This is the Pre-Published Version.

This article has been accepted for publication in a future issue of *Ad hoc & Sensor Wireless Networks*, 2016. Compared with the final manuscript, this is the latest author-reserved version, because content has been changed slightly.

Stochastic Analysis on the Deactivation-Controlled Epidemic Routing in DTNs with Multiple Sinks

Huawei Huang, Song Guo, Peng Li and Toshiaki Miyazaki
The University of Aizu, Japan

***Abstract*—To describe the delivery performance of message delivery in Delay Tolerant Networks (DTNs), a number of research efforts have been taken. However, the state-of-the-art analysis typically shares a common simplification that the pair-wise meeting rate between any two mobile nodes is exponentially distributed and supposed to be known in advance. In this paper, rather than depending on such assumption, we jointly consider the deactivation-rate of relay nodes and the number of sink nodes deployed in network as the primary tunable system parameters, aiming to study how such two parameters affect the performance of a message's delivery. Two groups of Ordinary Differential Equations (ODEs) based theoretical frameworks are proposed for replication based and Network Coding (NC) based epidemic routing protocols that can handle the small-sized and the large-sized message deliveries, respectively. Via extensive simulations, the accuracy of our analytical frameworks is verified.**

***Index Terms*—Deactivation Rate, Network Coding, Epidemic Routing, Message Delivery, DTN**

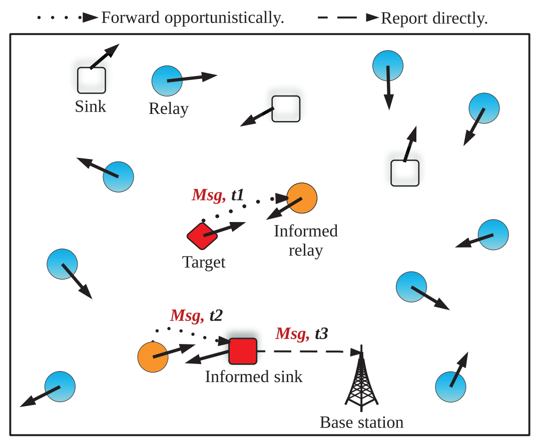


Fig. 1. The message dissemination using multiple-sink epidemic routing in a DTN, where $t_1 < t_2 < t_3$.

I. Introduction

The development of sensor network technology has enabled the implementation of target detection and monitoring [1] in a large scale environment, e.g., the tracking of mobile targets in wireless sensor networks [2], the detection of malicious intruders that invade the network domain [3], and the data collection for habitat and vital signs in wild animal sanctuaries [1], [4], [5]. Since the wireless sensors have the wireless communication capability, a Disruption/Delay Tolerant Network (DTN) can be constructed when these sensor-equipped animals or robots roam in the sanctuary. In such network, instead of continuous connection between two mobile objects, there is only intermittent connection when they opportunistically get into the transmission range of each other during their movements.

In this paper, we are primarily interested in data collection problem by considering both moving source-target (diamond-shaped node) and other mobile objects, namely the mobile relays (circle-shaped nodes) and multiple sinks (square-shaped nodes), in a DTN shown in Fig.1. In such a network, one message *Msg* is disseminated from the source target node to others periodically. Relay nodes help to forward copies of *Msg*. Once it is received by at least one sink node, it will be reported to the base station directly and immediately leveraging the long-distance communication capacity of sink node. Many practical applications in DTNs fall into such scenario, e.g., collection of the vital signs, habitat data and location information monitored by the sensor attached to the host animal in mobile sensor network, message dissemination in a battle field, and message forwarding in a pocket switched network in a campus.

Particularly, we are interested in the following two issues: 1) the evolution process of different types of mobile nodes during the message propagation, and 2) the relationship between delivery delay and the number of sinks deployed in the network. The evolution process means the constitution transformation of different types of mobile nodes as time goes by. Once we understand the evolution law of the message propagation, the deployment of mobile nodes can be controlled according to the theory. In the second issue, the delivery delay is defined as the duration from the generation instant of a message at the source target until its final reception instant by the first informed sink node. Without doubt low delivery delay is always preferred. The epidemic routing scheme has been proved as an efficient way to achieve minimum delivery delay and high delivery ratio [6]–[8] due to its greedy message forwarding manner, where a message is forwarded in a flooding style by fully exploring all communication opportunities such that it can be delivered to the sink as soon as possible.

Corresponding to the issues aforementioned, our objectives are: 1) establish the theoretical framework to describe the population evolution process of all types of mobile nodes; then 2) use the derived theoretical framework to find the connection between the number of sinks and delivery delay.

A. Motivation

Based on an extensive review on related literature, we note that most of the recent studies share a common simplification that the intermittent transmission opportunities can be described by an empirical exponential distribution of pairwise meeting interval between mobile objects [7], [9]–[12]. However, this assumption has the following limitations:

1) The pairwise inter-meeting time is often unknown in advance and its accurate value is hard to obtain.
2) Although it could be estimated under a certain mobility model, the analytical result is only applicable when the transmission range of a node is much smaller than the whole network region (Lemma 4 in [13]).
3) The analytical results can hardly be utilized by network operators for any optimization (e.g., minimizing delivery delay or energy consumption) because the inter-meeting rate between nodes cannot be controlled directly.

This motivates us to investigate the delivery performance of epidemic routing in DTNs as a function of parameters that can be controlled directly. One such parameter is the deactivation-rate over the relay nodes. Generally, message replication performed by pure epidemic routing paradigms incurs high storage overhead on wireless nodes [7] and very likely makes node run out of its buffer capacity. This is because each relay node potentially relays messages for many other targets simultaneously, and excessive copies of message generated from different source nodes can occupy too much of its buffer space [6], [8] if without any serving ability limitation. To tackle this problem, a reasonable method is to introduce a dedicated deactivation-rate [14], [15] parameter, which has a population control over the specified class of mobile nodes in network. Particularly, in this paper, we focus on the deactivation over relay nodes. If a relay node is deactivated, it stops relaying messages for other peers.

Now we describe how this parameter works over the specified nodes. In essential, the deactivation-rate is a probability within (0, 1) which plays an uniform role over each relay node. Every relay node maintains the same deactivation-rate in their memory. At any time slot during a message's propagation, if a relay node was selected by this probability, it enters into inactive status automatically. After that it will neither forward nor receive the message. Further, the message it has received shall be deleted to alleviate the buffer-space occupancy, as well as to be ready for forwarding packets towards other target messages. Therefore, it is significant to set an appropriate deactivation-rate over relay nodes.

Another parameter is the number of deployed sinks [15]–[18] in the network. It is believed that the more sinks are deployed, the lower delivery delay yields. However, it is a waste of expenditure if the network operator over-deploys sinks to collect the target message. Therefore, it is critical to understand the relationship between the number of sinks and the delivery performance. To this end, we propose the conventional ODEs based approach [7], [10], [11], [19] to construct the theoretical framework such that the delivery performance can be depicted accurately.

