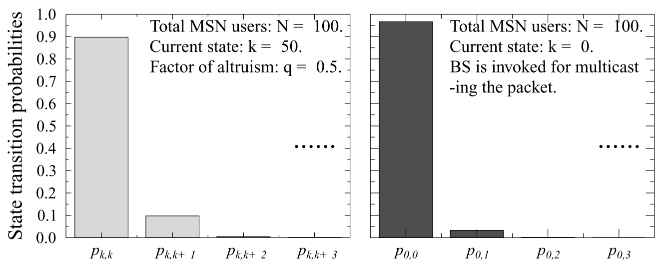


Fig. 4. State transition probabilities when $\Delta t = 0.001$ TS.

The terms in the square brackets of (8) have the same growth rate as $\overline{\mu}_s^2 \Delta t^2$. Compared to the first term $(N-k)n_k\overline{\mu}_s\Delta t$ in (8), the terms in the square brackets are negligibly low, when $\overline{\mu}_s\Delta t$ is close to zero. Hence, in this case, we can approximate $p_{k,k+1|n_k\neq 0}$ as $p_{k,k+1|n_k\neq 0} \approx (N-k)n_k\overline{\mu}_s\Delta t$. Similarly, when $\overline{\mu}_s\Delta t$ is close to zero, $p_{k,k+m|n_k\neq 0}$ associated with $m \geq 2$ in (7) can be approximated as $p_{k,k+m|n_k\neq 0} \approx 0$. Moreover, substituting $m = 0$ into (7), we obtain the probability of the PBMC sojourning in the current state k after the time interval Δt, which is $p_{k,k|n_k\neq 0} = (1-\overline{\mu}_s\Delta t)^{n_k(N-k)}$. Again, when $\overline{\mu}_s\Delta t$ is very close to zero, $p_{k,k|n_k\neq 0}$ can be approximated as $p_{k,k|n_k\neq 0} \approx 1 - n_k(N-k)\overline{\mu}_s\Delta t$.

Another scenario is that no POs are willing to forward the packet, corresponding to the case $n_k = 0$. Then the $(N-k)$ unserved MUs have to receive the packet directly from the BS. Similarly, we can also demonstrate that $p_{k,k+1|n_k=0} \approx (N-k)\overline{\mu}_b\Delta t$ and $p_{k,k|n_k=0} \approx 1-(N-k)\overline{\mu}_b\Delta t$, while $p_{k,k+m|n_k=0} \approx 0$ for $m \geq 2$, provided that $\overline{\mu}_b\Delta t$ is sufficiently small. Furthermore, it can be shown that $p_{0,1} \approx N\overline{\mu}_b\Delta t$, $p_{0,0} \approx 1 - N\overline{\mu}_b\Delta t$ and $p_{0,m} \approx 0$ for $m \geq 2$, provided that $\overline{\mu}_b\Delta t$ is sufficiently small.

According to the PHY layer parameters in TABLE II, we plot the state transition probabilities for state $k = 50$ and for state $k = 0$, respectively, in Fig. 4. We observe from Fig. 4 that the state transition probabilities of $p_{k,k+m}$ and $p_{0,m}$ for $m \geq 2$ are negligibly low, which demonstrates the high accuracy of the above approximations involved.

Therefore, assuming a sufficiently short time interval Δt, only adjacent-state transitions occur during the process modelled by the discrete-time PBMC, as shown in Fig. 3.

B. Delay of State Transition

In order to study the delay statistics of disseminating a specific packet, we need to know the specific delay that the PBMC spends in a particular state, which is termed as the *state transition delay*. As a result, the following lemma may be formulated:

Lemma 2: Given the state transition probability $\widetilde{\mu}_k\Delta t$ from the current state k to state $(k+1)$, the transition delay from state k to state $(k+1)$ obeys the exponential distribution with a mean of $1/\widetilde{\mu}_k$ TS, provided that Δt is sufficiently small. Here, $\widetilde{\mu}_k$ is termed as the transition rate.

Proof: The proof can be found in Appendix B. ■

Based on *Lemma* 2, the discrete-time PBMC seen in Fig. 3 can be further simplified to a continuous-time PBMC, which only has adjacent-state transitions. The transition rate of this continuous-time PBMC can be shown to be $p_{k,k+1}/\Delta t$, where $p_{k,k+1}$ is the adjacent-state transition probability derived in Section III-A.

Let us first consider the delay T_k of the transition from state k to $(k+1)$, when $k \geq 1$. Since each PO has a probability q of forwarding the packet, in the current state k, the number n_k $(0 \leq n_k \leq k)$ of POs willing to forward the packet obeys a Binomial distribution having a pair of parameters k and q, whose pmf is given by [31]

$$p(n_k) = \binom{k}{n_k} q^{n_k}(1-q)^{k-n_k}, \quad n_k = 0, 1, \ldots, k. \tag{9}$$

For the case of $n_k \neq 0$, we have $p_{k,k+1|n_k\neq 0} \approx n_k(N-k)\overline{\mu}_s\Delta t$. According to *Lemma* 2, the delay T_k of the transition from state k to state $(k+1)$ obeys an exponential distribution having a rate of $n_k(N-k)\overline{\mu}_s = n_k\mu_{s,k}$, where $\mu_{s,k} = (N-k)\overline{\mu}_s$. Hence, when $n_k \neq 0$, the conditional pdf, the mean and the second moment of T_k may be formulated as

$$f_{T_k|n_k}(t_k) = n_k\mu_{s,k} \cdot \mathrm{e}^{-n_k\mu_{s,k}t_k}, \ t_k \geq 0 \tag{10}$$

$$\mathcal{E}\left[T_k|\, n_k\right] = \int_0^\infty t_k f_{T_k|n_k}(t_k)dt_k = \frac{1}{n_k\mu_{s,k}}, \tag{11}$$

$$\mathcal{E}\left[T_k^2\middle|\, n_k\right] = \int_0^\infty t_k^2 f_{T_k|n_k}(t_k)dt_k = \frac{2}{(n_k\mu_{s,k})^2}. \tag{12}$$

For the case of $n_k = 0$, we have $p_{k,k+1|n_k=0} \approx (N-k)\overline{\mu}_b\Delta t$, as the MUs in the target social group have to receive the packet from the BS. According to *Lemma* 2, the delay T_k of the transition from state k to $(k+1)$ obeys an exponential distribution having a rate of $\mu_{b,k} = (N-k)\overline{\mu}_b$. Hence, given $n_k = 0$, the conditional pdf, the mean and the second moment of T_k are derived as

$$f_{T_k|n_k=0}(t_k) = \mu_{b,k} \cdot \mathrm{e}^{-\mu_{b,k}t_k}, \ t_k \geq 0 \tag{13}$$

$$\mathcal{E}\left[T_k|\, n_k=0\right] = \int_0^\infty t_k f_{T_k|n_k=0}(t_k)dt_k = \frac{1}{\mu_{b,k}}, \tag{14}$$

$$\mathcal{E}\left[T_k^2\middle|\, n_k=0\right] = \int_0^\infty t_k^2 f_{T_k|n_k=0}(t_k)dt_k = \frac{2}{\mu_{b,k}^2}. \tag{15}$$

