

Research Article

A Data Dissemination Method Based on Region Type Correlations for Mobile Opportunistic Networks

Linfeng Liu ^{1,2} and Daoliang Chen¹

¹Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

²School of Computer and Information, Hohai University, Nanjing 211100, China

Correspondence should be addressed to Linfeng Liu; liulinfeng@gmail.com

Received 12 May 2018; Revised 8 September 2018; Accepted 30 September 2018; Published 15 October 2018

Academic Editor: Carles Gomez

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

In Mobile Opportunistic Networks (MONs), due to the node movements and the uncontrollable on/off switches of the carried communication devices, the contacts between nodes may be scarce and momentary, and thus a data packet should be transferred through some discrete hops. To avoid the costly flooding of data packets, the data packets are typically disseminated to some relay nodes selected by data holders. However, the mobility patterns of nodes will become different in different types of regions (such as residential regions, commercial regions, scenery regions, or industrial regions); i.e., the movement directions and movement ranges of nodes are frequently varied when the nodes move among various regions. At present, the issues regarding the region types and region type correlations have not been investigated for the data dissemination in existing works. To this end, we propose a Region Type based Data Dissemination Method (RTDDM) for MONs, which exploits the region type correlations and selects the proper relay nodes through a Markov decision model. To verify the performance of RTDDM, we give some theoretical analysis as well as an elaborated simulation study, the results of which show that RTDDM can improve the delivery ratio and reduce the delivery delay, especially in the applications with various region types.

1. Introduction

Mobile Opportunistic Network (MON) [1, 2] is one of the emerging communication paradigms in wireless mobile communications. An MON is defined as a mobile network where the communications between nodes are challenged by sporadic and intermittent contacts, as well as the frequent disconnections and reconnections, and thus a stable communication path from the source to the destination usually cannot be obtained. The MON applications are diverse, such as the vehicle networks for traffic information sharing, the mobile sensor networks for wildlife tracking, and the pocket switched networks comprised of human-carried mobile devices.

In most traditional routing methods, the end-to-end paths must be established, and then the data packets are disseminated along these paths. Therefore, the global connectivity of network topology should be satisfied, such that at

least a stable end-to-end communication path can be found from the network topology. However, the MON nodes always move frequently, and the communication devices carried by nodes are switched between on/off statuses unpredictably. These characteristics of MONs make the end-to-end paths extremely difficult to be established, and hence the traditional routing methods cannot be applied to MONs. In MONs, the node contacts must be exploited for the data dissemination; i.e., the data packets are disseminated from the *data holders* (the nodes carrying data packets) to the encountered *ordinary nodes* (the nodes not carrying data packets) and finally are transferred to the destination node through discrete hops.

Therefore, the held data packets are allowed to be disseminated to some relay nodes by the data holders. Due to the unpredictability in future movements of nodes, the data dissemination method (especially how to select the proper relay nodes) becomes a vital issue for the

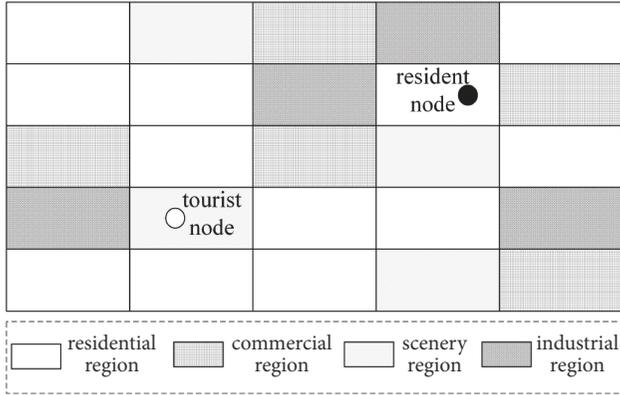


FIGURE 1: Example regions of different types.

improvement of delivery ratio and the reduction of delivery delay.

The node contacts are concerned with the region types. In many MON applications, the nodes are likely to appear in some regions with special types more frequently or stay in these regions for a longer period. For example, the tourists visit some scenery regions more frequently, and the residents are more likely to stay in some residential regions, as shown in Figure 1. These features can help improve the data dissemination in MONs; i.e., the future contacts between nodes can be predicted according to the region types, and the proper relay nodes can be selected to make the data packets be delivered to the destination nodes as soon as possible. At present, the issues regarding the region types and the region type correlations have not been paid enough attention for the data dissemination of MONs. Note that the mobility patterns of nodes will be different in the different types of regions, and the movement directions and movement ranges of nodes are frequently varied when the nodes move among various regions. Thus, the data dissemination can be improved through differentiating the region types and investigating the region type correlations.

Our research is also inspired by the following observations: (a) the movements of communication devices in MONs are controlled by their carriers (human beings), and the mobility of human beings is driven by their habits which are typically stable in long terms. Therefore, the movement directions and movement ranges in different types of regions (such as residential regions, commercial regions, scenery regions, or industrial regions) are quite different; (b) the randomness in node mobility makes the data dissemination problem be mapped into a sequential decision problem mathematically. Besides, the Markov decision method is a preferable decision method and is suitable for the sequence decision problem [3, 4].

This work is an extension of our previous conference paper [5]. In this paper, we improve the Region Type based Data Dissemination Method (RTDDM), especially in terms of the action determinations and message interactions, and some theoretical analysis is conducted for RTDDM. Besides, more simulation results are also provided to further clarify

the merits of RTDDM. The remainder of this paper is organized as follows. Section 2 discusses related studies. Section 3 describes the network model. Section 4 presents our algorithm. Section 5 provides some theoretical analysis regarding RTDDM. The simulation study is reported in Section 6. This paper is concluded by Section 7.