B. Major Contributions

The major contributions of this paper can be summarized as follows.

- To the best of our knowledge, we are the first to accurately analyze the delivery performance under general mobility model by viewing deactivation-rate, the number of deployed sinks as the primary system parameters. Particularly, time-varying infectious and recovery rates are carefully derived to capture the population evolutions of all different types of mobile nodes over the network.
- Two suites of ODE-based analytical frameworks are constructed for both replication based and Network Coding based epidemic routing protocols according to the size of the target message.
- The correlation between the number of sinks and the delivery delay is also investigated in a stochastic manner applying the proposed theoretical frameworks.
- The correctness and high accuracy of our analysis are verified by extensive simulations.

The rest of the paper is organized as follows. Section II summaries the related work. Section III introduces the system model. Sections IV and V detail our ODE-based stochastic analysis for small-sized and large-sized message dissemination, respectively. The theoretical findings are verified by experiments in Section VI. Finally, Section VII concludes this work.

II. Related Work

In this section, we analyze the differences between our proposal and existing approaches in the literature in terms of the following two aspects: *Epidemic Routing under Classical Assumption* and *Buffer Occupancy Control*.

A. Epidemic Routing under Classical Assumption

Various epidemic routing protocols and analytical models have been proposed in [7], [8], [10], [11], [19]–[23] in DTNs. We notice that the state-of-the-art work always assumes the pairwise inter-meeting rate or encounter probability between mobile nodes is known. For example, authors in [7], [11], [23], [24] assume that the inter-meeting time of any pair of nodes is an exponential random parameter with a given rate. However, as we have discussed, such empirical model is inaccurate and the rate is also difficult to be obtained in advance.

Under the classical assumption, a number of work has been conducted on the performance analysis of message dissemination in wireless mobile networks using epidemic routings based on ODE system. For example, Zhang et al. [7] develop a rigorous, unified framework to study the epidemic routing and its numerical variations. Then, Lin et al. [11] introduce an analytical system to study the delivery performance of epidemic routing using network coding in opportunistic networks. Zhuo et al. [24] study the traffic offloading for 3G network by building an incentive framework based on ODEs and Network Coding approach. The tradeoff between the amount of traffic being offloaded and user's satisfaction has been investigated. Zhang et al. [22] investigate the benefits of applying Random Linear Coding to unicast application in DTNs. The relation between the delivery delay and replication control has been well analyzed. Recently, Xu et al. [23] also develop an OED based analytical model to characterize social ties and user behaviors in Mobile Social Networks (MSNs).

Our study differs from the existing work mentioned above, because we novelly propose a stochastic theoretical framework, which considers the deactivation-rate over relay nodes and the number of sinks as the significant system parameters that can be controlled directly, rather than the empirical pair-wise inter-meeting rate. The developed analytical frameworks are able to well depict the complicated message propagation process under the deactivation-controlled epidemic routing protocols.

B. Buffer Occupancy Control

On the other hand, to avoid high buffer occupancy level of relay nodes while using epidemic routing, Feng et al. [8] propose an encounter count quota, using which packets can only be delivered to a certain number of relay nodes. Yao et al. [25] and Sugiyama et al. [26] use another controlled epidemic routing where each message is assigned a time-to-live residence time when caching in a relay node. In this way, messages can be deleted automatically after the residence time expires such that both the buffer occupancy and caching energy consumption can be reduced.

In contrast, we introduce another way to lower the high buffer occupancy, i.e., set a deactivation rate over all relay nodes when they are cooperating to forward packets for each target source node. Particularly, when the size of the target message is very large, we also study the deactivation-controlled epidemic routing using network coding technique. The buffer occupancy on wireless nodes can be further released significantly, because the disseminated messages are coded, rather than the native ones. Therefore, our analytical results obtained under the network coding framework can be extended to the scenario of big data oriented MSNs [27], aiming to analyze the delivery of mobile big data.

III. System Model

In this section, we introduce the system model used in this paper, which includes the basic network model, mobility model and some basic assumptions.

We consider a DTN with a number of mobile nodes that are independently distributed within a large scale of area. It should be noticed that, we mainly focus our study on the application layer by ignoring the consideration of effect of other layers, such as physical layer or MAC layer. Without loss of generality, we suppose that all mobile nodes own the uniform transmission range, denoted by r. Two mobile nodes can directly communicate with each other if and only if their Euclidean distance is no larger than r. Furthermore, we also assume that exactly one packet (native packet or coded packet) can be completely forwarded during one transmitting opportunity.

The Random Direction Mobility (RDM) [28] was created to overcome density waves in the average number of neighbors produced by the Random Waypoint Mobility Model [29]. It has been considered as a widely adopted random mobility model [5], [10], [13], [21], [25], [30]. In this model, any mobile node chooses its direction independently with an angle uniformly distributed within $[0, 2\pi)$, and then travels along that direction. Once a mobile node reaches the network border, it pauses for a specified duration and then randomly chooses its next angular “bounce” direction between $[0, \pi]$ and repeat the above process [29]. Note that, the moving speed of a node is uniform at any instant and the moving direction is randomly and independently distributed. Thus to any node, the possibility of locating at any position in any time instant is uniform in the network area. As a result, all mobile nodes can be uniformly and randomly distributed at any time.

TABLE I
NOTATIONS AND SYMBOLS

Notation	Description
N	the number of regular mobile relay nodes
K	the number of mobile sink nodes
$\mathcal{S}$	area of the whole network
μ	uniform deactivating rate over relay nodes
S-node	susceptible mobile node
I-node	infected mobile node
R-node	deactivated mobile node
Y-node	mobile relay node
K-node	mobile sink node
Φ-node	the infected Y-node
Θ-node	the infected K-node
X(t)	the number of X-nodes at time $t>0$ (X can be I, R, Y, K, Φ and Θ)
ρ_0	the density of all mobile nodes in network
$\rho_{X(t)}$	the density of X-nodes, X can be I, R, Y, K, Φ, Θ
C	the total number of innovative coded packets if decoding
$\Phi_c(t)$	the number of infected Y-nodes each holding c innovative coded packets at time $t>0$, $c=1,2,...,C$
$\Theta_c(t)$	the number of infected K-nodes each holding c innovative coded packets at time $t>0$, $c=1,2,...,C$
$\phi_c(t)$	the ratio of Φ_c-node, $c=1,2,...,C$
$\theta_c(t)$	the ratio of Θ_c-node, $c=1,2,...,C$
S_Λ	the total infectious area of all I-nodes
D	the delivery delay sensed by the first informed sink node
$F(t)$	the CDF of D in replication based epidemic routing at $t>0$
$\widetilde{D}$	the delivery delay at the first decoding sink node
$\widetilde{F}(t)$	the CDF of D in NC based epidemic routing at $t>0$

The other notations and symbols used in this paper are summarized in Table I.

