Message delivery in a mobile social network (MSN) is difficult due to the fact that the topology of such network is sparse and unstable. Various routing schemes for MSNs were proposed to make the message delivery robust and efficient. However, little research has been conducted to explore how much delay has to be tolerated for the message delivery from the source to the destination. Since the social relationships among nodes are stable during a certain period of time, it is expected that the delay of message delivery in MSNs could be modeled with a probability model. In this paper, we take the first step to address this issue. We firstly extract three routing models from the existing routing schemes for MSNs and then develop the probability models of the message transmission delay for each abstract routing model. The simulation results show that the theoretical models match very well the simulation trace statistics.

1. Introduction

Mobile social network is a type of delay-tolerant networks (DTN [1]) which take into consideration the sociality of the participating nodes of the networks [2]. If the terminals of MSN are smartphones facilitated with rich sensing capability, it can function as a typical distributed sensor network [3]. Similar to the traditional mobile ad hoc networks (MANETs), no fixed infrastructures are built in such networks, and communications between the nodes depend on short-distance wireless links, like Wi-Fi or even Bluetooth. Since equipment in an MSN is usually carried by people moving freely, those nodes may organize themselves into highly dynamic topology, and the links between the nodes are usually intermittent and connected by opportunity due to the mobility of the nodes. As a result, large transmission delay is allowed in MSN. So in some extent, MSN is considered as a type of delay-tolerant networks. Although routing in an MSN might adopt the similar store-and-forward strategies [4] that dominate DTNs, it is essentially different from the DTN routing because the sociality of the user behavior can be cooperated explicitly to improve the efficiency of networking and communication in MSNs. Previous works [5–7] have shown that the performance of routing in MSN may heavily depend on the users’ social behavior like that in a human social network, which is also the key characteristic for the design and analysis of the other issues in MSNs; see, for example [8, 9].

Traditional MANET routing protocols such as AODV [10], DSR [11], DSDV [12], and LAR [13] make assumption that the topologies of the networks are fully connected. These protocols will fail to route any message if there is no precomputed route from source to destination at the time of message being sent. In an opportunistic MSN, nodes carry the data to be forwarded and also are ready to forward data for other nodes. What is more, the mobility of nodes can be exploited to forward data opportunistically upon the encounter with each other. The key problem here is hence how to design appropriate relay selection strategies to improve the opportunity or probability of data forwarding. This improvement will make the data forwarding more efficient with less delay. Relay selection strategies depending on the history of contacts among nodes were proposed to find the right forwarding relays [14, 15].

While the end-to-end connections cannot be preset when the message is ready for transmission, social networks often demonstrate unique social characteristic like the so-called small-world phenomenon shown in the Milgram’s 1967 mail transfer experiment [16]. In a social network, two-people contacts frequently usually have social ties. Recently, several
sociality-aware routing strategies have been proposed to improve the routing efficiency for the MSNs [17–19]. Both contact history-based and social network metrics-based routing schemes designed for MSNs are best-effort, and there is no delay guarantee for the message delivery from the source to the destination. While MSNs are transmission delay-tolerant, it is still curious that how much delay must be tolerated for the message to be delivered from source to the destination or the delay bounds for message delivery in certain scenarios. These bounds are particularly interesting for the future multimedia application over the MSNs. While most of the work for data delivery in MSNs focuses on the design of new routing schemes, researches about how to evaluate the message delay remain rare and we try to make a first step to fill this gap.

In our opinion, the difficulty of the delay estimation for MSNs comes from two aspects: the dynamics of network topology and the uncertainty of the routing path. Though the mobility of nodes makes the networks' physical topology change dynamically, the nodes' social behavior tries to maintain regular contacts among nodes with social ties. We believe, therefore, it is possible to predict the routing behavior, at least the bounds of end-to-end delay of message delivery in an MSN.

In this paper, we illuminate a unified framework for evaluating the delay of message delivery in MSNs. Firstly, we propose general routing models extracted from the existing routing schemes. Then, we evaluate the delay with stochastic process theory based on the extracted routing model. The contributions of our paper are in threefold. First, we extract characteristics from the existing MSN routing strategies and classify them into three general models. Second, we propose a general method to evaluate the end-to-end delay in an MSN. And last, we design simulation experiments to validate the theoretical model. To our knowledge, this is the first effort to estimate the end-to-end delay of message delivery in an MSN via a probability framework.