According to the classic Bayesian principle [31], the pdf of T_k may be expressed as

$$f_{T_k}(t_k) = \sum_{n_k=1}^{k} f_{T_k|n_k}(t_k)\cdot p(n_k) + f_{T_k|n_k=0}(t_k)\cdot p(n_k=0)$$
$$= \sum_{n_k=1}^{k}\binom{k}{n_k} q^{n_k}(1-q)^{k-n_k}\cdot n_k\mu_{s,k}\mathrm{e}^{-n_k\mu_{s,k}t_k} + (1-q)^k\mu_{b,k}\mathrm{e}^{-\mu_{b,k}t_k}. \quad (16)$$

Moreover, the mean of T_k is formulated as

$$\mathcal{E}[T_k] = \mathcal{E}[T_k|\, n_k=0]\, p(n_k=0) + \sum_{n_k=1}^{k}\mathcal{E}[T_k|\, n_k]\, p(n_k)$$
$$= \underbrace{\frac{(1-q)^k}{\mu_{b,k}}}_{\mathcal{E}[T_{k,b}]} + \underbrace{\sum_{n_k=1}^{k}\binom{k}{n_k}\frac{q^{n_k}(1-q)^{k-n_k}}{n_k\mu_{s,k}}}_{\mathcal{E}[T_{k,s}]}, \quad (17)$$

where $\mathcal{E}[T_{k,b}]$ represents the average duration of the BS-aided multicasting invoked during the transition from state k to state $(k+1)$, where $\mathcal{E}[T_{k,s}]$ is the average duration of the social group multicasting during this state transition. Furthermore, the second moment of T_k is formulated as

$$\mathcal{E}\left[T_k^2\right] = \mathcal{E}\left[T_k^2\middle|\, n_k=0\right] p(n_k=0) + \sum_{n_k=1}^{k}\mathcal{E}\left[T_k^2\middle|\, n_k\right] p(n_k)$$
$$= \frac{2(1-q)^k}{\mu_{b,k}^2} + \sum_{n_k=1}^{k}\binom{k}{n_k}\frac{2q^{n_k}(1-q)^{k-n_k}}{(n_k\mu_{s,k})^2}. \quad (18)$$

From (17) and (18), we can also derive the variance of T_k by using the formula of $Var[T_k] = \mathcal{E}[T_k^2] - \{\mathcal{E}[T_k]\}^2$. Furthermore, we may simply derive the pdf, the mean and the second moment of the transition delay T_0 from state 0 to state 1 by substituting $k=0$ in (13), (14), and (15), respectively.

C. Dissemination Delay

Since the delay of the transition from a state to its successor is independent of any other state transition's delay, and given that the dissemination delay across the target social group is defined as $T_D = \sum_{k=0}^{N-1} T_k$, the mean of T_D can be expressed as

$$\mathcal{E}[T_D] = \sum_{k=0}^{N-1}\frac{(1-q)^k}{\mu_{b,k}} + \sum_{k=1}^{N-1}\sum_{n_k=1}^{k}\binom{k}{n_k}\frac{q^{n_k}(1-q)^{k-n_k}}{n_k\mu_{s,k}}, \quad (19)$$

while the variance of T_D can be formulated as $Var[T_D] = \sum_{k=0}^{N-1} Var[T_k]$.

There is no exact closed-form tdf for the dissemination delay T_D in this general case. However, given its mean and variance, we may approximate it as a random variable obeying the *Gamma distribution*, which is usually more accurate than its Gaussian counterpart, when non-negative random variables are concerned [32]. According to the theory of the Gamma distribution [33], it is uniquely and unambiguously described by its *shape parameter* $m = \{E[T_D]\}^2/Var[T_D]$ and *scale parameter* $\Theta = Var[T_D]/E[T_D]$. Then, given a delay threshold D_{th}, we may derive the approximate probability of the dissemination delay T_D exceeding D_{th} as

$$\Pr(T_D > D_{th}) \approx \frac{\Gamma\left(m, \frac{D_{th}}{\Theta}\right)}{\Gamma(m)} = \frac{\Gamma\left(\frac{\{E[T_D]\}^2}{Var[T_D]}, \frac{D_{th}E[T_D]}{Var[T_D]}\right)}{\Gamma\left(\frac{\{E[T_D]\}^2}{Var[T_D]}\right)}. \quad (20)$$

The accuracy of (20) will be verified by the Monte-Carlo simulation in Section V.

D. Individual User-Delay

A specific MU $\mathcal{A}$ in the target social group may receive the packet at any state spanning from 1 to N during the process of state transitions. When considering the transition from state $(k-1)$ to k $(1 \le k \le N)$, any of the $(N-k+1)$ unserved MUs may successfully receive the packet with a probability of $1/(N-k+1)$, and may not receive it with a probability of $(N-k)/(N-k+1)$. Specifically, the probability of $\mathcal{A}$ receiving the packet in state k, which naturally implies that $\mathcal{A}$ has not received the packets at any of the previous states, may be expressed as

$$p_k = \frac{1}{N-k+1}\cdot\prod_{i=1}^{k-1}\frac{N-i}{N-i+1} = \frac{1}{N}, \quad 1 \le k \le N. \quad (21)$$

Hence, given that $\mathcal{A}$ receives the packet in state k, the individual user-delay of $\mathcal{A}$ is expressed as $T_{\mathcal{A}|k} = \sum_{j=0}^{k-1} T_k$ and the conditional pdf of $T_{\mathcal{A}|k}$ is expressed as $f_{T_{\mathcal{A}|k}}(t_{\mathcal{A}}) = f_{T_0+\cdots+T_{k-1}}(t_{\mathcal{A}})$. According to the Bayesian principle [31], the pdf of the individual user-delay $T_{\mathcal{A}}$ can be expressed as:

$$f_{T_{\mathcal{A}}}(t_{\mathcal{A}}) = \sum_{k=1}^{N} f_{T_{\mathcal{A}|k}}(t_{\mathcal{A}})\cdot p_k = \sum_{k=1}^{N}\frac{f_{T_0+\cdots+T_{k-1}}(t_{\mathcal{A}})}{N}. \quad (22)$$

Furthermore, owing to the fact that $\{T_0, T_1, \ldots, T_{k-1}\}$ are independent of each other, the average of $T_{\mathcal{A}}$ can be obtained as

$$\mathcal{E}[T_{\mathcal{A}}] = \int_0^{\infty} t_{\mathcal{A}}\sum_{k=1}^{N}\frac{f_{T_0+\cdots+T_{k-1}}(t_{\mathcal{A}})}{N}dt_{\mathcal{A}} = \sum_{k=1}^{N}\frac{1}{N}\cdot\sum_{i=0}^{k-1}\mathcal{E}[T_i]$$
$$= \sum_{k=1}^{N}\frac{N-k+1}{N}\mathcal{E}\left[T_{k-1}\right], \quad (23)$$

where $\mathcal{E}\left[T_{k-1}\right]$ is given by (17). Furthermore, the second moment of $T_{\mathcal{A}}$ is given by