IV. Replication-based Dissemination Towards Small-Sized Message

In order to monitor the target in DTNs, firstly we need to know the evolution principle that all types of mobile nodes obey. Then, it is important to make it clear how the focused parameters, namely the deactivation-rate over relay nodes and the number of sinks, affect the delivery delay experienced by the first informed sink. Once we address the two problems, the parameters of the network can be set appropriately according to the quality of service. In this section, we apply stochastic analysis method to investigate them respectively.

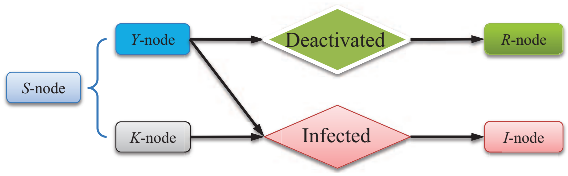


Fig. 2. The replication based *IRYK* epidemic routing protocol.

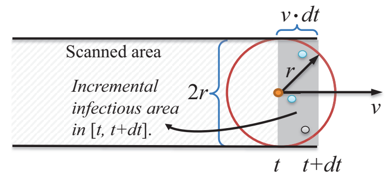


Fig. 3. The scanned area and infectious area of an I-node in its trajectory.

A. Replication based Epidemic Routing Protocol

In the focused DTN, the message generated in a source target node is disseminated to the entire network, and received by a sink at last. Initially, the target node creates a message and therefore this source node can be viewed as the first infectious node (I-node). The message can be received by any sink node (K-node) and any relay node (Y-node). Hence, all the K-nodes and Y-nodes can be viewed as susceptible nodes (S-node), which is illustrated by Fig.2. Once an S-node meets an I-node, it will receive a copy of the message, and become a new I-node, which will also be able to infect other susceptible nodes thereafter. An I-node always remains in infectious state. To receive the copy of the target message as soon as possible, we control all K-nodes keeping active. In other words, they can be always ready to receive the message. To the Y-nodes, to address the flooding problem because of excessive epidemic routing, we introduce a parameter μ, which denotes the deactivation-rate over all relay nodes serving a specified target node. Since the deactivation-rate is a probability within the range (0,1) that acts uniformly over all relay nodes, in any time slot, each Y-node can transform its state as a recovered node (R-node) with a probability μ. An R-node will neither forward nor receive the message any longer. Moreover, to relieve buffer occupancy, the message copy it has been received shall be removed. Therefore, we can see that this type of controlled epidemic routing is a variation of the famous SIR epidemic routing model [7], [19], [23]. For short, we call it the *IRYK* model.

Suppose that there are one target node, N-1 relay nodes and K sink nodes in the network initially. Let $I(t)$ $(i(t))$, $R(t)$ $(r(t))$, $Y(t)$ $(y(t))$ and $K(t)$ $(k(t))$ denote the number (ratio) of infectious nodes, recovered nodes, relay nodes and the sink nodes at time t, respectively. Apparently, we have $i(t) = \frac{I(t)}{N+K}$, $r(t) = \frac{R(t)}{N+K}$, $y(t) = \frac{Y(t)}{N+K}$ and $k(t) = \frac{K(t)}{N+K}$. Particularly, the equation $i(t) + r(t) + y(t) + k(t) \equiv 1, t \geq 0$ always holds in the message propagation.

B. Stochastic Analysis of Message Dissemination Procedure

We set the system time when a message is generated at the source target node as 0 (i.e., t=0). Since then, the source target starts *infecting* other mobile nodes opportunistically during its travel. An infected mobile node is then able to infect other susceptible nodes. On the other hand, due to the introduction of deactivation-rate to Y-nodes, the population of relay node declines gradually. At time $t = 0$, since all mobile nodes are Y-nodes and K-nodes except that the source target node is the only I-node, we have the initial condition:

$$\begin{cases} I(0) = 1, \\ R(0) = 0, \\ Y(0) = N - 1, \\ K(0) = K. \end{cases} \tag{1}$$

Thereafter, whenever a Y-node or K-node moves into the transmission area of any I-node, it shall be infected. We call such transmission area as *infectious area* in the *IRYK* epidemic routing model. The transition rate of $I(t)$ therefore can be written as:

$$\frac{dI(t)}{dt} = \frac{\rho_{Y(t)} \cdot S_\Lambda}{dt} + \frac{\rho_{K(t)} \cdot S_\Lambda}{dt},\ t \geq 0, \tag{2}$$

where the first term in the right hand side equation denotes the increasing differential contribution from Y-nodes, while the second represents the increasing differential contribution from K-nodes. We define $\rho_{Y(t)}$ and $\rho_{K(t)}$ as the density of Y-nodes and K-nodes, respectively, during the message propagation process. We further use S_Λ to denote the overall *incremental* infectious area of I-nodes. They can be derived as the follows.

$$\rho_{Y(t)} = \frac{Y(t) \cdot \rho_0}{N+K},\ t \geq 0, \tag{3}$$

$$\rho_{K(t)} = \frac{K(t) \cdot \rho_0}{N+K},\ t \geq 0, \tag{4}$$

$$S_\Lambda = I(t) \cdot d_S = I(t) \cdot 2rv \cdot dt,\ t \geq 0, \tag{5}$$

where $\rho_0 = (N+K)/\mathcal{S}$ denotes the density of all mobile nodes in DTN, d_S is the incremental infectious area of one I-node during the differential interval $[t, t+dt]$ as highlighted by the dark shadowed area in Fig. 3. This area is scanned by the mobile node with transmission diameter $2r$ with moving velocity v. Therefore, the *overall infectious area* of I-nodes is equal to the summation over d_S of each I-node, as calculated in (5).

As we know, any Y-node shall recover to an R-node once it is *selected* by the deactivation-rate in each time slot. Thus, we have the recovering rate during the message propagation process:

$$\frac{dR(t)}{dt} = \mu \cdot Y(t),\ t \geq 0. \tag{6}$$

From the infecting rate and recovering rate aforementioned, the transition rate of relay nodes includes the decreasing differential portions transforming to I-nodes and R-nodes, simultaneously. Since sink nodes always keep active, they shall only transform to I-nodes once they receive the message.

Therefore, the transition rates of Y-nodes and K-nodes can be derived as follows.

$$\begin{cases} \dfrac{dY(t)}{dt} = -\dfrac{\rho_{Y(t)} \cdot S_\Lambda}{dt} - \mu \cdot Y(t), \\ \dfrac{dK(t)}{dt} = -\dfrac{\rho_{K(t)} \cdot S_\Lambda}{dt}, \; t \geq 0. \end{cases} \tag{7}$$

C. Delivery Delay

Due to the unexpected dynamics of DTNs, it is difficult to obtain the exact delivery delay of a message to mobile nodes. In this subsection, we shall derive the delivery delay distribution by a stochastic analysis.

