RELAY SELECTION BASED ON SOCIAL RELATIONSHIP PREDICTION AND INFORMATION LEAKAGE REDUCTION FOR MOBILE SOCIAL NETWORKS

XIAOSHUANG XING
School of Computer Science and Engineering
Changshu Institute of Technology, Changshu, China
Provincial Key Laboratory for Computer Information Processing Technology
Soochow University, Suzhou, China

GAOFEI SUN* AND YONG JIN
School of Computer Science and Engineering
Changshu Institute of Technology, Changshu, China

WENYI TANG
the School of Computer Science and Engineering
University of Electronic Science and Technology of China, Chengdu, China

XIUZHEN CHENG
Computer Science, The George Washington University
Washington DC, USA

(Communicated by Zhipeng Cai)

Abstract. Despite the extensive study on relay selection in mobile social networks (MSNs), few work has taken both transmission latency (i.e. efficiency) and information leakage probability (i.e. security) into consideration. Therefore we target on designing an efficient and secure relay selection algorithm to enable communication among legitimate users while reducing the information leakage probability to other users. In this paper, we propose a novel mobility model for MSN users considering both the randomness and the sociality of the movements, based on which the social relationship among users, i.e. the meeting probabilities among the users, are predicted. Taken both efficiency and security into consideration, we design a network formation game based relay selection algorithm by defining the payoff functions of the users, designing the game evolving rules, and proving the stability of the formed network structure. Extensive simulation is conducted to validate the performance of the relay selection algorithm by using both synthetic trace and real-world trace. The results show that our algorithm outperforms other algorithms by trading a balance between efficiency and security.

2010 Mathematics Subject Classification. Primary: 58F15, 58F17; Secondary: 53C35.
Key words and phrases. Mobile social networks, routing, relay selection, information leakage, mobility model.

The first author is supported by National Natural Science Foundation of China (61602062) and the Natural Science Foundation of Jiangsu Province (BK20160410).
* Corresponding author: Gaofei Sun.
1. Introduction. Nowadays, people carry various mobile devices such as smartphones, PDAs, and laptops to leverage WiFi, Internet, and 3G/4G cellular networks for pervasive communication services. When such infrastructures are unavailable, these devices can form a self-organizing network and enable communication among users by using the opportunistic encounters between them. Such a network is termed as a mobile social network (MSN) [8]. Finding appropriate relays and forming suitable routing paths play critical roles for information dissemination in MSNs.

Extensive effort has been taken to design efficient relay selection algorithms and routing protocols for MSNs. Moreover, taking into account the fact that human social behavior heavily impacts the operation of MSNs, researchers have designed a number of relay selection algorithms by considering human social features such as the visiting frequencies of human beings to certain locations and the encounter frequencies among human beings [24, 25]. However, few literature has considered the fundamental problem of preventing information leakage to non-destination users. Despite the employment of encryption mechanism [3, 26], reducing the information leakage probability is still critical as encryption mechanism is not “perfectly secure”. Therefore, we target on designing an efficient and secure relay selection algorithm to enable communication among legitimate users while reducing the information leakage probability to other users in MSNs.

The main contributions of this paper are as follows: First, both the social relationship among the users and the overhearing probabilities of the users are considered in our design, through which the efficiency and the security of the relay selection algorithm can be ensured. Second, we propose a novel mobility model for MSN users considering both the randomness and the sociality of the movements, based on which the social relationship among users are predicted. Third, we propose a network formation game based relay selection algorithm. The payoff functions of the information source, the candidate relays, and the information destination are defined, the game evolving rules are designed, and the stability of the formed network structure is proved. Finally, extensive simulation is conducted to validate the performance of the proposed relay selection algorithm by using both synthetic trace and real-world trace. The results show that our algorithm outperforms other algorithms by trading a balance between the transmission latency (i.e. efficiency) and the information leakage probability (i.e. security).