$$\mathcal{E}[T_{\mathcal{A}}^2] = \int_0^{\infty}\sum_{k=1}^{N}\frac{t_{\mathcal{A}}^2 f_{T_0+\cdots+T_{k-1}}(t_{\mathcal{A}})}{N}dt_{\mathcal{A}}$$
$$= \sum_{k=1}^{N}\frac{\mathcal{E}\left[(T_0+T_1+\cdots+T_{k-1})^2\right]}{N} = \sum_{k=1}^{N}\sum_{i,j=0}^{k-1}\frac{\mathcal{E}[T_iT_j]}{N}$$
$$= \sum_{k=1}^{N}\frac{N-k+1}{N}\mathcal{E}[T_{k-1}^2] + \sum_{k=1}^{N}\frac{\boldsymbol{\xi}_k^T\left[\mathbf{H}_k - \mathbf{I}_k\right]\boldsymbol{\xi}_k}{N}, \quad (24)$$

where $\boldsymbol{\xi}_k = (\mathcal{E}[T_0], \mathcal{E}[T_1], \ldots, \mathcal{E}[T_{k-1}])^T$, $\mathbf{H}_k$ is a $k \times k$ matrix, whose elements are all ones, and $\mathbf{I}_k$ is a $k \times k$ identity matrix. Consequently, the variance of $T_{\mathcal{A}}$ can be expressed as $Var(T_{\mathcal{A}}) = \mathcal{E}[T_{\mathcal{A}}^2] - \{\mathcal{E}[T_{\mathcal{A}}]\}^2$. Hence, by substituting $\mathcal{E}[T_{\mathcal{A}}]$ and $Var[T_{\mathcal{A}}]$ into (20), we may obtain the approximate probability of $T_{\mathcal{A}}$ exceeding threshold D_{th}.

IV. Delay Metrics for Special Cases

A. Case 1: Conventional BS-Aided Multicast (q = 0)

In this pessimistic case, all the MUs in the target social group are selfish during the packet dissemination process. Hence, the BS has to disseminate the packet to all the MUs in the target social group.

1) Dissemination Delay: When FA is $q = 0$, according to Eqs.(13)∼(15) in Section III-B, the state transition delays $\{T_k, k = 0, 1, \ldots, (N-1)\}$ are the independent exponentially distributed variables associated with the rates of $\{\widetilde{\mu}_k = (N - k)\overline{\mu}_b, k = 0, 1, \ldots, (N-1)\}$. Since the dissemination delay is defined as $T_D = \sum_{k=0}^{N-1} T_k$, T_D obeys the *hypoexponential distribution* [34]. Furthermore, since the rates of $\{T_k, k = 0, 1, \ldots, (N-1)\}$ are different from each other, the pdf of T_D can be expressed as

$$f_{T_D|q=0}(t_D) = \sum_{k=0}^{N-1} \prod_{j=0, j\neq k}^{N-1} \frac{N-j}{k-j} (N-k)\overline{\mu}_b \mathrm{e}^{-(N-k)\overline{\mu}_b t_D}. \tag{25}$$

In order to derive the probability of T_D exceeding a given threshold D_{th}, we integrate the above pdf $f_{T_D|q=0}(t_D)$ over the region $[D_{th}, \infty)$, which is expressed as

$$\begin{aligned}\Pr(T_D > D_{th}|q=0) &= \int_{D_{th}}^{\infty} f_{T_D|q=0}(t_D) dt_D \\ &= \sum_{k=0}^{N-1} \prod_{j=0, j\neq k}^{N-1} \frac{N-j}{k-j} \mathrm{e}^{-(N-k)\overline{\mu}_b D_{th}}.\end{aligned} \tag{26}$$

2) Individual User-Delay: When the FA is $q = 0$, the individual user-delay is solely determined by the quality of the wireless link connecting the MU $\mathcal{A}$ to the BS. As a result, according to *Lemma 2*, the individual user-delay $T_{\mathcal{A}}$ obeys an exponential distribution having a mean of $1/\overline{\mu}_b$. Furthermore, the probability of $T_{\mathcal{A}}$ exceeding a given threshold D_{th} is derived as $\Pr(T_{\mathcal{A}} > D_{th}|q=0) = \exp(-\overline{\mu}_b D_{th})$.

B. Case 2: Fully Altruistic Behaviours (q = 1)

In this optimistic scenario, all the MUs in the target social group are completely altruistic. Since there are always some POs willing to forward the packet during the dissemination process, the BS is not invoked for multicasting the packet any more, once some of the MUs have initially received it from the BS.

1) Dissemination Delay: When the FA is $q = 1$, by substituting $n_k = k$ into Eqs.(10)∼(12) in Section III-B, we know that the state transition delays $\{T_k, k = 1, \ldots, (N-1)\}$ are independent exponentially distributed variables associated with the rates of $\{\widetilde{\mu}_k = k(N-k)\overline{\mu}_s, k = 1, 2, \ldots, (N-1)\}$. Furthermore, by substituting $k = 0$ into Eqs.(13)∼(15) in Section III-B, the initial state transition delay T_0 is also an exponentially distributed variable associated with a rate of $\widetilde{\mu}_0 = N\overline{\mu}_b$. Note furthermore that T_0 is also independent of $\{T_k, k = 1, \ldots, (N-1)\}$. Since the dissemination delay is defined as $T_D = \sum_{k=0}^{N-1} T_k$, T_D obeys the hypoexponential distribution.

However, the rates of $\{\widetilde{\mu}_k = k(N-k)\overline{\mu}_s, k = 1, 2, \ldots, (N-1)\}$ associated with $\{T_k, k = 1, \ldots, (N-1)\}$ exhibit a symmetric structure. For example, the rates of T_k and T_{N-k} share the same value of $k(N-k)\overline{\mu}_s$. Hence, the closed-form equation for the tdf of T_D may only be expressed in the form of a *continuous phase-type distribution* [35]. As a result, when $q = 1$, the transition rate matrix of the PBMC is expressed as

$$\mathbf{P} = \begin{pmatrix} -\widetilde{\mu}_0 & \widetilde{\mu}_0 & 0 & \cdots & 0 & 0 \\ 0 & -\widetilde{\mu}_1 & \widetilde{\mu}_1 & \ddots & 0 & 0 \\ \vdots & \ddots & -\widetilde{\mu}_k & \widetilde{\mu}_k & \ddots & \vdots \\ 0 & 0 & \ddots & -\widetilde{\mu}_{N-2} & \widetilde{\mu}_{N-2} & 0 \\ 0 & 0 & \cdots & 0 & -\widetilde{\mu}_{N-1} & \widetilde{\mu}_{N-1} \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{Q} & \mathbf{Q}_0 \\ \mathbf{0} & 0 \end{pmatrix}, \tag{27}$$

where $\mathbf{Q}$ is a $(N \times N)$-element matrix containing all the transition rates between transient states, $\mathbf{Q}_0$ is a $(N \times 1)$ column vector containing all the transition rates from transient states to the absorbing state N, whose last entry is $\widetilde{\mu}_{N-1}$ and finally, the remaining entries are all zeros. As shown in Fig.3, the packet dissemination process starts from the initial state 0. Thus, the probability of T_D exceeding a given threshold D_{th} is expressed as

$$\Pr(T_D \geq D_{th}|q=1) = \boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q}) \times \mathbf{1}_N. \tag{28}$$

Note that in (28), the $(N \times 1)$ column vector $\boldsymbol{\tau}_{k+1}$ $(0 \leq k \leq N-1)$, whose $(k+1)$th entry is one but all the others are zeros, indicates that the PBMC starts at state k, while the $(N \times 1)$ column vector $\mathbf{1}_{k+1}$, whose first $(k+1)$ entries are ones and the remaining entries are zeros, indicates that the PBMC process is absorbed at state $(k+1)$. The proof of (28) can be found in [36].