Let D denote the delivery delay experienced by the first informed sink node, and $F(t) = \Pr(D \leq t)$ be the Cumulative Distribution Function (CDF) of delivery delays. In order to obtain $F(t)$, we first need to calculate the probability that a K-node meets an I-node during a differential interval $[t, t + \triangle t], \triangle t > 0$.

Lemma 1. *Let $N_I(t) = \rho_0 \cdot I(t) \cdot d_S$ indicate the expected number of mobile nodes locating inside the overall incremental infectious areas, and P_0 denote the probability that a sink node meets an I-node during $[t, t + \triangle t]$. Then we have*

$$P_0 = \frac{\binom{K}{1} \cdot \binom{N_I(t)}{1}}{\binom{N+K}{1}}, \; t \geq 0. \tag{8}$$

Proof. The probability of a K-node meets an I-node in $[t, t + \triangle t]$ equals the probability that a K-node locates in the incremental infectious area of any I-node during $[t, t + \triangle t]$. Then, P_0 is equal to the probability of selecting one mobile node from the network such that this node is within $N_I(t)$ as well as it is a K-node. Apparently, the number of all combinations that this selected mobile node is within $N_I(t)$ and is one K-node simultaneously equals to $\binom{K}{1} \cdot \binom{N_I(t)}{1}$. Furthermore, the number of cases that selecting one mobile node from the network is $\binom{N+K}{1}$. Therefore, the probability of a sink node meets an I-node in $\triangle t$ is $\frac{\binom{K}{1} \cdot \binom{N_I(t)}{1}}{\binom{N+K}{1}}$. □

Then secondly, we continue to derive the transition rate of $F(t)$ during the differential interval aforementioned.

Lemma 2. *For any time instant $t \geq 0$, during a differential interval $\triangle t$, we have $\triangle F(t) = (2rv\rho_0 \cdot \frac{I(t) \cdot K}{N+K} \cdot \triangle t) \cdot (1 - F(t))$.*

Proof.

$$\begin{aligned} &\triangle F(t) = F(t + \triangle t) - F(t) \\ =&\Pr\{t < D \leq (t + \triangle t)\} \\ =&\Pr\{\text{a } K\text{-node meets an } I\text{-node during } \triangle t\} \cdot \Pr\{D > t\} \\ =&P_0 \cdot \Pr\{D > t\} \\ =&\frac{\binom{K}{1} \cdot \binom{N_I(t)}{1}}{\binom{N+K}{1}} \cdot (1 - F(t)) \\ =&(2rv\rho_0 \cdot \frac{I(t) \cdot K}{N+K} \cdot \triangle t) \cdot (1 - F(t)). \end{aligned}$$

□

Finally, letting $\triangle t \to 0$, we can get the following differential equation on $F(t)$:

$$\frac{dF(t)}{dt} = (2rv\rho_0 \cdot \frac{I(t) \cdot K}{N+K}) \cdot (1 - F(t)), \; t \geq 0. \tag{9}$$

Therefore, $F(t)$ can be solved in conjunction with the aforementioned ODEs (2), (6) and (7) with the initial condition $F(0) = 0$.

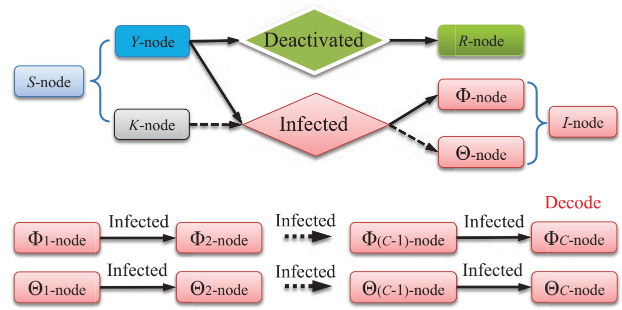


Fig. 4. The Network Coding based tunable *IRYK* epidemic routing protocol.

V. NC-based Dissemination Towards Large-sized Message

In this section, we consider the target message with a large size that it cannot be forwarded completely within one transmission opportunity. In such a scenario, Network Coding based approach [31] shows superior performance over the duplication-based one [22], [24]. Accordingly, we develop another set of analytical framework aiming to describe the dissemination of encoded packets.

A. Network Coding based Epidemic Routing Protocol

We consider a large-sized message partitioned into C native packets, each with a size that can be forwarded in a transmission opportunity. A coded packet p is a linear combination of the C native packets P_1, P_2, ..., P_C in the form: $p = \sum_{i=1}^{C} \alpha_i P_i$, where α_i are coding coefficients.

The NC based routing protocol is illustrated in Fig. 4. Suppose node a holds x $(1 \leq x \leq C)$ coded packets in buffer. When node a meets node b, it generates a coded packet p_a as:

$$p_a = \sum_{i=1}^{x} \beta_i p_i, \tag{10}$$

where β_i is randomly chosen from a Galois field, and then forwards p_a and the coding coefficients to node b. Finally, node b inserts the received coded packet into its buffer.

A received coded packet is said innovative if the rank of coefficient matrix at the receiver can be increased after its reception. Such matrix corresponds to all coded packets in the buffer. Once the rank reaches C at any sink node, all original C native packets can be recovered.

Let Φ_c-node and Θ_c-node denote the infected Y-node and K-node holding c innovative coded packets, respectively. Then, all the I-nodes can be classified into $2\times C$ types: Φ_c-node and Θ_c-node, $c=1,2,...,C$, as shown in Fig. 4. Letting $\widetilde{D}$ be the time at which any Θ_{C-1}-node receives an innovative coded packet at time $\widetilde{D}$, i.e., it can decode all C native packets, we say $\widetilde{D}$ is the delivery delay of the original large-sized message.

In the following, we conduct the stochastic analysis of NC-based message delivery under the classic approximation [11], [32]: *an I-node can transmit an innovative coded packet to another undecoded I-node with high probability*. This is because, in the case of abundant buffers, Deb et al. [33] have shown that the probability that a coded packet is useful to another node is 1-1/q, where q is the size of the Galois Field to generate random coding coefficients. In practice, q is usually sufficiently large such that the probability is very close to 1.