In the rest of the paper, we briefly summarize the related work on relay selection in MSNs in Section 2. We describe the problem to be solved in Section 3. In Section 4, we propose our mobility model by considering both the randomness and the sociality of the movements and predict the social relationship among users. An efficient and secure relay selection algorithm is proposed based on the network formation game in Section 5, where we define the payoff functions in Section 5.1, design the game evolving rules in Section 5.2, and prove the stability of the formed network structure in Section 5.3. A simulation study to validate the performance of the relay selection algorithm is reported in Section 6. Finally, we conclude the paper in Section 7.

2. Related work. The MSN routing schemes can be classified into four main categories: flooding based schemes, single-copy based schemes, utility based schemes, and quota based schemes. Flooding based schemes maintain low transmission latency at the expense of high storage cost and communication overhead [19]. On the contrary, single-copy based schemes reduce the storage cost and the communication overhead by allowing only one copy of an information packet to be transmitted.
in the network [22]. But they provide no performance guarantee on the transmission latency. To trade off between transmission latency and storage/communication cost, utility based schemes and quota based schemes are proposed. In utility based schemes, a user forwards the information only if it encounters another user with higher utility [6]. While in quota based schemes, the number of copies for each information packet is upper bounded by a fixed number [18].

Considering the fact that network operations are heavily impacted by human social behaviors, variants of the aforementioned routing schemes are proposed by exploiting different human social characteristics. [5] takes the rate of encountering other users, or simply the encounter rate, as the main metric for relay selection. A user will forward the information to an encountered user with higher encounter rate with the destination. [15] shares a similar idea with [5] while the encountering duration is taken as the relay selection metric. [21], [23], and [11] exploit the mobility properties of the users for relay selection. They intend to predict/assess the encounters among the users through mobility analysis, thus they can be considered as a variant of [5] and [15].

Despite these efforts to solve the relay selection problem in MSNs, none of them considers the fundamental problem of preventing information leakage to non-destination users [24, 2]. Although the encryption mechanism [20, 13] can be employed to protect the information transmission, reducing the information leakage probability is still critical as encryption mechanism is not “perfectly secure”. To tackle this challenge, a routing scheme is designed in [4] to reduce information leakage in social networks. However, the social relationship among users is not considered thus the transmission latency (i.e. routing efficiency) can not be guaranteed. To the best of our knowledge, we are the first to take both efficiency and security into consideration for relay selection in MSNs.

3. Problem description. Consider a mobile social network consisting of a set of users, denoted by \( N = \{1, 2, \cdots, n\} \). For each user \( i \in N \), a social relationship vector \( \Omega_i = \{\omega_{i1}, \omega_{i2}, \cdots, \omega_{in}\} \) is defined to describe its social relationship with others. A higher social relationship \( 0 \leq \omega_{ij} \leq 1 \) indicates a higher probability that user \( i \) meets user \( j \). Moreover, an overhearing probability \( p_i \) is defined for each user \( i \) to indicate the probability that \( i \) intercepts the information not for it. A higher \( 0 \leq p_i \leq 1 \) indicates a higher probability that \( i \) overhears others’ information.

We study the relay selection problem in the considered mobile social network. Assume that there is an information source \( I_s \in N \) trying to share its information with a set of destinations \( V_D \subseteq N \setminus I_s \). The information source and the destinations are assumed to be static. For the destinations within the source’s transmission range, the information source transmits the information directly to them. While for those destinations outside its transmission, the information source will explore the movements of the MSN users for information dissemination. This assumption is reasonable in kinds of practical scenarios. For example, a road side unit (RSU) in vehicular networks wants to share some traffic information or road information with other RSUs. If the destination RSU is within the source’s transmission range, direct transmission can be conducted. Otherwise, the source RSU should employ the moving vehicles and conduct relay transmission. In our work, we aim to find a relay path \( L_i \) for each destination \( i \) outside the transmission range of the information source. The relay path can be denoted by \( L_i = \{(i_l, i_{l+1})|l \in \{1, 2, \cdots, l_{in} - 1\}, i_1 = I_s, i_{l_{in}} = i\} \), with \((i_l, i_{l+1})\) being a relay link from user \( i_l \) to user \( i_{l+1} \) and \( l_{in} \) being...
the number of users on path \( L_i \). Our objective is to find the relay path \( L_i \) with small information leakage probability and transmission latency.