2) Individual User-Delay: Given an event that the MU $\mathcal{A}$ successfully receives the packet at state $(k+1)$ $(0 \leq k \leq N-1)$, the PBMC used for modelling the packet dissemination in Fig.3 is considered to be terminated at state $(k+1)$. According to the physical meaning of both $\boldsymbol{\tau}_{k+1}$ and $\mathbf{1}_{k+1}$, similar to (28), the probability of $T_{\mathcal{A}}$ exceeding the threshold D_{th}, given that $\mathcal{A}$ receives the desired packet at state $(k+1)$ for $(0 \leq k \leq N-1)$, is expressed as

$$\Pr(T_{\mathcal{A}} \geq D_{th}|q=1, k+1) = \boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q}) \times \mathbf{1}_{k+1}. \tag{29}$$

Since we have already derived the probability of $p_{k+1} = 1/N$ that $\mathcal{A}$ receives the packet at state $(k+1)$ in (21), according to the Bayesian principle [31], the probability of $T_{\mathcal{A}}$ exceeding the threshold D_{th} is derived as

$$\Pr(T_{\mathcal{A}} \geq D_{th}|q=1) = \sum_{k=0}^{N-1} \Pr(T_{\mathcal{A}} \geq D_{th}|q=1, k+1) \cdot p_{k+1}$$
$$= \sum_{k=0}^{N-1} \frac{\boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q}) \times \mathbf{1}_{k+1}}{N} = \frac{\boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q})}{N} \times \sum_{k=0}^{N-1} \mathbf{1}_{k+1}$$
$$= \frac{\boldsymbol{\tau}_1^T \times \mathbf{exp}(D_{th}\mathbf{Q}) \times \boldsymbol{\eta}}{N}, \tag{30}$$

where $\boldsymbol{\eta} = (N, N-1, \ldots, 1)^T$ is a $(N \times 1)$ column vector.

C. Case 3: Moderately Altruistic Behaviours (q = 0.5)

Unfortunately, we are unable to derive the exact tdf for the scenario, when the FA is set to $q = 0.5$. However, we are still able to offer some interesting insights concerning the delay metrics of this specific case. Substituting $q = 0.5$ into the second term of (17), the average duration of the social group multicast process during the transition from state k to $(k+1)$ for $k \geq 1$ can be given by

$$\mathcal{E}[T_{k,s}|q=0.5] = \frac{1}{2^k \cdot \mu_{s,k}} \sum_{n_k=1}^{k} \binom{k}{n_k} \frac{1}{n_k}. \tag{31}$$

According to Eq.(68.1) of [33], we arrive at the following lower bound for $\mathcal{E}[T_{k,s}|q=0.5]$, which is expressed as:

$$\mathcal{E}[T_{k,s}|q=0.5] > \frac{1}{2^k \cdot \mu_{s,k}} \left[\sum_{n_k=0}^{k} \binom{k}{n_k} \frac{1}{n_k+1} - 1 \right]$$
$$= \frac{1}{2^k \cdot \mu_{s,k}} \frac{2^{k+1} - k + 2}{k+1}. \tag{32}$$

Similarly, substituting $q = 1.0$ into the second term of (17), the corresponding formula of $\mathcal{E}[T_{k,s}|q=1.0]$ for this fully altruistic behaviour may be expressed as $\mathcal{E}[T_{k,s}|q=1.0] = 1/(k\mu_{s,k})$. As a result, the ratio $\mathcal{R}_{k,s}$ of these two expressions can be formulated as

$$\mathcal{R}_{k,s} = \frac{\mathcal{E}[T_{k,s}|q=0.5]}{\mathcal{E}[T_{k,s}|q=1.0]} > \frac{(2^{k+1} - k + 2)k}{2^k(k+1)}. \tag{33}$$

In the ideal scenario, when k tends to infinity, this ratio can be expressed as $\lim_{k\to\infty} R_{k,s} > 2$. Since the lower bound derived in (32) is very tight[8], we can summarise that by assuming moderately altruistic behaviours, the average duration of the social group multicasting during the transition from state k to $(k+1)$ is twice that of the fully altruistic scenario, provided that k is sufficiently high.

Let us now demonstrate the tightness of the lower bound (32) in terms of the average dissemination delay. Substituting (32) into (19), the lower bound of the average dissemination delay $\mathcal{E}[T_D|q=0.5]$ can be formulated as

$$\mathcal{E}[T_D|q=0.5] = \sum_{k=0}^{N-1} \frac{1}{2^k \mu_{b,k}} + \sum_{k=1}^{N-1} \frac{1}{2^k \mu_{s,k}} \sum_{n_k=1}^{k} \binom{k}{n_k} \frac{1}{n_k}$$
$$> \sum_{k=0}^{N-1} \frac{1}{2^k \mu_{b,k}} + \sum_{k=1}^{N-1} \frac{1}{2^k \cdot \mu_{s,k}} \frac{2^{k+1} - k + 2}{k+1}. \tag{34}$$

When we compute the exact result of $\mathcal{E}[T_D|q=0.5]$, which is represented by the first line of (34), and its lower bound, which is quantified by the second line of (34), then for a large social group size N, such as $N = 50\sim200$, using a set of other related parameters in line with those of Fig. 6, the root-mean-square-deviation (RMSD) of these two sets of results can be shown to be 0.094 TS. Hence, we can claim that for a large social group size, which represents our densely populated scenario, the lower bound expressed in (34) can be regarded as an approximate result of $\mathcal{E}[T_D|q=0.5]$. Furthermore, the tightness of the lower bound derived in (33) can also be readily demonstrated.

Similarly, with the aid of (32), we can also obtain the lower bound for the average individual user-delay $\mathcal{E}(T_{\mathcal{A}}|q=0.5)$.

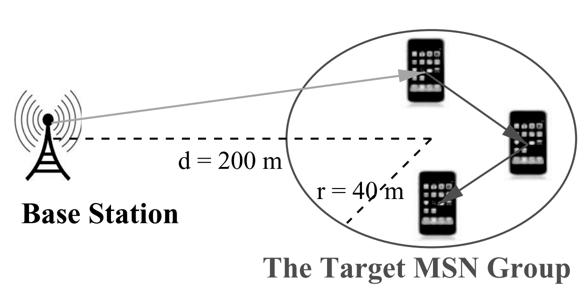


Fig. 5. Geographic features for obtaining numerical results.

V. Numerical Results

The parameters of the PHY layer are presented in TABLE II. The specific parameters used for transmissions from the BS to the MUs are in line with FDD-LTE standard[9], while the transmission parameters between the MUs are in line with the commonly used 802.11 protocol [18].

As shown in Fig. 5, we assume that all MUs in the target social group roam in a circular area having a radius of $r = 40$ m by obeying the *uniform mobility model*. The BS is $d = 200$ m away from the centre of the circular area. In this scenario, the pdf $f_{Y_s}(y_s)$ of the distance between a pair of MUs is given by Eq. (23) of [38], and $f_{Y_b}(y_b)$ between the BS and a MU can be found in our technical report [39]. Substituting $f_{Y_s}(y_s)$ and $f_{Y_b}(y_b)$ into (6), alongside the parameters offered in TABLE II, we may obtain the average SPRP $\overline{\mu}_s$ and $\overline{\mu}_b$, which further lead us to the analytical (ana) results for the various metrics. If we let $q = 0$ in our model, the corresponding analytical results are derived for conventional BS-aided multicast.