B. Stochastic Analysis of Message Dissemination Procedure

Recalling that, there are Y-nodes, K-nodes and only one Φ_1-node at time $t=0$, we have the initial condition:

$$\begin{cases} Y(0)=N-1, \\ K(0)=K, \\ R(0)=0, \\ \Phi_1(0)=1, \\ \Phi_x(0)=0,\ x=2,...C \\ \Theta_x(0)=0,\ x=1,2,...C. \end{cases} \tag{11}$$

The equations about $R'(t)$, $Y'(t)$, $K'(t)$ are the same as in the last section. In contrast, the construction of $I(t)$ is in a different way shown as follows.

$$\begin{cases} \Phi(t)=\sum_{c=1}^{C}\Phi_c(t),\ t\geq 0. \\ \Theta(t)=\sum_{c=1}^{C}\Theta_c(t),\ t\geq 0. \\ I(t)=\Phi(t)+\Theta(t),\ t\geq 0. \end{cases} \tag{12}$$

Particularly, in NC-based epidemic routing, the evolution of population of Φ_c-node, $c=1,2,...,C$, can be characterized as:

$$\begin{cases} \dfrac{d\Phi_1(t)}{dt} =\dfrac{S_\Lambda\cdot\rho_{Y(t)}-S_\Lambda\cdot\rho_{\Phi_1(t)}}{dt} \\ \qquad\quad =\dfrac{I(t)\cdot(\rho_{Y(t)}-\rho_{\Phi_1(t)})\cdot d_S}{dt} \\ \dfrac{d\Phi_j(t)}{dt} =\dfrac{I(t)\cdot(\rho_{\Phi_{j-1}(t)}-\rho_{\Phi_j(t)})\cdot d_S}{dt}, \\ \qquad\quad j=2,...,C-1 \\ \dfrac{d\Phi_C(t)}{dt} =\dfrac{I(t)\cdot\rho_{\Phi_{C-1}(t)}\cdot d_S}{dt},\ t\geq 0. \end{cases} \tag{13}$$

For each $\frac{d\Phi_j(t)}{dt}$, it constitutes two parts: the positive rate representing the transition from Φ_{j-1} to Φ_j nodes and the negative rate representing the transition from Φ_j to Φ_{j+1} nodes. Note that as special cases, we have $\Phi_0(t)=Y(t)$ at the beginning and no transitions from Φ_C nodes eventually.

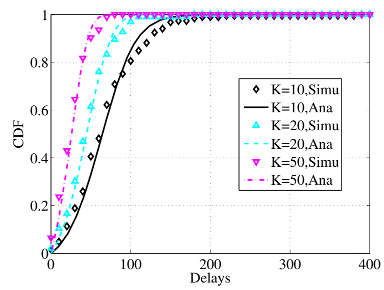


Fig. 7. CDF of delivery delays experienced by the first informed sink v.s. number of sink, while using the replication based epidemic routing.

In a similar manner, we have the equations for $\Theta_j'(t)$ as follows.

$$\begin{cases} \dfrac{d\Theta_1(t)}{dt} =\dfrac{S_\Lambda\cdot\rho_{K(t)}-S_\Lambda\cdot\rho_{\Theta_1(t)}}{dt} \\ \qquad\quad =\dfrac{I(t)\cdot(\rho_{K(t)}-\rho_{\Theta_1(t)})\cdot d_S}{dt} \\ \dfrac{d\Theta_j(t)}{dt} =\dfrac{I(t)\cdot(\rho_{\Theta_{j-1}(t)}-\rho_{\Theta_j(t)})\cdot d_S}{dt}, \\ \qquad\quad j=2,3,...,C-1 \\ \dfrac{d\Theta_C(t)}{dt} =\dfrac{I(t)\cdot\rho_{\Theta_{C-1}(t)}\cdot d_S}{dt},\ t\geq 0. \end{cases} \tag{14}$$

C. Delivery delay

Now we present the analysis of delivery delay distribution under the NC-based epidemic routing. Firstly, we define $\widetilde{F}(t)=\Pr(\widetilde{D}\leq t)$ as the CDF of delivery delays in the NC-based epidemic routing. Following the similar derivation as given in Lemma 1 and 2, we have the following results.

Lemma 3. *Letting $\widetilde{P}_0$ denote the probability that a Θ_{C-1}-node meets an I-node during $[t,t+\triangle t]$, we have*

$$\widetilde{P}_0=\frac{\binom{\Theta_{C-1}}{1}\cdot\binom{N_I(t)}{1}}{\binom{N+K}{1}},\ t\geq 0. \tag{15}$$

Lemma 4. *For any time instant $t\geq 0$, during a differential interval $\triangle t$, we have $\triangle\widetilde{F}(t)=(2rv\rho_0\cdot\frac{I(t)\cdot\Theta_{C-1}(t)}{N+K}\cdot\triangle t)\cdot(1-\widetilde{F}(t))$.*

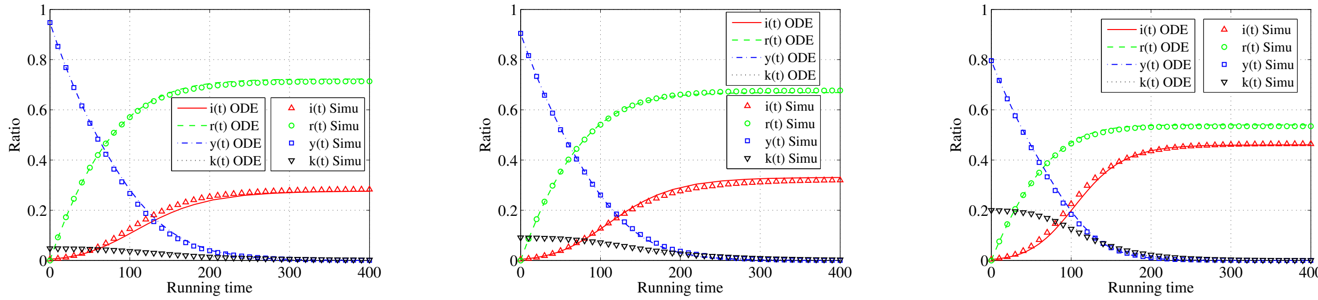
Then, letting $\triangle t\to 0$, we acquire the differential equation of $\widetilde{F}(t)$:

$$\frac{d\widetilde{F}(t)}{dt}=(2rv\rho_0\cdot\frac{I(t)\cdot\Theta_{C-1}(t)}{N+K})\cdot(1-\widetilde{F}(t)),\ t\geq 0. \tag{16}$$

Finally, $\widetilde{F}(t)$ can be solved in conjunction with the aforementioned ODEs (2), (6), (7), (12), (13) and (14) with the initial conditions (11) and $\widetilde{F}(0)=0$.

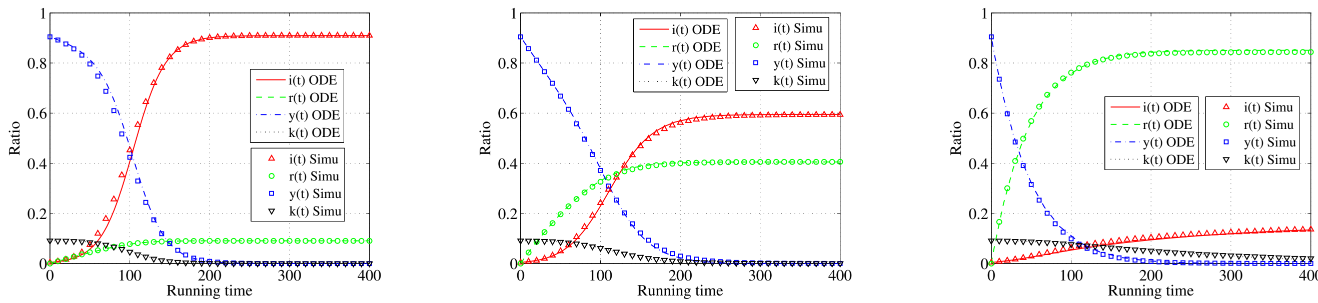
D. Buffer Occupancy of Coded Packets

We also study the buffer occupancy of coded packets in the NC-based epidemic routing. It is defined as the total amount of buffer space occupied by the coded packets over all the active infected nodes. According to the analysis of the packets



(a) μ=0.01, K=10 (b) μ=0.01, K=20 (c) μ=0.01, K=50

Fig. 5. Results of $i(t)$, $r(t)$, $y(t)$ and $k(t)$ v.s. number of sink, while using the replication based epidemic routing.