The rest of this paper is organized as follows. Section 2 reviews the existing work. Section 3 presents the abstract routing model of the MSNs. Section 4 analyzes the delay bounds of the three types of general routing models with stochastic process theory and gives out the estimation results. Section 5 validates the performance of the proposed probability delay model by comparing the theoretical result with the simulation, and Section 6 concludes the paper.

2. Related Work

There is much work on the routing schemes for MANETs [20], from which the DTNs were evolved. The researches about routing schemes in a DTN might originate from the work of Jain et al. [4], in which messages are to be moved end-to-end across a time-varying connected graph but the topology dynamics may be known in advance. After that, there is much progress on this topic [21, 22].

Epidemic routing [23] which is originally designed for MANETs could deliver data fast and robustly by forwarding data to any encounters. To satisfy the limitation of energy and memory of nodes, a variety of relay selection strategies specially for DTNs are proposed [14, 15, 17–19]. Lindgren et al. proposed the ProPhet [14], making use of contact history to predict the probability of the meeting of the two nodes. Only nodes that can reach the destination with a higher probability will get the data from the relay.

Data dissemination in the delay tolerant MSNs is a critical component of many applications, for example, content update in an online social networks [24]. Sociality-based routing schemes for the DTNs have also been studied in recent years. SimBet routing [17] uses two social metrics (centrality and similarity) to estimate or predict the probability that potential relay nodes may reach the destination. Unlike the ProPhet and Epidemic, only one copy of data exists in the networks by SimBet routing. After that, Daly and Haahr tried to improve SimBet by using multiple copy of data forwarding [18]. Recently, Hui et al. [19] proposed a novel sociality-based forwarding algorithm, BUBBLE, which employed two social and structural metrics, namely, centrality and community.

Though there have been many research works on the design of routing schemes for the delay tolerant MSNs, model-based performance evaluation of these routing schemes is relatively scarce. Boldrini et al. [25] considered a utility-based cooperative data dissemination system in which the utility of data was defined based on the social relationships between users. Specifically, they designed a Markov model to characterize the data dissemination process in both its stationary and transient regimes. The main result of their analysis is that the data distribution process always converges to one of two possible stationary regimes.

Our study is obviously different from above work. In this paper, we concentrate on how to evaluate the delay from the source to the destination in a delay tolerant MSN by a probabilistic way. Based on the abstracted routing models, the data disseminating process is modeled as a stochastic process to estimate the delay to be experienced in such routing process.

3. Models

3.1. Network Model. In an MSN, nodes may be in a moving status and links between some nodes do not always exist. These characteristics make the topology of an MSN dynamic. However, during a certain time, the social relationships among the nodes are fixed. Being similar with [26, 27], we assume the pairwise node intercontact time is exponentially distributed. We consider the links among nodes as the social relationships of their holders, and the network model can be described as follows.

Consider an MSN with $n$ mobile nodes, which can be denoted as a graph $G = (V, E)$, where $V$ is the set of nodes and $E$ is the set of links between nodes. If there is social relationship between node $i$ and node $j$, let $e_{ij} \in E$ denote the edge $e_{ij}$ between them. Letting $\lambda_{ij}$ denote the weight of the link $e_{ij}$, then the intercontact time between nodes $i$ and $j$ obeys exponential distribution with parameter $\lambda_{ij}$. And further, the contacts between nodes $i$ and $j$ form a homogeneous Poisson process with the contact rate $\lambda_{ij}$.

3.2. The Model of MSN Routing. An MSN may experience frequent, persistent link partitioning and may never have
a stable end-to-end path. Like in DTNs, routing in MSN employs the store-and-forward strategies over the opportunistic links. From various routing schemes existing currently, we extracted the following essential attributes which play the key roles for message delivery in an MSN.

(a) The relay selection strategy: this is without doubt the most important issue in the design of MSN routing, which directly affects the routing performance. Generally, we divided the strategies into two categories: nonstrategic that forwards data to any nodes it meets, and strategic that only forwarding data to those nodes which have better forwards quality.