In order to obtain a reliable statistical characterization of the simulation performance (sim), we repeatedly run Monte-Carlo

[8]The tightness of this lower bound will be demonstrated in the following paragraph in terms of the average dissemination delay.

[9]We assume a 1.8 GHz carrier frequency in line with the LTE networks operated by the British company EE [37].

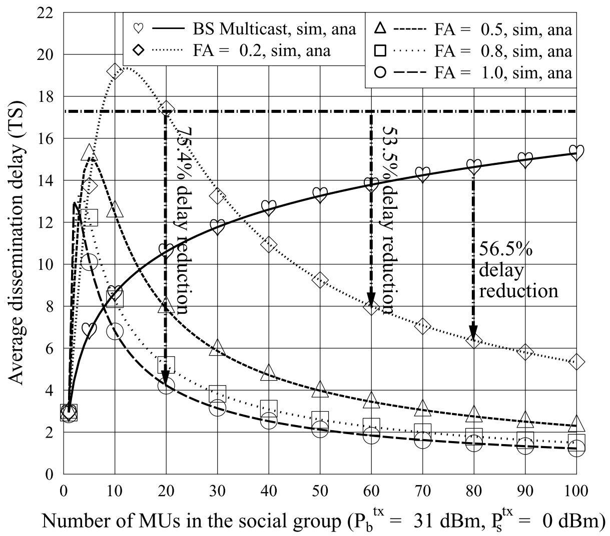


Fig. 6. Average dissemination delay affected by the number of MUs in the target social group, which is parameterized by the FA. The analytical results were evaluated from Eq. (19).

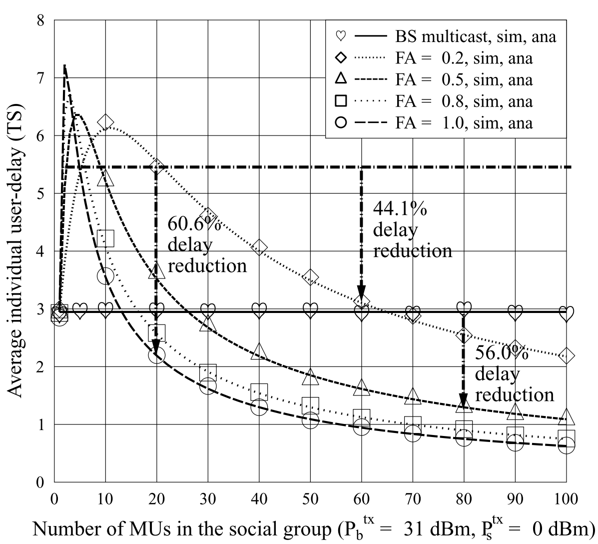
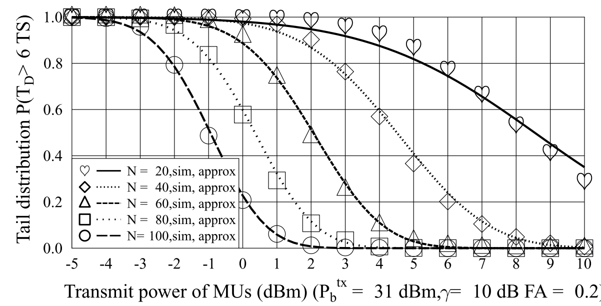
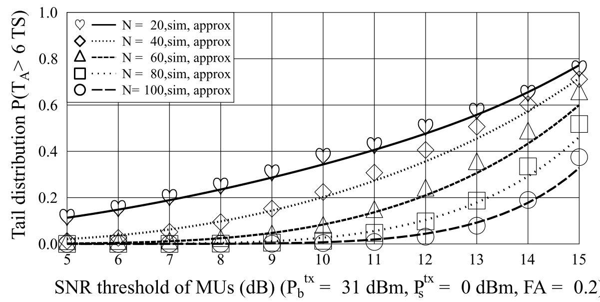


Fig. 7. Average individual user-delay as a function of the number of MUs in the target social group, which is parameterized by the FA. The analytical results were evaluated from (23).



(a) Dissemination delay



(b) Individual user-delay

Fig. 8. The tail distribution of the delay versus (a) the transmit power and (b) the SNR threshold for successful reception, which is parameterized by the number of MUs in the target social group. The analytical results were either directly or indirectly derived from Eq.(20).

simulations 10 000 times and set the time-interval of our system to be $\Delta t = 0.001$ TS, where a TS can be considered as a packet duration. All the delay related metrics are evaluated by the number of TSs. In the numerical results of Figs. 6–8, we study the impact of the social group size N on the delay metrics of the packet dissemination process without considering any specific content popularity.

A. Delay Metrics for Uniform Mobility Model

As shown in Fig. 6, when FA $\neq 0$, the average dissemination delay firstly increases, as the number of MUs is increased. When only a few MUs are in the target social group, a longer period is required for disseminating the packet to all of the group members due to the increasing content demand of the unserved MUs. However, by further increasing the number of MUs, the diversity gain incurred by the cooperation of the multiple multicasters becomes sufficiently high to mitigate the adverse effect of the increasing content demand. As a result, we observe that the average dissemination delay decays after reaching its peak, as the number of MUs is further increased. For example, for FA $= 0.2$, the delay is reduced by 53.5%, as the number of MUs is increased from $N = 20$ to 60. Furthermore, a higher FA incurs a lower delay, since more POs are willing to forward the packet after they successfully receive it. For example, for $N = 20$, the average dissemination delay is reduced by 75.4%, as the FA is increased from 0.2 to 1. By contrast, when FA $= 0$, the conventional BS-aided multicast technique is invoked. However, as the number of the MUs increases, the average dissemination delay also increases. We observe from Fig. 6 that our approach is capable of reducing the average dissemination delay of the conventional BS-aided multicast by 56.5% for $N = 80$, when a small FA value of 0.2 is assumed.

As shown in Fig. 7, when only a few MUs are in the target social group and the FA is non-zero, due to the users' selfishness, fewer than two POs are willing to forward the packet during the dissemination process. Therefore, we observe from Fig. 7 that the average individual user-delay initially increases, because it does not benefit from any diversity gain. However, as we further increase the number of MUs, an increasing number of POs become willing to forward the packet, which substantially reduces the average individual user-delay, as observed from Fig. 7. For example, for FA $= 0.2$, the average individual user-delay is reduced by 44.1%, as the number of MUs is increased from $N = 20$ to 60. Nevertheless, when the conventional BS-aided multicast is invoked, the average individual user-delay, which only relies on the link connecting this

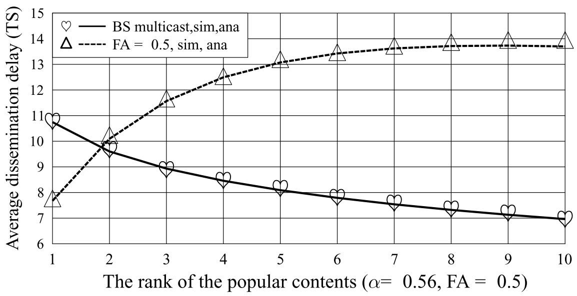


Fig. 9. Average dissemination delay as a function of the rank of the popular content. The transmit power of the BS is $P_b^{tx} = 31$ dBm and the transmit power of a MU is $P_t^{tx} = 0$ dBm. $\mathcal{N} = 100$ MUs independently request $\mathcal{M} = 10$ ranked-popularity pieces of contents according to the request probabilities listed in TABLE I when $\alpha = 0.56$. The analytical results were evaluated from Eq.(3).