4. Prediction of the social relationship \( \omega_{ij} \) in MSNs. In order to find the relay path with low transmission latency, we need to know the meeting probabilities among the users, i.e. the social relationship \( \omega_{ij} \) among users. However, this information is hard to obtain due to the huge number of users in MSNs. To tackle this challenge, we will propose a social relationship prediction method based on analysing the mobility model of the MSN users.

4.1. Mobility model of MSN users. As mentioned in Section 1, the users in MSNs are the smartphones, PADs, and laptops carried by human beings. As a result, the mobility pattern of the MSN users corresponds to that of their holders. Empirical observation reveals that the mobility of human beings has the characteristic of both randomness and sociality. Randomness means people can change their moving speed and direction arbitrarily, while sociality signifies that there exists certain regulations for the movements. In this subsection, we consider both the randomness and the sociality, and focus on modeling the moving direction and the moving speed of the MSN users. We assume the users’ movement can be divided into a sequence of time intervals with length \( \epsilon \), the users do not change their moving direction or speed within one interval \([10]\).

Based on the basic idea of the virtual grid proposed in \([1]\), we participate the whole area \( \mathcal{A} \) into multiple square zones \( \mathcal{A} = \{ A_1, A_2, \ldots \} \) with each square \( A_\tau \in \mathcal{A} \) being indicated by the coordinates of its center \((x_\tau, y_\tau)\). Each zone \( A_\tau = (x_\tau, y_\tau) \) is determined by its functions, such as home, school, working place, gymnasium, shopping mall, restaurant, and so on. For different users, the probability of moving towards each zone is different. For example, the probability of user \( i \) moving towards \( A_\tau \) will be high if its home or working place is within \( A_\tau \). Let \( \zeta_{i\tau} \) being the probability for user \( i \) moving towards \( A_\tau \). It will be easy to get the value of \( \zeta_{i\tau} \) based on the analysis of user \( i \)’s hobbies and daily life traces. Through this way, we can get the probability function of choosing different moving directions for the users.

In normal life, human beings usually move with an average speed while the probability of moving with an extremely low or high speed is quite small. For example, the average speed for walking is about 5km/hour and the average speed for driving cars is about 100km/hour on high ways. According to this observation, it is reasonable to assume the moving speed of the users follow normal distribution. Let \( v_i \) denote the moving speed of user \( i \), then \( v_i \sim \mathcal{N}(\mu, \sigma) \). Here \( \mathcal{N}(\mu, \sigma) \) is a normal distribution with the mean being \( \mu \) and the standard deviation being \( \sigma \). In our model, we define the average moving speed of the users as \( \mu \). According to the 3\( \sigma \) rules of the normal distribution, \((\mu - 3\sigma, \mu + 3\sigma)\) can be considered as the value space for \( v_i \). Since the minimum speed of a user is 0, we let \( \mu - 3\sigma = 0 \) and get \( \sigma = \frac{\mu}{3} \). Therefore, we model the moving speed of user \( i \) as \( v_i \sim \mathcal{N}(\mu, \frac{\mu}{3}) \).

4.2. Prediction of \( \omega_{ij} \). Due to the fading property of wireless signals, each user has a limited communication range. As shown in Fig. 1, the communication range of user \( j \) is indicated by a dashed circle centered on \( j \). Assume that the radius of the communication range for each user is the same, we draw the following conclusion. Only if user \( i \) moves into the communication range of user \( j \), these two users can communicate with each other. Such case is defined as a meeting between user \( i \)
and user \( j \). As mentioned in Section 3, the social relationship \( \omega_{ij} \) indicates the probability that user \( i \) meets user \( j \). In this subsection, we will predict \( \omega_{ij} \) through deriving the meeting probability between \( i \) and \( j \) based on the mobility model described in last subsection.