2. Related Work

Recently, some relevant research has been conducted on the data dissemination in MONs. Mcgeehan *et al.* propose a message routing approach ChitChat [6], which exploits users' direct and transient social interests via discriminatory gossiping to penetrate messages deeper into the network. ChitChat enables message carriers to make opportunistic and distributed routing decisions based on the likelihood of forwarding the messages to the destination. In [7], a geographic-based spray-and-relay (GSaR) routing scheme is proposed. By estimating the movement range of the destination via the historical geographic information, GSaR expedites the message being sprayed toward this range and postpones that out of this range. Reference [8] gives two sociality based routings, named as the contact based routing and community aware two-hop routing, respectively. These routings are designed based on the graph properties (such as the average path length and modularity) and the node encounter process. Cai *et al.* construct a cross-layer distributed opportunistic routing protocol [9], which considers the channel sensing strategy, the forwarder selection for each secondary user, and the package division scheme on each link.

To overcome the communication problems of changing topology and missing infrastructure, [10] presents a lightweight and scalable routing protocol that exploits the limited social circle of each node. In [11], the Community-Based Forwarding (CBF) considering the community interactions is developed. CBF is able to explore the intermediate connections between clusters to route the messages with more balanced node participation and higher levels of reliability. To improve the dissemination efficiency, [12] presents a routing algorithm called Popular Node Gateway Protocol (PNGP). PNGP is validated through simulations, the results of which indicate that it performs satisfactorily in terms of delivery ratio, delivery delay, and transmission cost. In [13], an analytical model of message dissemination by applying the concept of mean free path is proposed, and each node contact is represented as a collision between two particles. In order to relieve the congestion without affecting the delivery delay, Ciobanu *et al.* present an interest-based dissemination algorithm where nodes tend to group together based on interests [14]. Reference [15] proposes a stochastic approach for modeling the message dissemination in opportunistic networks to choose the appropriate relay nodes according to the features of the application scenario.

In this paper, the issues of region type correlations and region weights are specially exploited and formulated, and then a Markov decision model is applied to determine the relay nodes of data holders. Thus, the data packets could be continuously propagated to the proper relay nodes to improve the delivery ratio and reduce the delivery delay.

3. Network Model

The time is divided into discrete time slots with an equal length τ_s . Each data packet needs to be disseminated from the source node to the destination region during t^* time slots.

3.1. Nodes. Suppose there are N mobile nodes in the network, denoted by $1, 2, \dots, i, \dots, N$, and each node is with the same communication range r_c . The coordinate of node i at the t -th time slot is denoted by $C(i)^{(t)}$. The distance between node i and node j at the t -th time slot is marked as $d(i, j)^{(t)}$. If $d(i, j)^{(t)} \leq r_c$, then i and j are neighbors.

3.2. Regions. The network area Ω is divided into K regions $R_1, R_2, \dots, R_k, \dots, R_K$. Each region R_k is denoted by a triple (x_k, y_k, T_k) , where (x_k, y_k) is the center coordinate of R_k , and T_k denotes the region type of R_k . Suppose there are M different region types, such as residential regions, business regions, industrial regions, and educational regions. The distance between two regions R_k and $R_{k'}$ is calculated as

$$|R_k - R_{k'}| = \sqrt{(x_k - x_{k'})^2 + (y_k - y_{k'})^2}. \quad (1)$$

The human mobility in MONs is analyzed by a real-world dataset (Dartmouth Colleges wireless local area network (WLAN) trajectories [16]) which are recorded as the connections or disconnections of nodes with some access points (APs). Figure 2 illustrates that the nodes usually visit a few APs frequently, and note that the average probability of visiting these APs is even larger than 0.7. The results inspire us to put forward the notion *home regions* for nodes; i.e., the nodes will appear in the home regions with very large probabilities. The home regions of node i are marked as a region set $\mathbf{H}(i)$, where the probability of i appearing in each region belonging to the set $\mathbf{H}(i)$ is larger than a preset threshold p_h .

Besides, the distance between $\mathbf{H}(i)$ and $\mathbf{H}(j)$ is denoted by D_{ij} , and D_{ij} is computed as the minimum distance between $\mathbf{H}(i)$ and $\mathbf{H}(j)$,

$$D_{ij} = |\mathbf{H}(i) - \mathbf{H}(j)| = \min |R_k - R_{k'}|, \quad (2)$$

$$\forall R_k \in \mathbf{H}(i), R_{k'} \in \mathbf{H}(j)$$

3.3. Region Type Correlation and Region Weight. The nodes are prone to appear in the correlated regions. The correlation of different region types is calculated based on the movement trajectories of nodes. With regard to each region R_k , the correlation between R_k and $R_{k'}$ is computed as

$$Cor(R_k, R_{k'}) = \begin{cases} 1, & \text{if } T_k = T_{k'}, \\ \frac{\sum_{1 \leq i \leq N} B(i, R_k, R_{k'})}{N}, & \text{otherwise,} \end{cases} \quad (3)$$

where $B(i, R_k, R_{k'})$ is set as 1 if there are at least two regions belonging to $\mathbf{H}(i)$ simultaneously; otherwise, $B(i, R_k, R_{k'})$ is set as 0.