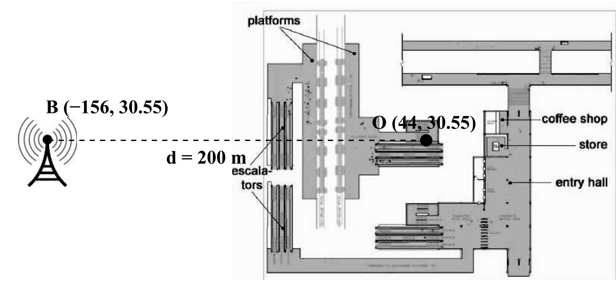


Fig. 10. A densely popluated subway station.

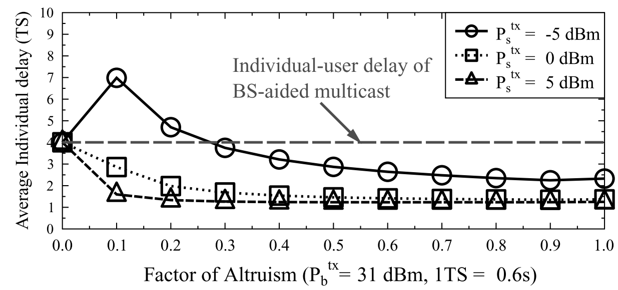


Fig. 11. Average individual user-delay in a subway station when all the MUs in the subway station form a grand social group for downloading a content of common interest.

specific MU to the BS, remains near-constant at 2.95 TS, as the number of MUs increases. Furthermore, the average individual user-delay is improved, when we increase the value of the FA. For example, given $N = 20$ MUs in the target social group, the average individual user-delay is reduced by 60.6%, as the FA is increased from 0.2 to 1.0. Additionally, given $N = 80$ MUs in the target social group, the average individual user-delay drops from 2.95 TS to 1.3 TS, comparing the conventional BS-aided multicast to our approach associated with FA $= 0.5$.

Observe in Fig. 8(a) that the probability of the dissemination delay exceeding a threshold of $D_{th} = 6$ TS reduces upon increasing the transmit power of each MU. By contrast, as portrayed in Fig. 8(b), the probability of the individual user-delay exceeding the same threshold increases upon increasing the SNR threshold to be exceeded for ensuring successful packet reception. Our Gamma-distribution-based approximations match the simulation results.

Then, we study the average dissemination delay as a function of the specific popularity of the pieces of contents in Fig. 9. Observe from Fig. 9 that as a piece of contents becomes less popular, the average dissemination delay of our scheme increases, when we have a moderate degree of altruism associated with FA $= 0.5$. When a piece of content is less popular, fewer MUs may request this content, hence the resultant smaller social group fails to provide sufficient cooperative multicast opportunities for rapidly disseminating the packet across the social group. By contrast, since a less popular piece of contents results in a lower content demand, the average dissemination delay of the BS-aided multicast reduces, as the content becomes less popular. Furthermore, as shown in Fig. 9, our scheme associated with FA $= 0.5$ outperforms the conventional BS-aided multicast in terms of its delay of disseminating the most popular content. Nevertheless, the BS-adied multicast is more suitable for disseminating the less popular pieces of contents.

B. Investigations Using Real Mobility Traces

Let us now study the content dissemination performance in a densely-populated subway station scenario [40]. The mobility traces for this scenario can be downloaded from the CRAWDAD database[10]. The active area in this scenario is 1921 m^2. After analysing the mobility traces, the centre O of the active area is found to be at the coordinates of (44, 30.55) m, as shown in Fig. 10. In our simulations, we placed the BS at the point (−156, 30.55) m, which is 200 m away from the centre of the subway station. Since the MUs arrive/depart either through the entrances or during the arrival/departure of trains, the number of MUs is dynamic during the simulation time. As a result, we cannot readily obtain the dissemination delay in this scenario. However, we are still able to evaluate the individual user-delay, when our content dissemination scheme and conventional BS-aided multicast scheme are invoked. Again, the physical layer parameters are summarised in TABLE II. Since the positions of the MUs are captured every 0.6 s in this mobility trace, in our simulations we set the basic time interval of $\Delta t = 0.6$ s as a single TS, which can be considered as a packet's duration. Then the delay was evaluated in terms of the number of TSs.

[10]http://crawdad.cs.dartmouth.edu/kth/walkers/

We first assume that all the MUs in the subway station form a large social group in order to download the train schedule of common interest. Observe from Fig. 11 that for the cases of $P_s^{tx} = 0$ dBm and $P_s^{tx} = 5$ dBm, the average individual user-delay is reduced, as we increase the FA from 0.0 to 1.0. For $P_s^{tx} = -5$ dBm, when FA is increased from 0.0 to 0.1, we observe an increasing average individual user-delay. This is because the SPRP between the MUs is low and also, because fewer POs are willing to forward the packet. As FA becomes higher, more POs may join to assist the packet dissemination process, which significantly reduces the average individual user-delay. Specifically, when FA $= 0$, conventional BS-aided multicast is invoked for disseminating the packets. For $P_s^{tx} = 0$ or 5 dBm, if the MUs become only modestly altruistic, say we have FA $= 0.1$, our content dissemination scheme outperforms

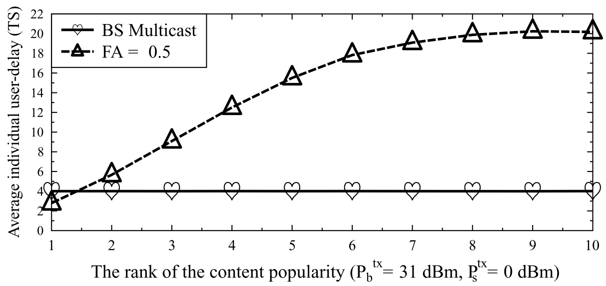


Fig. 12. Average individual user-delay in a subway station when the MUs in the subway station independently request $\mathcal{M} = 10$ ranked-popularity pieces of contents according to the probabilities listed in TABLE I when $\alpha = 1$.

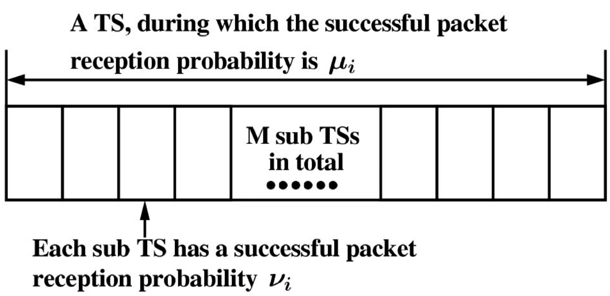


Fig. 13. The structure of a TS.

the conventional BS-aided multicast. For $P_s^{tx} = -5$ dBm, our scheme starts to outperform the classic BS-aided multicast, provided that FA is higher than 0.4.