Refer to Fig. 1, user \( i \) is a mobile user and user \( j \) is one of the destinations outside the communication range of the information source. Let us assume the current time interval is \( t_0 \), and the current location of user \( i \) is \((x_i^0, y_i^0)\) which is obtained via GPS or other techniques. Denote the current moving direction and speed of user \( i \) by \((x_a, y_a)\) and \( v_a \) respectively, with \((x_a, y_a)\) being the center of the zone that user \( i \) is currently moving towards. Since the users do not change their moving direction or speed within one time interval \([10]\), we can get the location of user \( i \) at the end of \( t_0 \) denoted by \((x_i^1, y_i^1)\). For next time interval \( t_1 \), user \( i \) will change its moving direction and speed according to the probability function of \( \zeta_i \) and the probability density function of \( v_i \) respectively. As shown in Fig. 1, it is possible for user \( i \) to meet user \( j \) (i.e., user \( i \) may move into the communication range of user \( j \)) if it’s moving direction in \( t_1 \) is \((x_b, y_b)\) or \((x_c, y_c)\) while the meeting between user \( i \) and user \( j \) will never happen during \( t_1 \) if user \( i \)’s moving direction is \((x_d, y_d)\). More specifically, when the moving direction of user \( i \) is \((x_c, y_c)\), it will meet user \( j \) if it’s location at the end of \( t_1 \) is no nearer than \((x_{min}, y_{min})\) and no further than \((x_{max}, y_{max})\). Therefore, the minimum speed of user \( i \) should be

\[
v_{min} = \frac{\sqrt{(x_{min} - x_i^1)^2 + (y_{min} - y_i^1)^2}}{\epsilon}
\]

while the maximum speed of user \( i \) should be

\[
v_{max} = \frac{\sqrt{(x_{max} - x_i^1)^2 + (y_{max} - y_i^1)^2}}{\epsilon}
\]
Thus, when the moving direction of user $i$ is $(x_c, y_c)$, the probability that user $i$ meets user $j$ during $Int_1$ can be derived as

$$
\phi_{ij} = \int_{v_{\min}}^{v_{\max}} \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(v_i - \mu)^2}{2\sigma^2} \right) dv_i
= \frac{1}{2\sigma} \left( \text{erf} \left( \frac{v_{\max} - \mu}{\sqrt{2}\sigma} \right) - \text{erf} \left( \frac{v_{\min} - \mu}{\sqrt{2}\sigma} \right) \right)
$$

(3)

Here, $\sigma = \frac{v_0}{3}$ and $\text{erf}(\cdot)$ is the Gauss error function. Let $C$ be the set of moving directions that makes the meeting between user $i$ and $j$ possible. For example, $(x_s, y_b) \in C$, $(x_c, y_c) \in C$, while $(x_d, y_d) \notin C$. The social relationship between $i$ and $j$ can be predicted as

$$
\omega_{ij} = \sum_{(x_c, y_c) \in C} \zeta_{ic} \phi_{ij}
$$

(4)

Here, $\zeta_{ic}$ is the probability that user $i$ choose direction $(x_c, y_c)$ and $\phi_{ij}$ is the probability that user $i$ meet user $j$ when its moving direction is $(x_c, y_c)$.

5. **Design of the relay selection algorithm.** In this section, we model the relay selection problem as a network formation game, where the users, also referred to as the players in the game, have the discretion in forming the relay paths. Each user calculates its payoff as the difference between the benefit and the cost for jointing a specific relay path and decides its strategy that could maximize its own payoff. More specifically, we will define the payoff functions for the information source, the relays, and the destinations in Section 5.1, design the game evolving rules (that is the procedure of the relay selection algorithm) in Section 5.2, and prove the stability of the network structure formed by the proposed algorithm in Section 5.3.

5.1. **Payoff functions.** For any game, one basic problem is to inspire the users to participate. In this subsection, we design the payoff functions to capture the incentives for the information source, the relays, and the destinations.