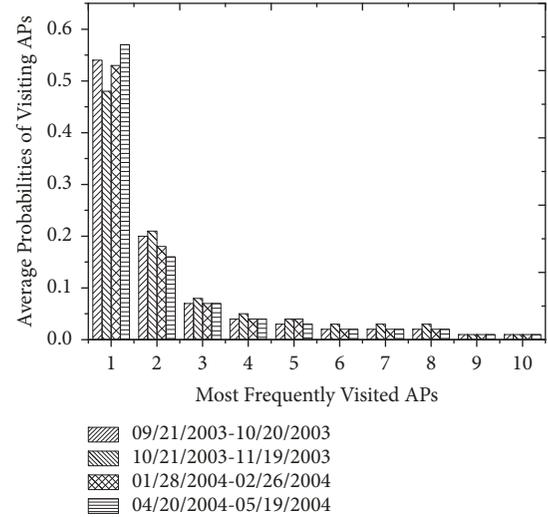


FIGURE 2: Average probabilities of visiting APs. The movement trajectories during the periods 09/21/2003-10/20/2003, 10/21/2003-11/19/2003, 01/28/2004-02/26/2004, and 04/20/2004-05/19/2004 are analyzed. Ten most frequently visited APs are selected and sorted for each node, and then the probabilities of nodes visiting these APs are averaged and given in this figure.

The region weight of R_k is denoted by W_k , where $W_k \in [0, 1]$. The value of W_k indicates the possibility of data packets being disseminated from R_k to the destination region R_d . W_k is expressed as

$$W_k = Cor(R_k, R_d)^\alpha \cdot \left\{ \frac{|R_k - R_d|}{\max_{1 \leq k' \leq K} |R_{k'} - R_d|} \right\}^\gamma, \quad (4)$$

where α and γ are preset exponents. The value of α and γ reflects the influences of region type and region distance on the region weight, respectively.

At the t -th time slot, the weight of j receiving and storing the data packets is expressed as

$$w^{(t)}(j) = \frac{\sum_{\forall R_k \in \mathbf{H}(j)} [P(j, R_k) \cdot W_k]}{\sum_{\forall R_k \in \mathbf{H}(j)} P(j, R_k)}. \quad (5)$$

4. Data Dissemination Method

In this section, we propose a Region Type based Data Dissemination Method (RTDDM) for MONs. In particular, in RTDDM, the Markov decision model is used to determine the relay nodes whose home regions are with larger region weights.

4.1. Dissemination Process Based on Markov Decision. The dissemination process based on Markov decision is described as follows:

(1) Decision stages: $\mathcal{T} = \{1, \dots, F\}$, which indicates that the node encounters F neighbors during a time slot.

(2) State set: $\mathcal{S} = \{0, 1, \Delta\}$ is a finite set of states, where "1" indicates that the current encountered node is the best relay node and "0" indicates that the current encountered

node is not the best relay node. Δ indicates the termination of dissemination at current time slot; i.e., the data holder has already encountered F nodes or the data packets have been disseminated to the encountered nodes.

(3) Action set: the action set $\mathcal{A}(s)$ is defined as

$$\mathcal{A}(s) = \begin{cases} \{R, E\}, & \text{if } s \in \{0, 1\}, \\ R, & \text{if } s = \Delta, \end{cases} \quad (6)$$

where s denotes the current state. E is the action of the current encountered node storing the data packets, and R is the action of the current encountered node rejecting the data packets.

(4) Reward function: when the number of nodes is large enough, the nodes are considered to be evenly distributed approximatively. At the t -th time slot, the obtained reward is expressed as

$$r_m^{(t)}(s, a) = \begin{cases} 0, & \text{if } s \in \{0, 1\} \text{ and } a = R, \\ w^{(t)}(h) \cdot \frac{m}{F}, & \text{if } s \in \{0, 1\}, a = E \text{ and } m < F, \\ 0, & \text{if } s = \Delta, \end{cases} \quad (7)$$

where m/F is the probability of the best relay node falling into the set of the first m encountered nodes and h denotes the m -th encountered node.

(5) Transition probability: the transition probability is denoted by $P_m^{(t)}(s' | s, a)$, where s' and s denote the future

state and the current state, respectively. The expression of $P_m^{(t)}(s' | s, a)$ is written as

$$P_m^{(t)}(s' | s, a) = \begin{cases} \frac{1}{m+1}, & \text{if } s \in \{0, 1\}, a = R \text{ and } s' = 1, \\ \frac{m}{m+1}, & \text{if } s \in \{0, 1\}, a = R \text{ and } s' = 0, \\ 1, & \text{if } s \in \{0, 1\}, a = E \text{ and } s' = \Delta, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

According to (8), the transition probability is equal to $1/(m+1)$ when the $(m+1)$ -th encountered node is by far the best relay node.

Then, we give the solution process for the above problem. At the t -th time slot, the maximum probability of the best relay node having been selected during the period from the current decision stage to the end stage is denoted by $u_m^{(t)}(1)$; i.e., the m -th encountered node is the best one in the previous encountered nodes. Likewise, from the next decision stage to the end stage, the maximum probability of selecting the best relay node is denoted by $u_m^{(t)}(0)$; i.e., the m -th encountered node is not the best one in the previous encountered nodes. $u_F^{(t)}(1)$ and $u_F^{(t)}(0)$ should satisfy the following equations:

$$\begin{aligned} u_F^{(t)}(1) &= 1, \\ u_F^{(t)}(0) &= 0, \\ u_F^{(t)}(\Delta) &= 0. \end{aligned} \quad (9)$$