We then study the impact of the specific content popularity on the average individual user-delay in a subway station. Observe from Fig. 12 that when disseminating the most popular content in the subway station, our dissemination scheme associated with FA $= 0.5$ outperforms the conventional BS-aided multicast. However, the BS-aided multicast is more suitable for disseminating less popular content in this scenario. The reason behind this trend is the same as that associated with Fig. 9.

VI. Conclusions

In this paper, we proposed a social group multicast aided content dissemination scheme as a supplement to the conventional cellular system. The content popularity is modelled by a Zipf distribution and the concept of FA was introduced for the sake of quantifying the probability of a PO forwarding a packet of the content of common interest. In our scheme, the BSs are invoked for multicasting the packet at the initial stage, as well as when no POs are willing to share the packet with others. By modelling the packet dissemination process as a PBMC, closed-form expressions were derived for the statistical properties of the various delay metrics. We demonstrated that our approach outperforms the conventional BS-adied multicast in terms of both the dissemination delay and the individual-user delay, especially when the density of MUs in a target group is high. Furthermore, we found that our approach is more suitable for disseminating a more popular content. By contrast, the conventional BS-aided multicast performs better for disseminating a less popular content.

Appendix A
The Proof of Lemma 1

As shown in Fig. 13, a TS is divided into M sub-TSs, each of which has a duration of $\Delta t = 1/M$ TS. We assume that the SPRP in a sub-TS is ν_i. As a result, given the SPRP $\overline{\mu}_i$ in a TS, we may derive the relation between $\overline{\mu}_i$ and ν_i, which is expressed as

$$\overline{\mu}_i = \sum_{j=1}^{M} (1-\nu_i)^{j-1} \nu_i = 1 - (1-\nu_i)^M. \tag{35}$$

Rewriting the above expression, we obtain

$$\nu_i = 1 - (1-\overline{\mu}_i)^{1/M} = 1 - (1-\overline{\mu}_i)^{\Delta t}, \tag{36}$$

where the second equality is derived according to $\Delta t = 1/M$ TS. If we expand $(1-\overline{\mu}_i)^{\Delta t}$ according to the Taylor series, we have

$$(1-\overline{\mu}_i)^{\Delta t} = \sum_{n=0}^{\infty} \binom{\Delta t}{n} (-\overline{\mu}_i)^n = 1 - \overline{\mu}_i \Delta t + O(\overline{\mu}_i^2), \tag{37}$$

where $O(\overline{\mu}_i^2)$ is the infinitesimal by small quantity on the same order as $\overline{\mu}_i^2$. Substituting the above equation into (36), we have

$$\nu_i = \overline{\mu}_i \Delta t + O(\overline{\mu}_i^2) \approx \overline{\mu}_i \Delta t. \tag{38}$$

According to our experiments, if we vary $\overline{\mu}_i$ from 0 to 0.8, the *root-mean-square-deviation* (RMSD) between the exact ν_i given by (36) and the approximated ν_i given by (38) is 9.45×10^{-4}. As a result, it is reasonable to claim that $\nu_i \approx \overline{\mu}_i \Delta t$.

Appendix B
The Proof of Lemma 2

During a time interval Δt, the PBMC may transit from state k to $(k+1)$ with a probability of $\widetilde{\mu}_k \Delta t$. Naturally, the successful state transition first occurring during the $(M_k = m_k)$-th Δt interval obeys a *geometric distribution*. According to the PMF of a geometric distribution having a parameter of $\widetilde{\mu}_k \Delta t$, we arrive at:

$$\Pr(M_k \Delta t \le m_k \Delta t) = \sum_{m=1}^{m_k} (1 - \widetilde{\mu}_k \Delta t)^{m-1} \widetilde{\mu}_k \Delta t, \tag{39}$$

$$\Pr(M_k \Delta t \le (m_k+1) \Delta t) = \sum_{m=1}^{m_k+1} (1 - \widetilde{\mu}_k \Delta t)^{m-1} \widetilde{\mu}_k \Delta t. \tag{40}$$

The continuous-valued delay of the adjacent-state transition is denoted as $T_k = M_k \Delta t$, which is associated with a specific value of $t_k = m_k \Delta t$. Hence, we may derive the pdf of T_k as:

$$\begin{aligned} f_{T_k}(t_k) &= \lim_{\Delta t \to 0} \frac{\Pr(T_k \le t_k + \Delta t) - \Pr(T_k \le t_k)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\Pr(M_k \Delta t \le (m_k+1)\Delta t) - \Pr(M_k \Delta t \le m_k \Delta t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{(1-\widetilde{\mu}_k \Delta t)^{m_k} \widetilde{\mu}_k \Delta t}{\Delta t} = \lim_{\Delta t \to 0} \widetilde{\mu}_k \mathrm{e}^{-m_k \widetilde{\mu}_k \Delta t} \\ &= \widetilde{\mu}_k \cdot \mathrm{e}^{-\widetilde{\mu}_k t_k}, \end{aligned} \tag{41}$$

where the last two lines are derived based on $\lim_{\Delta t \to 0} \widetilde{\mu}_k \Delta t = 1 - \mathrm{e}^{-\widetilde{\mu}_k \Delta t}$ and $m_k = t_k/\Delta t$, respectively.

REFERENCES

[1] N. Kayastha, D. Niyato, P. Wang, and E. Hossain, "Applications, architectures, and protocol design issues for mobile social networks: A survey," *Proc. IEEE*, vol. 99, no. 12, pp. 2130–2158, Dec. 2011.

[2] A. Vahdat and D. Becker, "Epidemic routing for partially-connected ad hoc networks," Master thesis, Dept. Comput. Sci., Duke Univ., Durham, NC 27708 USA, Tech. Rep., 2000.

[3] Y.-K. Ip, W.-C. Lau, and O.-C. Yue, "Performance modeling of epidemic routing with heterogeneous node types," in *Proc. IEEE Int. Conf. Commun. (ICCŠ08)*, May 2008, pp. 219–224.

[4] A. Clementi, F. Pasquale, and R. Silvestri, "Opportunistic manets: Mobility can make up for low transmission power," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 610–620, Feb. 2013.

[5] H. Sun and C. Wu, "Epidemic forwarding in mobile social networks," in *Proc. IEEE Int. Conf. Commun. (ICC'12)*, Jun. 2012, pp. 1421–1425.

[6] J. Whitbeck, V. Conan, and M. D. de Amorim, "Performance of opportunistic epidemic routing on edge-markovian dynamic graphs," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1259–1263, May 2011.

[7] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. IEEE INFOCOM*, 2009, pp. 1422–1430.

[8] C. Boldrini, M. Conti, and A. Passarella, "Contentplace: Social-aware data dissemination in opportunistic networks," in *Proc. 11th Int. Symp. Model. Anal. Simul. Wireless Mobile Syst. (MSWiM'08)*, 2008, pp. 203–210.

[9] D. Niyato, P. Wang, W. Saad, and A. Hjϕandrungnes, "Controlled coalitional games for cooperative mobile social networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1812–1824, May 2011.

[10] K. Akkarajitsakul, E. Hossain, and D. Niyato, "Cooperative packet delivery in hybrid wireless mobile networks: A coalitional game approach," *IEEE Trans. Mobile Comput.*, vol. 12, no. 5, pp. 1–15, May 2013.