(a) K=20, μ=0.001 (b) K=20, μ=0.005 (c) K=20, μ=0.02

Fig. 6. Results of $i(t)$, $r(t)$, $y(t)$ and $k(t)$ v.s. deactivation rates, while using the replication based epidemic routing.

transmission procedure, buffer occupancy is intensively related to the number of infected nodes and C. By letting $B(t), t > 0$ denote the total buffer occupancy over all the infected nodes, we have:

$$B(t) = \sum_{c=1}^{C}(c \cdot (\Phi_c(t) + \Theta_c(t))),\ c = 1, ..., C,\ t \geq 0. \quad (17)$$

VI. Performance Evaluation

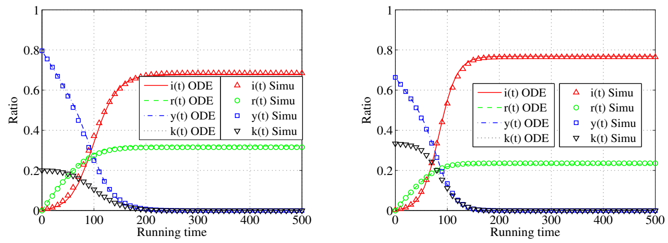
A. Basic Settings

We have developed a DTN simulator with C++ that strictly follows the system model described in Section III. The RDM mobility model and the controlled epidemic routing protocols are also implemented in the simulator to validate the accuracy of our analysis. In the simulations, mobile nodes are first randomly deployed sparsely in a $500m \times 500m$ square region. We set N and K to 200 and $\{10,\ 20,\ 50\}$, respectively, to evaluate the delivery performance on different network densities. After the initial deployment, nodes start moving according to RDM mobility model. The velocity v is set as 10 m/s. The uniform transmission range of mobile nodes is fixed to 3 meters. For each simulation setting, 200 instances are conducted to obtain the average value. Besides the simulation results, we also build the corresponding ODEs and obtain the ODE-based analytical results applying the numerical ODE45 solving tool provided by Matlab.

B. Results of Replication-based Messages Dissemination

To investigate the accuracy of our analysis on message propagation process, we firstly conduct a suite of simulations under RDM mobility model with settings μ=0.01 and $K \in \{10, 20, 50\}$. The evaluation results under variation settings of parameter K are presented in Fig. 5, which depicts the evolution of $i(t)$, $r(t)$, $y(t)$ and $k(t)$. Obviously, our analysis shows high accuracy for all cases. There are several other observations we can find. In the first, with the same deactivation rate μ=0.01, $y(t)$ shows as a deceasing function over system time in all three cases. Secondly, $i(t)$ converges faster and its converged peak value also becomes higher with the increasing K. Such a phenomenon can be attributed to that sink nodes will never be deactivated like relays, and with more sinks in the network, the transmitting opportunity becomes higher. As a result, the message is delivered faster and more relays and sinks become I-nodes. Simultaneously, fewer relays transform to R-nodes. Correspondingly, the peak value of $r(t)$ degenerates lower as the figures show.

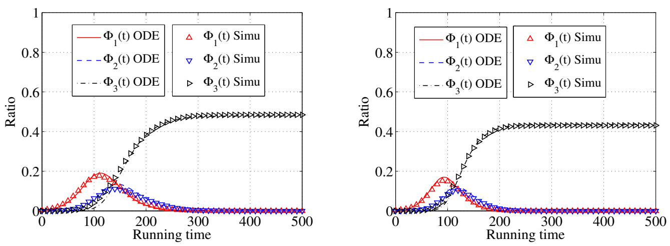
Next, we study the delivery process under different μ. Fig. 6 shows the evaluation results under the settings $K = 20$, $\mu \in \{0.001, 0.005, 0.02\}$. From all these figures, clearly we see that $y(t)$ falls down sharper with the increasing μ. Correspondingly, more relays are deactivated to be R-nodes with a faster speed, and the converged peak value of $r(t)$ grows higher. This is because the transmitting opportunity decreases following the decreasing population of relays in the network. As a result, the message is delivered more slowly.



(a) μ=0.005, C=3, K=50

(b) μ=0.005, C=3, K=100

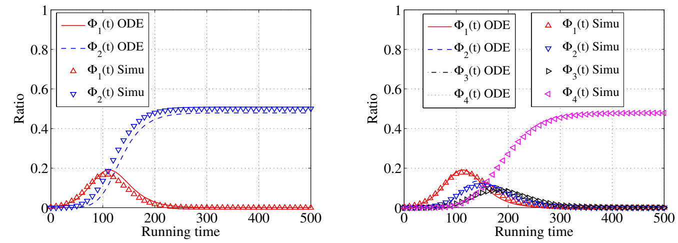
Fig. 8. Results of $i(t)$, $r(t)$, $y(t)$ and $k(t)$ v.s. number of sink, while using NC-based epidemic routing.



(a) μ=0.005, C=3, K=50

(b) μ=0.005, C=3, K=100

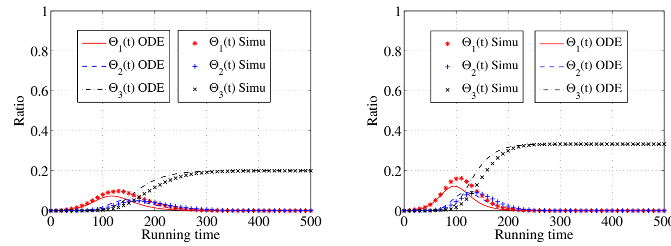
Fig. 9. Ratios of infected relays v.s. number of sink, while using NC-based epidemic routing.



(a) μ=0.005, K=50, C=2

(b) μ=0.005, K=50, C=4

Fig. 10. Ratios of infected relays v.s. size of target message, while using NC-based epidemic routing.



(a) μ=0.005, C=3, K=50

(b) μ=0.005, C=3, K=100

Fig. 11. Ratios of infected sinks v.s. number of sink, while using NC-based epidemic routing.

Due to the same reason, the number of I-nodes in the network grows slower, too. Furthermore, more relays are deactivated to become R-nodes before being infected potentially. That is why the peak value of $i(t)$ becomes lower as Fig. 6 shows.