For $m = 1, 2, \dots, F-1$, $u_m^{(t)}(1)$ and $u_m^{(t)}(0)$ can be written as

$$\begin{aligned} u_m^{(t)}(1) &= \max \left\{ r_m^{(t)}(1, E) + u_{m+1}^{(t)}(\Delta), r_{m+1}^{(t)}(1, E) \cdot \left[P_m^{(t)}(1 | 1, R) \cdot u_{m+1}^{(t)}(1) + P_m^{(t)}(0 | 1, R) \cdot u_{m+1}^{(t)}(0) \right] \right\} \\ &= \max \left\{ r_m^{(t)}(1, E), r_{m+1}^{(t)}(1, E) \cdot \left[\frac{u_{m+1}^{(t)}(1)}{m+1} + \frac{m \cdot u_{m+1}^{(t)}(0)}{m+1} \right] \right\}, \end{aligned} \quad (10)$$

$$\begin{aligned} u_m^{(t)}(0) &= \max \left\{ r_m^{(t)}(0, E) + u_{m+1}^{(t)}(\Delta), r_{m+1}^{(t)}(0, E) \cdot \left[P_m^{(t)}(1 | 0, R) \cdot u_{m+1}^{(t)}(1) + P_m^{(t)}(0 | 0, R) \cdot u_{m+1}^{(t)}(0) \right] \right\} \\ &= \max \left\{ r_m^{(t)}(0, E), r_{m+1}^{(t)}(0, E) \cdot \left[\frac{u_{m+1}^{(t)}(1)}{m+1} + \frac{m \cdot u_{m+1}^{(t)}(0)}{m+1} \right] \right\}. \end{aligned} \quad (11)$$

In particular, there is $u_m^{(t)}(\Delta) = 0$. Thus, the optimal actions can be obtained by solving (10) and (11).

4.2. RTDDM. In RTDDM, to restrict the dissemination cost, the maximum number of disseminated copies at each time slot is set as N_d ($N_d \leq F$). At each time slot, the relay nodes will be selected through a Markov decision model as described in Section 4.1. In particular, in RTDDM each data holder makes the decision about whether to disseminate the held data packets to the neighbors through several decision

stages. In each decision stage, each data holder disseminates the held data packets to at most N_d neighbors, a part of which will become the relay nodes. Thus, the decision stages will be repeated until N_d relay nodes have been selected or all the neighbors have received the data packets. The detailed description of RTDDM is given as follows.

Step 1. According to the previous trajectories of nodes, each node determines its home regions. The region type correlations and the region weights are calculated.

Type	Time	Self-Information	Holder	Message Set	Receiver
<i>inquire_msg</i>	t	i	$C(i)^{(t)}$	Yes	$data_list(i)$ Neighbors
<i>reply_msg</i>	t	i	$C(j)^{(t)}$	No	$data_list(j)$ Neighbors
<i>data_msg</i>	t	i	$C(i)^{(t)}$	Yes	NULL Selected Relay Nodes

FIGURE 3: Message structures.

Step 2. At each time slot (suppose the t -th time slot), each node i broadcasts an inquire message *inquire_msg*, as shown in Figure 3, where $data_list(i)$ denotes the list of held data packets by i at the current time slot.

Step 3. After the reception of *inquire_msg* (as shown in Figure 3) from i , each neighboring node j responds with a *reply_msg*. If the destination node has received an *inquire_msg* from i , it will send the data packets to destination node immediately; otherwise, Step 4 is carried out.

Step 4. Each data holder disseminates the held data packets *data_msgs* to N_d nearest nodes. With regard to the receivers of the data packets, if the current state of the m -th neighbor (denoted by s) is equal to 1, and there is $w^{(t)}(h) \cdot m/F \geq u_m^{(t)}(0)$, then the action E will be taken; otherwise, if s is equal to 0 or there is $w^{(t)}(h) \cdot m/F < u_m^{(t)}(0)$, then the action R will be taken. As mentioned above, the action $a(s)$ is determined by the following expression:

$$a(s) = \begin{cases} E, & \text{if } s = 1 \text{ and } w^{(t)}(h) \cdot \frac{m}{F} \geq u_m^{(t)}(0) \\ R, & \text{if } s = 0 \text{ or } w^{(t)}(h) \cdot \frac{m}{F} < u_m^{(t)}(0). \end{cases} \quad (12)$$

Step 5. If k neighbors have chosen the action E at the last decision stage, then m and N_d should be updated as

$$\begin{aligned} m &\leftarrow m + N_d, \\ N_d &\leftarrow N_d - k. \end{aligned} \quad (13)$$

After that, if m and N_d satisfy that $m < F$ and $N_d > 0$, Steps 4 and 5 will be repeated; if F and N_d satisfy that $m = F$ or $N_d = 0$, the dissemination process at the current time slot is terminated.

Step 6. The above steps will be repeated until the data packets have been disseminated to the destination region, or the period of t^* time slots has been expired. If the data packets have been delivered to destination region before the deadline (t^* time slots), an announcement message will be propagated to all data holders to stop the further dissemination. The

size of announcement message is very small and can be piggybacked with other messages.

The pseudocode of RTDDM is depicted in Algorithm 1.

5. Theoretical Analysis

In this section, we give a theoretical analysis for RTDDM.

5.1. RTDDM Complexity. In Step 2, each node broadcasts an *inquire_msg*; therefore the total message amount will reach N . In Step 3, there will be $N^2 \cdot (\pi \cdot r_c^2)/\Omega$ *reply_msgs* to be transmitted. Steps 4 and 5 make each data holder disseminate at least N_d copies of data packets, and thus the number of transmitted messages is $O(N)$. As a result, the total number of transmitted messages of RTDDM is at most $O(N^2)$.