[11] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Trans. Mobile Comput.*, vol. 11, no. 5, pp. 821–834, May 2012.

[12] Y. Li, M. Qian, D. Jin, P. Hui, Z. Wang, and S. Chen, "Multiple mobile data offloading through disruption tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 7, pp. 1579–1596, Jul. 2014.

[13] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Perform. Eval.*, vol. 62, nos. 1–4, pp. 210–228, Oct. 2005.

[14] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, "Power law and exponential decay of intercontact times between mobile devices," *IEEE Trans. Mobile Comput.*, vol. 9, no. 10, pp. 1377–1390, Oct. 2010.

[15] J. Hu, L.-L. Yang, and L. Hanzo, "Mobile social networking aided content dissemination in heterogeneous networks," *China Commun.*, vol. 10, no. 6, p. 1, 2013.

[16] J. Wang, S. Park, D. Love, and M. Zoltowski, "Throughput delay tradeoff for wireless multicast using hybrid-ARQ protocols," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2741–2751, Sep. 2010.

[17] J. Seo, T. Kwon, and V. Leung, "Social groupcasting algorithm for wireless cellular multicast services," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 47–50, Jan. 2013.

[18] Information Technology–Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks–Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, ISO/IEC/IEEE Standard 8802–11:2012(E), Nov. 2012, pp. 1–2798.

[19] Z. Gong and M. Haenggi, "Interference and outage in mobile random networks: Expectation, distribution, and correlation," *IEEE Trans. Mobile Comput.*, vol. 13, no. 2, pp. 337–349, Feb. 2014.

[20] H. Kwon and B. G. Lee, "Cooperative power allocation for broadcast/multicast services in cellular OFDM systems," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3092–3102, Oct. 2009.

[21] D. Feng, L. Lu, Y. Yuan-Wu, G. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3541–3551, Aug. 2013.

[22] R. E. Hattachi and J. Erfanian, "NGMN 5G White Paper version 1.0," *Next Generation Mobile Networks (NGMN)*, NGMN Alliance, Frankfurt, Germany, 2015.

[23] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas. (IMC'07)*, Oct. 2007, p. 1.

[24] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Characteristics of YouTube network traffic at a campus network—Measurements, models, and implications," *Comput. Netw.*, vol. 53, no. 4, pp. 501–514, Mar. 2009.

[25] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[26] M. Taghizadeh, K. Micinski, S. Biswas, C. Ofria, and E. Torng, "Distributed cooperative caching in social wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 6, pp. 1037–1053, Jun. 2013.

[27] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.

[28] H. Zhang, Z. Zhang, and H. Dai, "Gossip-based information spreading in mobile networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5918–5928, Nov. 2013.

[29] X. Wang, Q. Peng, and Y. Li, "Cooperation achieves optimal multicast capacity-delay scaling in MANET," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3023–3031, Oct. 2012.

[30] J. Hu, L.-L. Yang, and L. Hanzo, "Maximum average service rate and optimal queue scheduling of delay-constrained hybrid cognitive radio in Nakagami fading channels," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2220–2229, Jun. 2013.

[31] P. V. Mieghem, *Performance Analysis of Communications Networks and Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[32] M. Derakhshani and T. Le-Ngoc, "Aggregate interference and capacity-outage analysis in a cognitive radio network," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 196–207, Jan. 2012.

[33] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. New York, NY, USA: Elsevier/Academic, 2007.

[34] W. Gao, Q. Li, B. Zhao, and G. Cao, "Social-aware multicast in disruption-tolerant networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1553–1566, Oct. 2012.

[35] M. W. Fackrell, "Characterization of matrix-exponential distributions," Ph.D. dissertation, Faculty of Eng., Comput. Math. Sci., School of Appl. Math., Univ. Adelaide, Adelaide, South Australia, 5005, Australia, 2003.

[36] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Philadelphia, PA, USA: SIAM, 1999.

[37] Office of Communications. "Notice of proposed variation of Everything Everywhere's 1800 MHz spectrum licences to allow use of LTE and WiMAX technologies," Ofcom, Mar. 2012 [Online]. Available: http://stakeholders.ofcom.org.uk/consultations/variation-1800mhz-lte-wimax/summary.

[38] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, "Stochastic properties of the random waypoint mobility model," *Wireless Netw.*, vol. 10, no. 5, pp. 555–567, 2004.

[39] J. Hu, L.-L. Yang, and L. Hanzo, "Stochastic geometry in the cellular networks," Univ. Southampton, SO17 1BJ, U.K., Tech. Rep., 2015 [Online]. Available: http://eprints.soton.ac.uk/id/eprint/374932.

[40] O. Helgason, S. T. Kouyoumdjieva, and G. Karlsson, "Opportunistic communication and human mobility," *IEEE Trans. Mobile Comput.*, vol. 13, no. 7, pp. 1597–1610, Jul. 2014.

Jie Hu (S'11–M'16) received the B.Eng. degree in communication engineering and the M.Eng. degree in communication and information system from the School of Communication and Information Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2011, respectively, and the Ph.D. degree from the Southampton Wireless Group, University of Southampton, Southampton, U.K., in 2015.

Since March 2016, he has been working with the School of Communication and Information Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, as a Lecturer. He has a broad range of interests in wireless communication and networking, such as cognitive radio and cognitive networks, queuing analysis, resource allocation and scheduling, ad hoc wireless networks, and mobile social networks.

Lie-Liang Yang (M'98–SM'02–F'16) received the B.Eng. degree in communications engineering from Shanghai TieDao University, Shanghai, China, in 1988, and the M.Eng. and Ph.D. degrees in communications and electronics from Beijing (Northern) Jiaotong University, Beijing, China, in 1991 and 1997, respectively. From June 1997 to December 1997, he was a Visiting Scientist at the Institute of Radio Engineering and Electronics, Academy of Sciences of the Czech Republic, Prague, Czech Republic. Since December 1997, he has been with the University of Southampton, Southampton, U.K., where he is currently a Professor of Wireless Communications with the School of Electronics and Computer Science. His research interests include wireless communications, wireless networking, and signal processing for wireless communications. He has authored more than 300 research papers in journals and conference proceedings, authored or coauthored three books, and also published several book chapters.

Prof. Yang is a Fellow of The Institution of Engineering and Technology, U.K. He served as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the *Journal of Communications and Networks*, and is currently an Associate Editor of the IEEE ACCESS and the *Security and Communication Networks Journal*.

Lajos Hanzo (M'91–SM'92–F'03) received the degree in electronics, in 1976, the Doctorate degree, in 1983, and the D.Sc. degree. In 2009, he received the Honorary Doctorate "Doctor Honoris Causa" from the Technical University of Budapest, Budapest, Hungary. During his 35-year career in telecommunications, he has held various research and academic posts in Hungary, Germany, and the U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K., where he holds the Chair in telecommunications. He has successfully supervised 80 Ph.D. students, coauthored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10,000 pages, published over 1500 research entries at IEEE Xplore, acted both as TPC and General Chair of the IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently, he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) U.K., the European IST Programme and the Mobile Virtual Centre of Excellence (VCE), U.K. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. From 2008 to 2012, he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing, China. His research is funded by the European Research Council's Senior Research Fellow Grant. He is the Fellow of the REng, IET, and EURASIP.