(b) The data copying strategy: to improve the successful ratio of message delivery, the relays usually forward multiple copies of the data to their neighbors concurrently. However, this strategy significantly increases the consumption of the network resource. The data copying strategy here can be classified as single-copy and multicopy in general. The multicopy strategies can be further divided into finite-copy and infinite-copy strategies. While the routing scheme with finite-copy strategy usually forward fixed number of copies to the neighbors, the routing schemes with infinite-copy strategy simply forward the data to all the nodes it encounters.

(c) Clustering: while a good relay selection strategy can restrain the number of data copy to be forwarded, clustering is an effective approach to limit the over-consumption of the resource in a large scale MSN. According to whether the nodes will be clustered, the routing schemes for MSNs can be classified into layered routing and plain routing.

In this paper, only plain routing is considered. We extract three types of routing models from existing instances of routing schemes designed for MSNs based on the first two attributes described above.

3.2.1. Single-Copy Strategic Routing (SCSR). The routing strategy refers that node carrying data will not indiscriminately forward data to whichever it encounters but only chooses those nodes that can forward the data to destination with one or more hops. The usual routing strategies include what we have mentioned before, for example, the contacts history-based and social metric-based relay selection. A typical routing scheme belonging to this type is SimBet.

3.2.2. Multicopy Routing without Strategies (MCR-WS). This is the simplest routing schemes. Mobile node will forward its carried data to any nodes it encounters and keep the data at the same time. An infinite-copy strategy is more popular than the finite multiple copy strategy. Epidemic routing belongs to this type of routing schemes.

3.2.3. Multicopy Strategic Routing (MCSR). In this type of multicopy routing schemes, strategies are considered which makes routing decisions more complex. Since it combines the simplicity of multicopy routing and efficiency of strategic routing, it attracts more attention and most of the routing schemes belong to this type. Compared with MCR-WS routing, if the relay selection strategy is designed to be effective enough, MCSR could achieve similar performance while consuming less system resources. ProPhet is a typical routing scheme of this type.

3.3. Some Assumptions of the Routing Model. We make the following assumptions listed from weak to strong.

(1) There is enough time for the nodes to exchange their data when they contact.

(2) For the intercontact time is much longer than the data transmission time over the link, the latter would be ignored when evaluating the transmission delay.

(3) In the case of multicopy routing strategy, nodes’ buffer capacity is large enough thus no packet will be discarded due to the lack of memory.

(4) TTL (time-to-live) field is set as time limitation and the data will be discarded actively if it does not arrive at the destination after TTL time’s forwarding.

(5) All nodes will not receive the same data for two times by numbering the message with a global ID created by a Hash function.

4. Analysis

We will analyze the delay of message delivery in MSNs based on the three routing models in Section 3. Before the detailed analysis, we firstly present an important result in Lemma 1. Assume that the source intends to send data to the destination D through the intermediate node set \(\{N_1, N_2, \ldots, N_r\}\) as shown in Figure 1. Denote \(\lambda_1, \lambda_2, \ldots, \lambda_{r+1}\) as the weight of intermediate links in sequence, and the corresponding intercontact time is \(X_1, X_2, \ldots, X_{r+1}\). As described in Section 3.1, the intercontact time between nodes follows exponential distribution, so the probability density function of \(X_k\) is \(f_{X_k}(x) = \lambda_k e^{-\lambda_k x}\). Then the maximal aggregated time needed to forward data from S to D is

\[ Y = \sum_{i=1}^{r+1} X_i, \]  

which is hypoexponentially distributed [28] according to the following lemma.

Lemma 1 (section 5.2.4 of [28]). For an opportunistic path with r hops, the corresponding edges weight as \(\lambda_1, \lambda_2, \ldots, \lambda_{r+1}\), then the probability density function of \(Y\) as in (1) is

\[ f_Y(x) = \sum_{k=1}^{r+1} C_k (r+1) f_{X_k}(x), \]  