5.2. Expected Delivery Ratio and Expected Delivery Delay. With regard to each data packet, the total number of copies during t^* time slots is expressed as $(N_d + 1)^{t^*}$. Similar to [17], we apply the continuous-time Markov model to analyze RTDDM, as shown in Figure 4, where $Q = (N_d + 1)^{t^*}$. The state n indicates that there are n data holders in network, and the state O indicates that the data packets have been delivered to the destination region. $p(n | n-1)$ denotes the probability of the transition from state $n-1$ to state n .

When there are n data holders, the probability of the data packet being disseminated to the neighbors is expressed as $((N-n) \cdot N_d)/(N \cdot F)$. The average contact rate between nodes is marked as β which can be computed from the historical trajectories. Hence, $p(O | n)$ is expressed as $(\beta \cdot \sum_{i=1}^n n w^{(t)}(i))/N$, and $p(n | n-1)$ is expressed as

$$p(n | n-1) = \begin{cases} n \cdot \frac{(N-n) \cdot N_d}{N \cdot F} \cdot \beta, & \text{if } 1 \leq n < Q, \\ 0, & \text{if } n = Q. \end{cases} \quad (14)$$

Let $F_n(t)$ and $F_o(t)$ be the probability distribution of delivery delay of the state n and the state O , respectively. According to Kolmogorov equation [18], we obtain the first-order derivative of $F_n(t)$ as

$$F_n'(t) = \begin{cases} -[p(n | n-1) + p(O | n)] \cdot F_n(t), & \text{if } n = 1, \\ -[p(n | n-1) + p(O | n)] \cdot F_n(t) + p(n-1 | n-2) \cdot F_{n-1}(t), & \text{if } 2 \leq n \leq Q-1, \\ -p(O | n) \cdot F_n(t) + p(n-1 | n-2) \cdot F_{n-1}(t), & \text{if } n = Q, \\ \sum_{q=1}^Q p(O | q) \cdot F_q(t), & \text{if } n = O. \end{cases} \quad (15)$$

```

1:  $t \leftarrow 1$ .
2: while  $t \leq t^*$ , and data packets have not been delivered do
3:   for each node  $1 \leq i \leq N$  do
4:      $i$  broadcasts an inquire_msg.
5:     Each receiver responds a reply_msg.
6:   end for
7:   Neighbor sets of nodes are updated.
8:   for each data holder  $i$  do
9:     while  $m < F$  and  $N_d > 0$  do
10:       $i$  disseminates the held data packets to  $N_d$  neighbors;
11:      for each neighbor having received the data packets do
12:         $a(s)$  is calculated;
13:        The received data packets are stored or rejected according to  $a(s)$ ;
14:      end for
15:       $k$  ( $k \leq N_d$ ) neighbors become the relay nodes.
16:       $m \leftarrow m + N_d$ ,  $N_d \leftarrow N_d - k$ .
17:    end while
18:  end for
19:   $t \leftarrow t + 1$ .
20: end while

```

ALGORITHM 1: RTDDM pseudo-code.

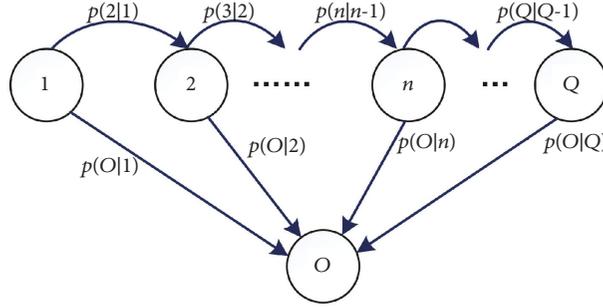


FIGURE 4: Continuous-time Markov model.

Because there are $F_1(0) = 1$ and $F_n(0) = F_o(0) = 0$, we can further obtain the probability density function of delivery delay through Laplace transform and inverse Laplace transformation:

$$f_o(t) = \sum_{n=1}^Q p(Q|n) \otimes_{q=1}^n p(q-1|q-2) \cdot e^{-[p(q|q-1) - p(O|q)]t}, \quad (16)$$

where \otimes denotes the convolution operation. Therefore, the expected delivery delay is expressed as

$$\mathbb{E}(\mathcal{D}_t) = \int_0^{\infty} t \cdot f_o(t) dt, \quad (17)$$

and the expected delivery ratio is expressed as

$$\mathbb{E}(\mathcal{R}_t) = \int_0^{t^*} f_o(t) dt. \quad (18)$$

The theoretical delivery delay and theoretical delivery ratio are calculated from (17) and (18) to compare against the simulation results of RTDDM in the next section.

6. Simulations

In this section, we present a thorough performance evaluation of our proposed RTDDM, along with a comparison study with the binary spray [19] (labeled by BINARY), BUBBLE Rap [20] (labeled by BUBBLE), and GSaR [7]. The performance metrics include the delivery ratio and the delivery delay. We develop a simulator using C++ language, and the mechanisms of overhearing and collisions (including the messages interactions for synchronization and backoff) have been realized in our simulator. Besides, the simulation results are averaged over 500 runs. The main parameter settings are shown in Table 1.