Overall, from both Fig. 5 and 6, we can always observe that the analysis on $i(t)$, $r(t)$, $y(t)$, and $k(t)$ matches the simulation results quite well. This validates the high accuracy of the proposed ODEs based analysis for the single-message *IRYK*-like propagation process in DTNs.

We also apply the same evaluation method to investigate the accuracy of the analysis on delivery delay experienced by the first sink. The CDF of delivery delays is also obtained through 200 instances for each simulation setting.

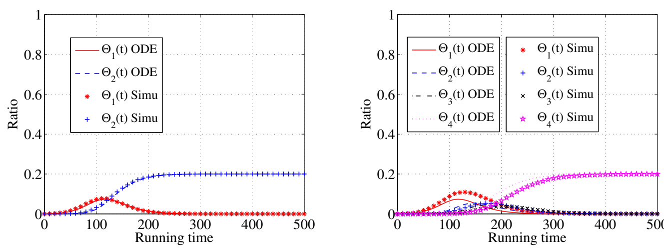
Fig. 7 shows the CDF of analytical and simulation delivery delays under various values of $K \in \{10, 20, 50\}$ when $r = 3m$, $v = 10m/s$ and $N = 200$, respectively. It is observed that the expected delivery delays of sink becomes smaller while K grows. For example in Fig. 7, there are 90% of the delivery delays occur at t=50 when K=50. However, these values drop to 50% and 30% when K=20 and 10, respectively. It is because the more sinks exist in the network, the higher transmitting opportunity becomes. In other words, the message can be delivered faster.

C. Results of NC-based Message Dissemination

Fig. 8 shows the delivery procedure in terms of 4 categories of nodes while the target message composed of C=3 native packets is forwarded under settings μ=0.005 and K={50, 100}. On the other hand, under various numbers of sinks $K \in$ {50, 100} and various size of the target message are plotted in Fig. 9 and Fig. 10, respectively. We have the following observations. (1) As shown in Fig. 9, where K increases from 50 to 100, the curve of Φ_3 converges faster. (2) As shown in Fig. 10, where C grows from 2 to 4, the converging speed of Φ_C is decreasing. This is because more sinks on shorter message can accelerate the dissemination procedure. (3) Both figures show that the theoretical results based on our analytical model match the simulation results very well. The similar observation on the evolutions of infected sinks, i.e., the Θ-nodes, can be also made as shown in Fig. 11 and Fig. 12.

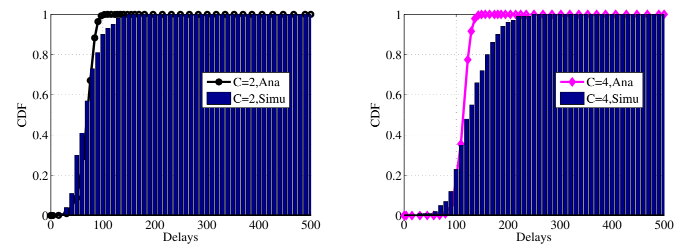
Next, we study the performance of delivery delays under the NC-based epidemic routing shown in Fig. 13 and Fig. 14 by varying K and C, respectively. From the former figure, we can see that the decoding delays of Θ_{C-1}-node becomes lower while increasing number of sinks. It is because the receiving opportunity increases with more sinks deployed in network, making the Θ_{C-1}-node transform to Θ_C-node earlier. For example, we can observe from Fig. 13(a) and Fig. 13(b) , there are 40% delivery delays lower than 100 seconds when K=50, but the proportion increases to 60% when K=100. The latter figure indicates that the decoding delay becomes longer while increasing C, i.e., the number of native packets of the target message. For instance, as shown in Fig. 14(a) and Fig. 14(b), 90% delivery delays are lower than 100 seconds when C=2, while the proportion is only 10% when C=4. This observation suggests that the target message is encouraged be encoded with a smaller C only if the opportunistic transmission capacity is preserved.

Finally, we study the buffer occupancy. By varying $K \in \{10, 20, 50, 100\}$ and fixing μ=0.005, C=3, Fig. 15(a) shows that the buffer occupancy is an increasing function over the



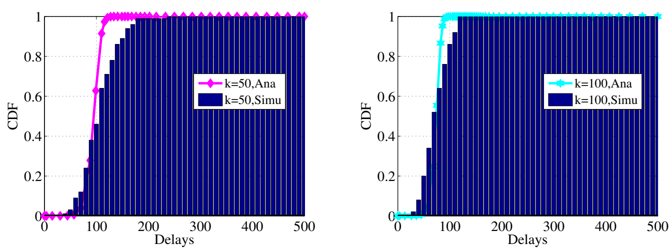
(a) μ=0.005, K=50, C=2 (b) μ=0.005, K=50, C=4

Fig. 12. Ratios of infected sinks v.s. size of target message, while using NC-based epidemic routing.



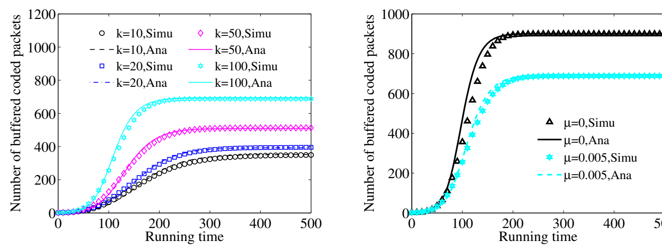
(a) μ=0.005, K=50, C=2 (b) μ=0.005, K=50, C=4

Fig. 14. CDF of delays, i.e., $\widetilde{F}(t)$, v.s. size of target message, while using NC-based epidemic routing.



(a) μ=0.005, C=3, K=50 (b) μ=0.005, C=3, K=100

Fig. 13. CDF of delays, i.e., $\widetilde{F}(t)$, v.s. number of sink, while using NC-based epidemic routing.



(a) μ=0.005, $K \in \{10, 20, 50, 100\}$ (b) K=100, $\mu \in \{0, 0.005\}$

Fig. 15. Buffer occupancy v.s. deactivation rates and number of sink, while fixing C=3 and using NC-based epidemic routing.

number of sinks. A good match of analytical and simulation results is also observed. We also testify the correlation between deactivation-rate and the buffer occupancy shown as Fig. 15(b). When the parameter μ is set as 0, it indicates that there is no deactivation operation over relays. Both simulation and analytical results show that the buffer space can be saved by approximate 25% when N=200, K=100 and μ=0.005. This also testifies that the deactivation control over relays is an efficient approach to lower the buffer occupancy while forwarding the target message using epidemic routing protocol.

VII. Conclusion

In this paper, we investigate the message dissemination performance in DTNs by introducing deactivation-rate over relay nodes and leveraging multiple sink nodes. The message propagation is modeled as an SIR-like routing procedure, which is able to characterize the evolution of different categories of mobile nodes over the network. Considering the size of the target message, we propose two suites of ODE-based analytical frameworks towards the replication based and the Network Coding based epidemic routings. A notable feature of our analytical frameworks is that the deactivation-rate over relay nodes and the number of sinks are tunable system parameters. Furthermore, we derive the correlations between the delivery delay as well as the buffer occupancy ratio and the tunable system parameters, i.e., deactivation-rate and the number of sinks. The correctness of our theoretical frameworks are also verified via extensive experiments.