where \(C_k (r+1) = \Pi_{s=1, s \neq k}^{r+1} (\lambda_s / (\lambda_s - \lambda_k)).\)
any node, these two paths are independent with each other. However, all paths in \( \{Y_1, Y_2, \ldots, Y_m \} \) from S to D usually are not independent, and some correlated paths should be fixed approximately. Assuming that \( P_1 = \{N_1^{(i)}, N_2^{(i)}, \ldots, N_m^{(i)} \} \) and \( P_j = \{N_1^{(j)}, N_2^{(j)}, \ldots, N_n^{(j)} \} \) are two paths sharing intermediate node, the shared node \( N_1^{(i)} \) or \( N_1^{(j)} \) is called broken node if \( N_1^{(i)} = N_1^{(j)} \) and \( N_1^{(i)} \neq N_1^{(j)} \). In contrast, \( N_k^{(i)} \) or \( N_k^{(j)} \) is called backing node if \( N_k^{(i)} = N_k^{(j)} \) and \( N_k^{(i)} \neq N_k^{(j)} \). By the position of the broken node and the next backing node, two subpaths can be found which have the same source \( S' \) (broken node) and the same destination \( D' \) (backing node). Because the two subpaths share no common intermediate nodes, they are independent. As a result, through Lemma 1 and Proposition 3, the expectation of transmission delay from \( S' \) to \( D' \) can be calculated as \( T' \), building a dummy connection between \( S' \) and \( D' \) instead of the paths which connecting them. The distribution of intercontact time \( T' \) between \( S' \) and \( D' \) is approximated as exponential distribution with parameter \( \lambda_{S'D'} = 1/T' \). Through the above approximate treatment, two related paths could be merged into one path. What is more, if the same merging process is done on all the paths \( \{Y_1, Y_2, \ldots, Y_m \} \), a group of independent paths will be constructed. Finally, transmission delay evaluation of MCR-WS will be done using Proposition 3 directly.

4.2. Delay of Single-Copy Strategic Routing. Now we consider the second case, SCSR. The strategy here refers to that relaying nodes will not forward data to any encounters but select those that could forward the data to the destination with better quality. There are several different relay selection strategies. Some maintain a local encountering probability by contact history and others calculate the utility value using social metrics. Despite of the distinct of the strategies, no strategy can guarantee that the relay selection will always reach the destination. It is assumed that the probability of selecting a right relay is a constant \( P_{\text{Eval-W}} \). We will show how to estimate the value of \( P_{\text{Eval-W}} \) later.

Let \( V \) be the node forwarding data and \( N(V) \) its neighbor set. Because not all neighbors can lead to the destination, let \( W(V) \subseteq N(V) \) denote nodes that can forward data to the destination and tend to be selected as the next relay by the strategy. Due to the single-copy strategy, only one of the nodes in \( N(V) \) will receive data from \( V \). To make sure the data could be forwarded to the destination with highest probability, the intercontact history between node \( V \) and the candidates of relays is used to decide the probability of relay selection. Let \( P_i(N_i) \) denote the probability that node \( N_i \in W(V) \) is selected by \( V \) as the next relay. For node \( N_j \in W(V) \), it is obvious that the more frequently contacting with \( V \), the higher probability of \( P_i(N_i) \). So we have

\[
\sum_{i \in W(V)} P_i(N_i) = P_{\text{Eval-W}}
\]

\[
P_i(N_j) = \omega(N_i, V) : \omega(N_j : V) \quad \forall i, j \in W(V).
\]
In most cases, there are many possible paths connecting the source $S$ and the destination $D$. All the possible path set can be found and it is denoted as $\text{PathSet}(S, D) = \{(P_1, \text{Prob}_1), (P_2, \text{Prob}_2), \ldots, (P_m, \text{Prob}_m)\}$, in which $\text{Prob}$ denotes the probability of selecting path $P_i$ for message delivery. Assuming the path $P_i$ includes the following nodes $N_1$ (source), $N_2$, $\ldots$, $N_{m-1}$, $N_m$ (destination), then

\[
\text{Prob}_i = \prod_{j=2}^{m} P_i(N_j). \quad (6)
\]

Because there is only one data copy in the network, with the assumption of independent probability of possible path selection, we can calculate the probability of the successful message delivery:

\[
P_{\text{Eval}} = \sum_{i=1}^{m} \text{Prob}_i. \quad (7)
\]

What is more, it is easy to get the probability of successful message delivery of the real scenario and we denote it as $P_{\text{Real}}$. Then letting $P_{\text{Eval}} = P_{\text{Real}}$, with (5), (6), and (7), the probability of selecting the proper relay $P_{\text{Eval-W}}$ can be calculated.

Through above discussion, the preparation for the evaluation of SCSR has been done and then Proposition 4 is given as follows.