Before the simulations, we collect the historical trajectories of node movements during the first 40 time slots, so that the average contact rate between nodes can be calculated, and the home regions of nodes can be determined. We compare

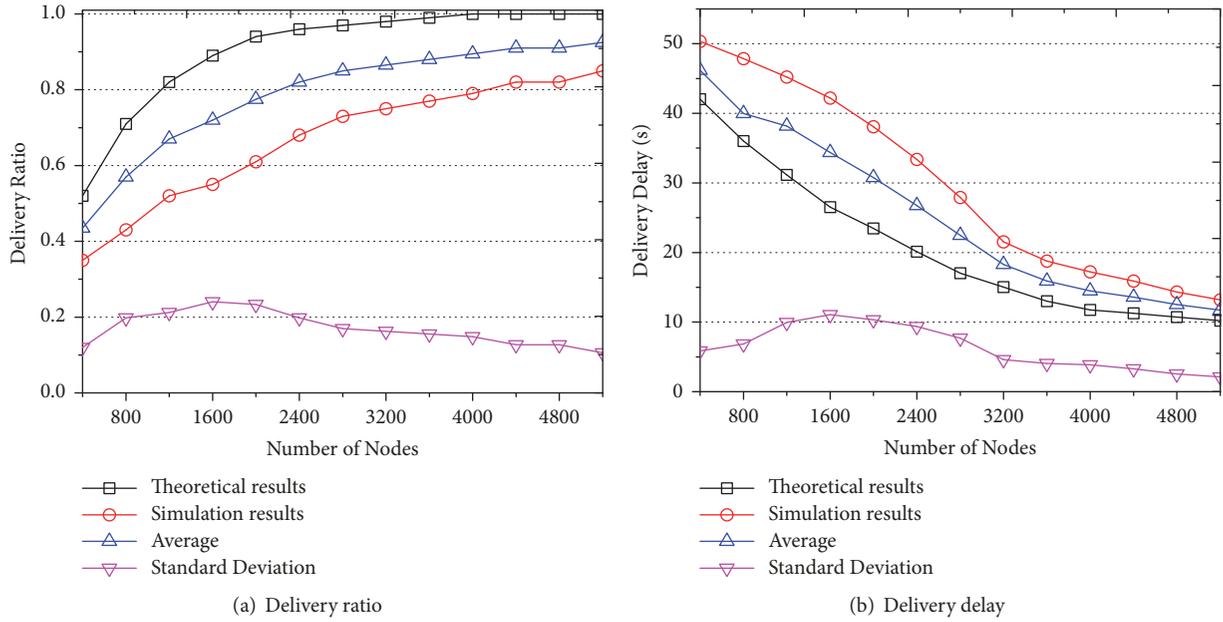


FIGURE 5: Theoretical results versus simulation results of RTDDM.