References

[1] J. Krause, S. Krause, R. Arlinghaus, I. Psorakis, S. Roberts, and C. Rutz, "Reality mining of animal social systems," *Trends in ecology & evolution*, vol. 28, no. 9, pp. 541–551, 2013.

[2] E. Xu, Z. Ding, and S. Dasgupta, "Target tracking and mobile sensor navigation in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 177–186, 2013.

[3] Y. Wang, W. Fu, and D. P. Agrawal, "Gaussian versus uniform distribution for intrusion detection in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 2, pp. 342–355, 2013.

[4] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. ACM, 2002, pp. 88–97.

[5] G. Keung, B. Li, Q. Zhang, and H.-D. Yang, "The target tracking in mobile sensor networks," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2011, pp. 1–5.

[6] A. Vahdat, D. Becker *et al.*, "Epidemic routing for partially connected ad hoc networks," Technical Report CS-200006, Duke University, Tech. Rep., 2000.

[7] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, vol. 51, no. 10, pp. 2867–2891, 2007.

[8] Z. Feng and K.-W. Chin, "On the performance of epidemic based routing protocols for delivering multicast bundles in delay tolerant networks," *International Journal of Wireless Information Networks*, vol. 21, no. 2, pp. 133–153, 2014.

[9] M. Abdulla and R. Simon, "Controlled Epidemic Routing for Multicasting in Delay Tolerant Networks," in *IEEE International Symposium on Modeling, Analysis and Simulation of Computers and Telecommunication Systems (MASCOTS '08)*. IEEE, 2008, pp. 1–10.

[10] Z. Zhang, G. Mao, and B. Anderson, "On the information propagation in mobile ad-hoc networks using epidemic routing," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference (GLOBE-COM)*, 2011, pp. 1–6.

[11] Y. Lin, B. Li, and B. Liang, "Stochastic analysis of network coding in epidemic routing," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 794–808, june 2008.

[12] M. Khouzani, S. Eshghi, S. Sarkar, N. Shroff, and S. Venkatesh, "Optimal energy-aware epidemic routing in DTNs," in *Proceedings of the 13th ACM international symposium on Mobile Ad Hoc Networking and Computing (Mobihoc '12)*. ACM, 2012, pp. 175–182.

[13] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, no. 1, pp. 210–228, 2005.

[14] E. Altman, A. P. Azad, T. Basar, and F. De Pellegrini, "Optimal activation and transmission control in delay tolerant networks," in *Proceedings of the 29th Annual Conference of the IEEE Computer and Communications (INFOCOM '10)*. IEEE, 2010, pp. 1–5.

[15] M. Di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: A survey," *ACM Transactions on Sensor Networks (TOSN)*, vol. 8, no. 1, p. 7, 2011.

[16] A. Bari, D. Teng, R. Ahmed, and A. Jaekel, "Relay node placement with energy and buffer constraints in wireless sensor networks using mobile data collector," *International Journal of Sensor Networks*, vol. 8, no. 3, pp. 147–159, 2010.

[17] W. Alsalih, H. Hassanein, and S. Akl, "Placement of multiple mobile data collectors in wireless sensor networks," *Ad Hoc Networks*, vol. 8, no. 4, pp. 378–390, 2010.

[18] M. Zhao, M. Ma, and Y. Yang, "Efficient data gathering with mobile collectors and space-division multiple access technique in wireless sensor networks," *IEEE Transactions on Computers*, vol. 60, no. 3, pp. 400–417, 2011.

[19] T. Britton, "Stochastic epidemic models: A survey," *Mathematical Biosciences*, vol. 225, no. 1, pp. 24–35, 2010.

[20] Q. Guan, F. R. Yu, S. Jiang, and G. Wei, "Prediction-based topology control and routing in cognitive radio mobile ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4443–4452, 2010.

[21] M. T. Islam, M. Akon, A. Abdrabou, and X. S. Shen, "Modeling epidemic data diffusion for wireless mobile networks," *Wireless Communications and Mobile Computing*, 2012.

[22] X. Zhang, G. Neglia, J. Kurose, D. Towsley, and H. Wang, "Benefits of network coding for unicast application in disruption-tolerant networks," *IEEE/ACM Transactions on Networking*, vol. 21, no. 5, pp. 1407–1420, 2013.

[23] Q. Xu, Z. Su, K. Zhang, P. Ren, and X. S. Shen, "Epidemic information dissemination in mobile social networks with opportunistic links," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 3, pp. 399–409, 2015.

[24] X. Zhuo, W. Gao, G. Cao, and Y. Dai, "Win-coupon: An incentive framework for 3g traffic offloading," in *2011 19th IEEE International Conference on Network Protocols (ICNP)*. IEEE, 2011, pp. 206–215.

[25] H. Yao, H. Huang, D. Zeng, B. Li, and S. Guo, "An energy-aware deadline-constrained message delivery in delay-tolerant networks," *Wireless Networks*, pp. 1–13, 2014.

[26] K. Sugiyama, T. Kubo, A. Tagami, and T. Hasegawa, "Mean free path applied to message dissemination in opportunistic networks," in *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2015, pp. 420–425.

[27] Z. Su, Q. Xu, and Q. Qi, "Big data in mobile social networks: a qoe-oriented framework," *IEEE Network*, vol. 30, no. 1, pp. 52–57, 2016.

[28] E. M. Royer, P. M. Melliar-Smith, and L. E. Moser, "An analysis of the optimum node density for ad hoc mobile networks," in *IEEE International Conference on Communications, (ICC)*, vol. 3. IEEE, 2001, pp. 857–861.

[29] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless communications and mobile computing*, vol. 2, no. 5, pp. 483–502, 2002.

[30] Z. Zhang, G. Mao, and B. Anderson, "On information dissemination in infrastructure-based mobile ad-hoc networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '12)*. IEEE, 2012, pp. 1743–1748.

[31] L. Wang, Y. Yang, W. Zhao, L. Xu, and S. Lan, "Multi-rate network coding for energy-efficient multicast in heterogeneous wireless multi-hop networks," *Ad Hoc & Sensor Wireless Networks (AHSWN)*, vol. 32, no. 3-4, pp. 197–219, 2016.

[32] Y. Lin, B. Liang, and B. Li, "Performance modeling of network coding in epidemic routing," in *Proceedings of the 1st international MobiSys workshop on Mobile opportunistic networking*. ACM, 2007, pp. 67–74.

[33] S. Deb, M. Médard, and C. Choute, "Algebraic gossip: A network coding approach to optimal multiple rumor mongering," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2486–2507, 2006.