**Proposition 4.** In the type of SCSR, assume the possible routing path set is $\text{Path}(S, D) = \{(P_1, \text{Prob}_1), (P_2, \text{Prob}_2), \ldots, (P_m, \text{Prob}_m)\}$, in which $\text{Prob}$ denotes the probability of selecting path $P_i$. And let $Y_i$ denote the transmission delay over path $P_i$. Then the delay $T$ needed to send data from source to destination successfully has the following probability density function:

\[
f_T(x) = \sum_{i=1}^{m} f_{Y_i}(x) \times \frac{\text{Prob}_i}{P_{\text{Real}}}, \quad (8)
\]

where $P_{\text{Real}} = \sum_{i=1}^{m} \text{Prob}_i$ and $f_{Y_i}(x)$ is the probability distribution function of the variable $Y_i$.

**Proof.** Firstly, it is proved that $f_T(x)$ satisfies the property of probability density function. Obviously, for all $x \in R$, $f_T(x) \geq 0$ and it has

\[
\int_{-\infty}^{+\infty} f_T(x) \, dx = \int_{-\infty}^{+\infty} \left( \sum_{i=1}^{m} f_{Y_i}(x) \times \frac{\text{Prob}_i}{P_{\text{Real}}} \right) \, dx = \sum_{i=1}^{m} \text{Prob}_i \times \left( \int_{-\infty}^{+\infty} f_{Y_i}(x) \, dx \right) = \sum_{i=1}^{m} \frac{\text{Prob}_i}{P_{\text{Real}}} = 1.
\]

So $f_T(x)$ can be probability density function of variable $T$. Then we prove that the relationship in (8) makes sense. Because of the single-copy strategy, there is only one path that is selected to transmit data. The probability Prob of each path being chosen can be calculated according to (5), (6), and (7), and the distribution functions of transmission delay $Y_i$ on each path are known; we have

\[
p(T = x) = \sum_{i=1}^{m} p(Y_i = x) \times \text{Prob}_i. \quad (10)
\]

By the influence of policy, the data may not be sent to the destination for selecting the wrong path, which makes the transmission delay infinite. If the case of choosing the broken path is not taken into consideration, the sum of chosen probability of all the possible left paths $\sum_{i=1}^{m} \text{Prob}_i$ is less than 1, which makes the component of $T$ uncompleted. But if we normalize the probability of selecting the proper path, that is to say, letting the probability Prob be normalized as $\text{Prob}_i/P_{\text{Real}}$, which would keep the component of $T$ completed, then we can get the probability density function

\[
f_T(x) = \sum_{i=1}^{m} f_{Y_i}(x) \times \frac{\text{Prob}_i}{P_{\text{Real}}} \quad \text{under the condition of successful message delivery}.
\]

By Proposition 4, the expected time delay from the source node $S$ to the destination $D$ in this type of routing schemes can be easily got as

\[
E[T(S, D)] = \sum_{i \in \text{Path}(S, D)} E[Y_i] \times \frac{\text{Prob}_i}{P_{\text{Real}}}. \quad (11)
\]

**4.3 Delay of Multicopy Strategic Routing.** The multicopy strategy improves the probability of successful message delivery comparing with the SCSR. And the relay selection strategy improves the probability of successful message delivery.

By normalizing the probability of selecting the proper path, the expected time delay from the source node $S$ to the destination $D$ can be easily got as

\[
E[T(S, D)] = \sum_{i \in \text{Path}(S, D)} E[Y_i] \times \frac{\text{Prob}_i}{P_{\text{Real}}}. \quad (11)
\]

The method of dealing with relay selection strategy is similar to that in SCSR. We still assume the probability of selecting the proper relay is a constant $P_{\text{Eval-W}}$. However, due to the multicopy strategy, every node in $W(S)$ is selected with the same probability $P_{\text{Eval-W}}$. All paths connecting the source $S$ and destination $D$ will be denoted as $\text{Path}(S, D) = \{(P_1, \text{Prob}_1), (P_2, \text{Prob}_2), \ldots, (P_m, \text{Prob}_m)\}$. Then the probability of transferring data on path $P_i$ is

\[
\text{Prob}_i = (P_{\text{Eval-W}})^{|P_i| - 1}, \quad (12)
\]