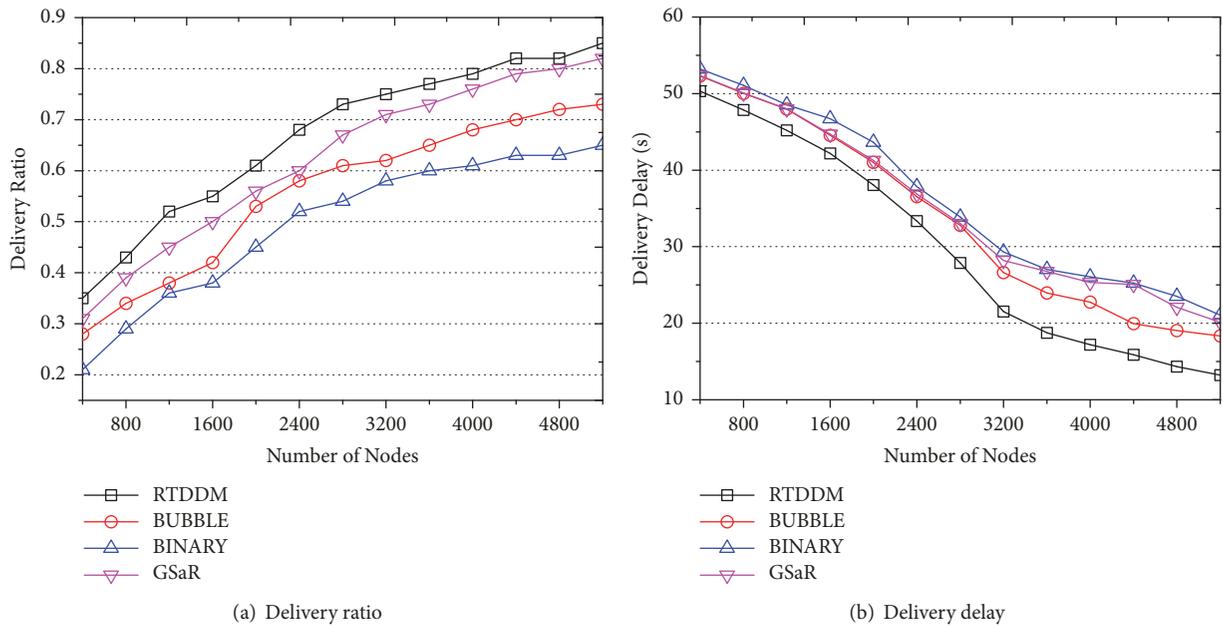
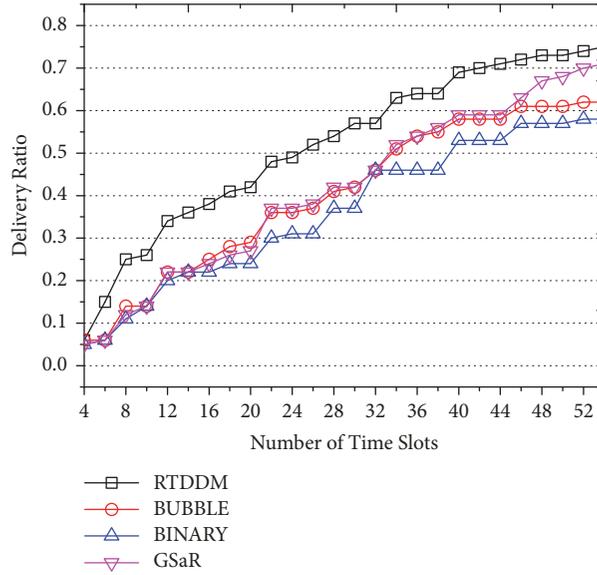
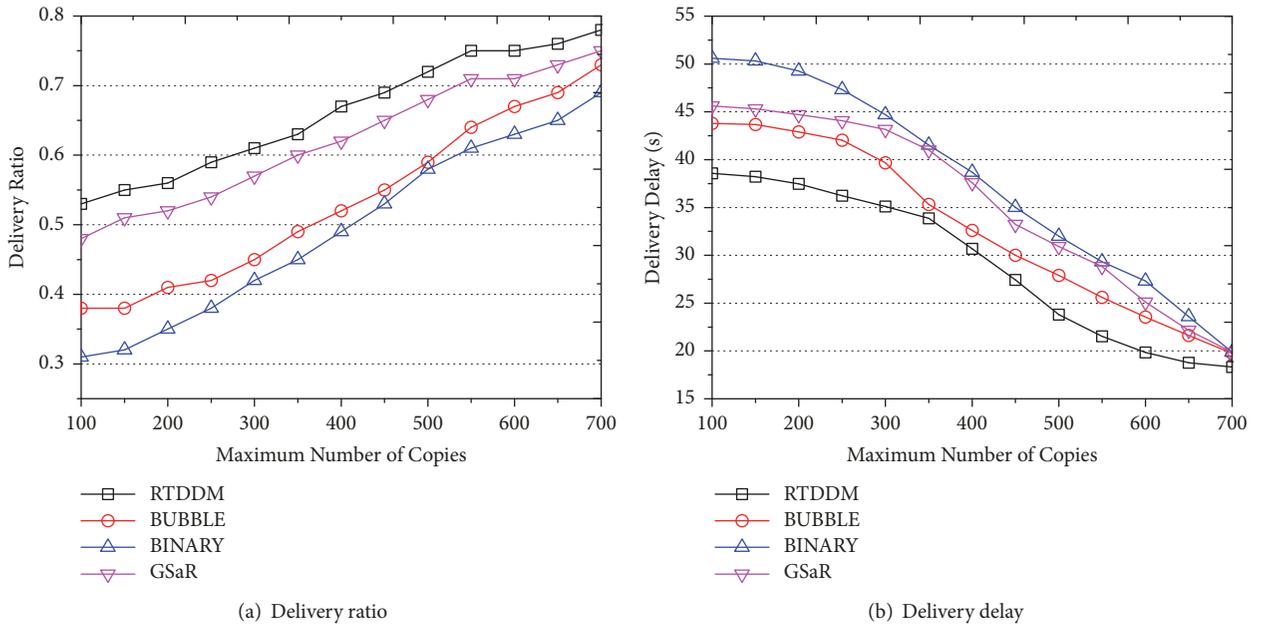


FIGURE 6: Impacts of the number of nodes.

the simulation results of RTDDM with theoretical results which are calculated by (17) and (18). From Figure 5, we observe that the overall trends of the two kinds of results are consistent, and the theoretical results are superior to the simulation results of RTDDM. This is because the theoretical results are always the optimal results which are obtained from centralized computations. Note that the standard deviation curves drop gradually with the increase of the number of nodes, which indicates that the simulation results of RTDDM can approach the optimal results when the nodes are deployed densely.

Figure 6 shows the results of RTDDM, BINARY, BUBBLE, and GSaR. Obviously, the delivery ratios of three algorithms are increasing as the number of nodes increases. This is because a denser deployment of nodes usually provides more opportunities for data holders to disseminate the held data packets. As illustrated in Figure 6(a), RTDDM achieves a larger delivery ratio than the other algorithms. Similarly, in Figure 6(b), RTDDM also outperforms other algorithms in terms of delivery delay, and this phenomenon is attributed to the fact that RTDDM exploits the region type correlations which can reflect the movement preferences of nodes, and

FIGURE 7: Impact of t^* on delivery ratio.FIGURE 8: Impacts of the maximum number of copies Q .

thus the data packets can be delivered to the destination region more quickly.

The delivery ratios of three algorithms perform better when the number of allowable times slots becomes larger, and the reason is that more data packets can be delivered to the destination region during a longer period, as shown in Figure 7.

Besides, we observe the impact of the maximum number of copies Q , and the results are given in Figure 8. We find that a larger Q (i.e., more copies of each data packet) gives rise to a larger delivery ratio, along with a shorter delivery delay, although the dissemination cost will be accordingly

increased. In some practical applications (especially the cost-sensitive applications), the dissemination cost must be taken into account, and the value of Q should be restricted to make a preferable tradeoff among the metrics of delivery ratio, delivery delay, and dissemination cost.

7. Conclusions

This study explores the data dissemination problem for the delivery ratio enhancement and delivery delay reduction in MONs. The region type correlations and region weights are explored and utilized to select the relay nodes for data

TABLE I: Simulation Parameters.

Parameter	Description	Value
N	Number of nodes	3,000
$ D $	Network space	$100 \times 100\text{m}^2$
r_c	Communication range	15 m
t^*	Number of allowable time slots	9
τ_s	Duration of each time slot	6 s
K	Number of regions	100
M	Number of region types	10
p_h	Threshold of home region	0.35
α	Exponent parameter	0.6
γ	Exponent parameter	0.4

holders. Then, a Region Type based Data Dissemination Method (RTDDM) is proposed. In RTDDM, the region type correlations are first calculated, and then the relay nodes are determined through a Markov decision model. Simulation results suggest that RTDDM can improve the delivery ratio and reduce the delivery delay, and it is especially suitable for the MONs where there are various region types concerning the node contacts, such as the pocket switched networks comprised of human-carried mobile devices.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by National Natural Science Foundation of China under Grant Nos. 61872191, 61801215, and 61872193; Postdoctoral Science Foundation of China under Grant Nos. 2014M560379 and 2015T80484; Natural Science Foundation of Jiangsu Province under Grant No. BK20160812.

References

- [1] D. Zhao, H. Ma, S. Tang, and X.-Y. Li, "COUPON: a cooperative framework for building sensing maps in mobile opportunistic networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 392–402, 2015.
- [2] M. Tasfe and A. Chakrabarty, "Gossip: A social interest based routing algorithm for pocket switched network," in *Proceedings of the 20th International Conference of Computer and Information Technology*, Dhaka, Bangladesh, 2017.
- [3] O. Alagoz, H. Hsu, A. J. Schaefer, and M. S. Roberts, "Markov decision processes: A tool for sequential decision making under uncertainty," *Medical Decision Making*, vol. 30, no. 4, pp. 474–483, 2010.
- [4] K. Chatterjee, R. Majumdar, and T. A. Henzinger, "Markov decision processes with multiple objectives," in *STACS 2006*, vol. 3884 of *Lecture Notes in Comput. Sci.*, pp. 325–336, Springer, Berlin, 2006.
- [5] D. Chen, L. Liu, and J. Wu, "A Region Type-based Data Dissemination Method for Mobile Opportunistic Networks," in *Proceedings of the 18th IEEE International Conference on Communication Technology*, Chongqing, China, 2018.
- [6] D. McGeehan, D. Lin, and S. Madria, "ChitChat: An Effective Message Delivery Method in Sparse Pocket-Switched Networks," in *Proceedings of the IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, Nara, Japan, June 2016.
- [7] Y. Cao, Z. Sun, N. Wang, M. Riaz, H. Cruickshank, and X. Liu, "Geographic-based spray-and-relay (GSaR): an efficient routing scheme for DTNs," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1548–1564, 2015.
- [8] S. Batabyal and P. Bhaumik, "Analysing social behaviour and message dissemination in human based delay tolerant network," *Wireless Networks*, vol. 21, no. 2, pp. 513–529, 2014.
- [9] Z. Cai, Y. Duan, and A. G. Bourgeois, "Delay efficient opportunistic routing in asynchronous multi-channel cognitive radio networks," *Journal of Combinatorial Optimization*, vol. 29, no. 4, pp. 815–835, 2015.
- [10] M. Fida, M. Ali, and A. Adnan, "Socialcircle: A message forwarding technique for pocket switched networks," in *Proceedings of the 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Denton, Texas, USA, July 2015.
- [11] K. Rasul, D. Makaroff, and K. G. Stanley, "Hybrid Community-Based Forwarding: A complete energy efficient algorithm for pocket switched networks," in *Proceedings of the 2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*, pp. 760–768, Clearwater Beach, Fla, USA, October 2015.
- [12] R. Barua, S. Shadman, and A. Chakrabarty, "PNGP : A social relationship based routing algorithm for pocket switched network," in *Proceedings of the 19th International Conference on Computer and Information Technology, ICCIT 2016*, pp. 25–30, Bangladesh, December 2016.
- [13] K. Sugiyama, T. Kubo, A. Tagami, and T. Hasegawa, "Mean free path applied to message dissemination in opportunistic networks," in *Proceedings of the IEEE INFOCOM 2015 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*, pp. 420–425, Hong Kong, Hong Kong, April 2015.
- [14] R.-I. Ciobanu, R.-C. Marin, C. Dobre, and V. Cristea, "Interest-awareness in data dissemination for opportunistic networks," *Ad Hoc Networks*, vol. 25, pp. 330–345, 2015.
- [15] J. Orozco, R. Santos, S. F. Ochoa, and R. Meseguer, "A Stochastic Approach for Modeling Message Dissemination in Opportunistic Networks," *Wireless Personal Communications*, vol. 97, no. 2, pp. 2207–2228, 2017.
- [16] P. Santi, *Mobility Models for Next Generation Wireless networks: Ad Hoc, Vehicular and Mesh Networks*, Wiley Publishing, 2012.
- [17] W. Gao and G. Cao, "On exploiting transient contact patterns for data forwarding in Delay Tolerant Networks," in *Proceedings of the 2010 18th IEEE International Conference on Network Protocols (ICNP)*, pp. 193–202, Kyoto, Japan, October 2010.
- [18] L. Chen and J.-Q. Sun, "The closed-form solution of the reduced Fokker-Planck-Kolmogorov equation for nonlinear systems," *Communications in Nonlinear Science and Numerical Simulation*, vol. 41, pp. 1–10, 2016.

- [19] R. Diana and E. Lochin, "Modelling the delay distribution of Binary Spray and Wait routing protocol," in *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, San Francisco, Calif, USA, 2012.
- [20] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: social-based forwarding in delay-tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1576–1589, 2011.