where $|P_i| - 1$ is the number of relays selected. Unless all the possible paths fail, the data will fail to be transmitted to the destination. So the probability of successfully sending data is then $P_{\text{Eval-W}}$, and each probability $\text{Prob}_i$ can be calculated. To evaluate the transmission delay in the MCSR, Proposition 5 is given as follows.
Proposition 5. In the type of MCSR, node $S$ sends message to node $D$ and the all possible routing path set is $\text{Path}(S, D) = \{ (P_1, \text{Prob}_1), (P_2, \text{Prob}_2), \ldots, (P_m, \text{Prob}_m) \}$. If the routing paths are independent with each other, the transmission delay $T$ follows:

$$F_T(x) = 1 - \prod_{i=1}^{m} (1 - F_{Y_i}(x)) \times \text{Prob}_i,$$

(14)

Proof. We consider all the cases of $T > x$. Firstly, if the data is sent successfully to the destination, it requires all transmission delay on every possible routing path that has $Y_i > x$. Because all $Y_i$ are independent with each other, through the distribution function of $Y_i$, the probability of $Y_i > x$ for $i = 1, 2, \ldots, m$ is $\prod_{i=1}^{m} (1 - F_{Y_i}(x))$. For every possible routing path is selected at the probability of $\text{Prob}_i$, the probability of $Y_i > x$ should be multiplied by $\text{Prob}_i$. Then we have $F_T(x) = 1 - \prod_{i=1}^{m} (1 - F_{Y_i}(x)) \times \text{Prob}_i$. Secondly, we verify whether $F_T(x)$ fulfills the definition of distribution function. Obviously, it is true that $\lim_{x \to \infty} F_{Y_i}(x) = 1$, for $\lim_{x \to \infty} F_{Y_i}(x) = 1$, for all $i = 1, 2, \ldots, m$. \qed

Proposition 5 requires that all possible correlated routing paths are independent with each other. Some fixing similar to MCR-WS should be done. Two related routing paths that share intermediate nodes will be merged. The merging process is as same as the process in MCR-WS, but the probability of the new merged routing path should be re-calculated. Assume, in the merging process, two subpaths which have $m_1$ and $m_2$ nodes, respectively, we use their common head node $S'$ and tail node $D'$. After they have been merged, the probability of having data flow from $S'$ to $D'$ is $1 - (1 - (P_{\text{Eval-W}})^{m_1}) \times (1 - (P_{\text{Eval-W}})^{m_2})$. Through the path merging and probability modifying process, a new independent routing path will be got. Then using Proposition 5, the transmission delay in MCSR can be calculated.

5. Simulation

In this section, we validate our delay model by comparing the theoretical delay model with the simulation results. All of the three routing schemes corresponding to each routing model are included: Epidemic [23] for MCR-WS, SimBet [17] for SCSR, and ProPhet [14] for MCSR.

5.1. Simulation Setup. The simulation experiment is developed based on the general discrete event simulation platform OMNET++ [29]. Though the scale of network in our simulation could be set as any size only if it is within the resource constraints of the simulator, the theoretical delay model has no relevance to the scale. In the simulation, 30 nodes are randomly placed inside an area of 500 $\times$ 500 $\text{m}^2$. IEEE 802.11b DCF is used as the MAC layer protocol. The radio propagation range of all nodes is set at 100 meters. The nodes move according to the Gauss-Markov mobility models [30], with a minimum speed of 0 m/s and a maximum speed of 5 m/s. When two nodes are in the radio propagation range, it is considered that they meet and can communicate with each other. As models omit the transmission delay of the channel, the transmission delay of all channels is set as 0.

During the simulation, the source and destination nodes are selected randomly from the node set. Data packets are generated by the source node every five seconds and 3000 packets will be sent during this interval. The time-to-live fields of these packets are set as 15s, which means the message will be discarded if it has not arrived at the destination after 15 seconds. The intercontact time of nodes meeting each other is recorded to provide the social ties for the model-based numerical experiment. We approximate the parameter of the exponential distribution as the reciprocal of the average intercontact time. The parameter setting of ProPhet is as same as in [14], that is, $P_{\text{init}} = 0.75$, $\beta = 0.25$, and $\gamma = 0.98$. And the parameters of SimBet are just like in [17], that is, $\alpha = \beta = 0.5$. To simulate the packet routing process, every data packet is forwarded on the meeting of the expected nodes according to the routing schemes and the transmission delay is traced.

5.2. Model Evaluation. We try to evaluate the accuracy of the theoretical model by comparing the estimated delay with the corresponding statistics of the simulation result. We have already derived the probability density function or cumulative distribution function of the transmission delay of each routing model, respectively, as in (4), (8) and (14). Denote all the delays traced in the simulation as $\{d_1, d_2, \ldots, d_m\}$ and divide the time interval $[0, \text{TTL}]$ into constant $k$ parts equally. The constant $k$ should be carefully set. Then counting the number $c_i$ of $d_i$ belongs to each subinterval and let $c_i/m$ be the probability of the subinterval. In one simulation, the comparison results are shown in Figures 2, 3, and 4. It is obvious that the delay in simulation matches very well the delay calculated by the proposed theory model.

Then we run the simulation for 20 times of each routing model, each of which includes sending 3000 data packets. Each time the network scale is also $N = 30$ but with different network topology. Through the simulation, the average transmission delay is obtained and the expected delay by the model could also be calculated. Pearson Correlation Coefficient [31] is a measure of the correlation between two variables $X$ and $Y$, giving a value between $+1$ and $-1$ inclusive. The more close to $+1$ or $-1$, the more related they are. It is commonly denoted with $r$ and can be obtained from the following formula:

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}.$$  

(15)

We use Pearson Correlation Coefficient to show further how much the average transmission delay in simulation matches the theoretical result. The result presented in Table 1...

| Routing model | Pearson Correlation Coefficient |
|---------------|---------------------------------|
| MCR-WS        | 0.8591                          |
| SCSR          | 0.8021                          |
| MCSR          | 0.9065                          |

Table 1: Pearson Correlation Coefficient.
shows that the simulation delays and theory-calculated delays are highly related.

6. Conclusions

In this paper, we proposed a probability model to estimate the delay of message delivery in the delay tolerant mobile social networks. We firstly extracted three general routing models from the existing various routing schemes for MSNs. According to an elegant result in probability theory, we constructed probability delay models for each of the three routing models. Then the simulation experiments were designed to validate the accuracy of the theoretical delay model. It was found that the delay statistics from the simulation trace matched very well the theoretical results, which means that proposed model is quite accurate for the prediction of delay of message delivery in an MSN.

During the model construction, we omitted the storage limitation of the nodes in an MSN. In the future work we will enhance the proposed delay model considering the message loss from the storage limit. For only the plain routing schemes for MSNs being considered in the current work, delay estimation for the layered routing schemes is also a valuable extension to this work.

Acknowledgments

The work was supported by National Science Foundation of China under Grant no. 61070170, Science Research Program of Universities in Jiangsu Province under Grant no. 11KJB520017, and Application Foundation Research of Suzhou (Jiangsu, China) under Grant no. SYG201238.

References

[1] K. Fall, “A delay-tolerant network architecture for challenged internets,” in Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 27–34, August 2003.

[2] H. Falk, “Applications, architectures, and protocol design issues for mobile social networks: a survey,” Proceedings of the IEEE, vol. 99, no. 12, pp. 2125–2129, 2011.
[3] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, “A survey of mobile phone sensing.” IEEE Communications Magazine, vol. 48, no. 9, pp. 140–150, 2010.

[4] S. Jain, K. Fall, and R. Patra, “Routing in a delay tolerant network,” in Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 145–158, New York, NY, USA, September 2004.

[5] E. Yoneki, P. Hui, S. Chan, and J. Crowcroft, “A Socio-Aware Overlay for publish/subscribe communication in delay tolerant networks,” in Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems, pp. 225–234, New York, NY, USA, October 2007.

[6] W. Gao, Q. Li, B. Zhao, and G. Cao, “Multicasting in delay tolerant networks: a social network perspective,” in Proceedings of the 10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc’09), pp. 299–308, New York, NY, USA, May 2009.

[7] T. Henderson, D. Kotz, and L. Abyzov, “The changing usage of a mature campus-wide wireless network,” in Proceedings of the 10th Annual International Conference on Mobile Computing and Networking, pp. 187–201, New York, NY, USA, October 2004.

[8] A. Beach, M. Gartrell, S. Akkala et al., “WhozThat? Evolving an ecosystem for context-aware mobile social networks,” IEEE Network, vol. 22, no. 4, pp. 50–55, 2008.

[9] W. D. Yu and A. Siddiqui, “Towards a wireless mobile social network system design in healthcare,” in Proceedings of the 3rd International Conference on Multimedia and Ubiquitous Engineering, pp. 429–436, June 2009.

[10] C. E. Perkins and E. M. Royer, “Ad-hoc on-demand distance vector routing,” in Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, pp. 90–100, February 1999.

[11] D. Johnson and D. Maltz, “Dynamic source routing in ad-hoc wireless networks,” Mobile Computing, vol. 353, pp. 153–181, 1996.

[12] C. E. Perkins and P. Bhagwat, “Highly dynamic destination-sequence distance-vector routing (DSDV) for mobile computers,” in Proceedings of the Conference on Communications Architectures, Protocols and Applications, vol. 24, no. 4, pp. 234–244, New York, NY, USA, 1994.

[13] Y.-B. Ko and N. H. Vaidya, “Location-aided routing (LAR) in mobile ad hoc networks,” Wireless Networks, vol. 6, no. 4, pp. 307–321, 2000.

[14] A. Lindgren, A. Doria, and O. Schelen, “Probabilistic routing in intermittently connected networks,” Mobile Computing and Communications Review, vol. 7, no. 3, pp. 19–20, 2003.

[15] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Spray and wait: an efficient routing scheme for intermittently connected mobile networks,” in Proceedings of the ACM SIGCOMM Workshop on Delay-Tolerant Networking, pp. 252–259, New York, NY, USA, August 2005.

[16] S. Milgram. “The small world problem,” Psychology Today, vol. 1, no. 1, pp. 60–67, 1967.

[17] E. M. Daly and M. Haahr, “Social network analysis for routing in disconnected delay-tolerant MANETs,” in Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 32–40, New York, NY, USA, September 2007.

[18] E. M. Daly and M. Haahr, “Social network analysis for information flow in disconnected delay-tolerant MANETs,” IEEE Transactions on Mobile Computing, vol. 8, no. 5, pp. 606–621, 2009.

[19] 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.

[20] E. M. Royer and C.-K. Toh, “A review of current routing protocols for ad hoc mobile wireless networks,” IEEE Personal Communications, vol. 6, no. 2, pp. 46–55, 1999.

[21] J. Ott, D. Kutscher, and C. Dwertmann, “Integrating DTN and MANET routing,” in Proceedings of the ACM SIGCOMM Workshop on Challenged Networks, pp. 221–228, September 2006.

[22] Y. Zhu, B. Xu, X. Shi, and Y. Wang, “A survey of social-based routing in delay tolerant networks: positive and negative social effects,” IEEE Communications Surveys & Tutorials, vol. 15, no. 1, pp. 387–401, 2013.

[23] A. Vahdat and D. Becker, “Epidemic routing for partially connected ad hoc networks,” Tech. Rep. CS-200006, Duke University, 2000.

[24] S. Ioannidis, A. Chaintreau, and L. Massoulie, “Optimal and scalable distribution of content updates over a mobile social network,” in Proceedings of the 28th Conference on Computer Communications (IEEE INFOCOM ’09), pp. 1422–1430, April 2009.

[25] C. Boldrini, M. Conti, and A. Passarella, “Modelling data dissemination in opportunistic networks,” in Proceedings of the 3rd ACM Workshop on Challenged Networks, pp. 89–96, New York, NY, USA, September 2008.

[26] L. C. Freeman, “A set of measures of centrality based on betweenness,” Sociometry, vol. 40, no. 1, pp. 35–41, 1977.

[27] L. C. Freeman, “Centrality in social networks conceptual clarification,” Social Networks, vol. 1, no. 3, pp. 215–239, 1979.

[28] S. M. Ross, Introduction To Probability Models, Academic Press, 2006.

[29] OMNeT++ Community Site, http://www.omnetpp.org/.

[30] D. B. Johnson and D. A. Maltz, “Dynamic source routing in Ad Hoc wireless networks,” in Mobile Computing, vol. 353 of The Kluwer International Series in Engineering and Computer Science, pp. 153–181, 1996.

[31] http://en.wikipedia.org/wiki/Pearson_correlation_coefficient.