First, we design the payoff function for the information source $I_s$. Let $N(I_s)$ denote the set of users that are within the transmission range of $I_s$. Then the destination set $V_D$ could be partitioned into two subsets: $V_D^{in} \subseteq N(I_s)$ and $V_D^{out} = V_D \setminus V_D^{in}$. For any destination $i \in V_D^{in}$, $I_s$ could build a direct transmission link $(I_s, i)$ and gain a unit of payoff through finishing the transmission to destination $i$. If $I_s$ selects a relay $j \neq i$ for destination $i$ instead of performing the direct transmission, $i$ will not be reached at this time interval. Thus, we assume that $I_s$ will gain a payoff of $-\infty$ if $I_s$ selects a relay $j \neq i$ for any $i \in V_D^{in}$. For any destination $i \in V_D^{out}$, $I_s$ can either hold the transmission until it moves into the destination’s transmission range or employ some other users, which are within its transmission range, as relays. For any destination $i \in V_D^{out}$, the probability that $I_s$ will meet $i$ can be denoted by $\omega_{I_s i}$. The probability that destination $i$ will be reached in next time interval if user $j \in N(I_s)$ is selected as the relay can be denoted by $\omega_{ji}$. Thus, when $\omega_{ji} > \omega_{I_s i}$, the benefit of employing user $j$ as the relay to destination $i$ can be defined as:

$$
B^*_i(j) = \omega_{ji} - \omega_{I_s i}
$$

(5)

However, everything has two sides. Compared with direct transmission, relay transmission may cause a certain probability of information leakage on the relay paths. For a certain relay, the information may be intercepted by itself or be overheard.
by the users within its transmission range. Therefore, the probability that the information for destination \( i \) is leaked at relay \( j \in L_i \) can be defined as:

\[
C_i^r(j) = \begin{cases} 
1 -(1 - p_j) \prod_{k \notin V_D} (1 - \omega_{jk} p_k), & \text{if } j \notin V_D; \\
1 - \prod_{k \notin V_D} (1 - \omega_{jk} p_k), & \text{if } j \in V_D.
\end{cases}
\] (6)

Where \( p_j \) is the overhearing probability of user \( j \).

Calculating the difference between \( B_i^s(j) \) and \( C_i^r(j) \), we get the payoff earned by \( I_s \) for employing user \( j \) as a relay to destination \( i \). Taking into account all the discussed cases, we define the payoff function of the information source as follows:

\[
U_i^s(j) = \begin{cases} 
1 & \text{if } i \in N(I_s), j = i; \\
-\infty & \text{if } i \in N(I_s), j \neq i; \\
B_i^s(j) - C_i^r(j) & \text{if } i \notin N(I_s), j \in N(I_s).
\end{cases}
\] (7)

Here, the first row indicates the case where destination \( i \in V_D^in \) and \( I_s \) performs the direct transmission; the second row indicates the case where destination \( i \in V_D^in \) and \( I_s \) selects a relay \( j \neq i \) for \( i \); and the third row indicates the case where \( i \in V_D^out \) and \( I_s \) selects a relay \( j \in N(I_s) \) for \( i \).

When considering the network formation from the view point of the relays, the wages for acting as a relay and the cost for maintaining the relay links are two major concerns. In this work, we assume that each source-destination pair \((I_s, i)\), \( i \in V_D \), provides a wage budget of \( \alpha \).

Each relay \( j \) on path \( L_i \) earns a wage \( B_i^r(j) \) through sharing the total budget according to a predefined rule, which will be detailed in next subsection. As to the cost, we assume that a cost \( \beta \) is needed for user \( j \) to maintain each link. Then, the cost of user \( j \) for acting as a relay to destination \( i \) can be denoted by

\[
C_i^r(j) = k_{ij} \cdot \beta
\] (8)

Where \( k_{ij} \) is the number of links that should be added by \( j \) to perform the relay function for \( i \).

Therefore, the payoff function of user \( j \) for acting as a relay to destination \( i \) can be defined as:

\[
U_i^r(j) = B_i^r(j) - C_i^r(j)
\] (9)

Finally, we define the payoff function for the destinations. For any destination \( i \in V_D \), the payoff function \( U_i^d \) is defined as:

\[
U_i^d = \begin{cases} 
1 & \text{if destination } i \text{ is reached; } \\
0 & \text{otherwise. }
\end{cases}
\] (10)

That is, destination \( i \in V_D \) earns a unit of payoff if \( i \) is reached and earns nothing otherwise.

5.2. Relay selection based on the network formation game. In the network formation game, the objective of each player (user) is to maximize its payoff \([12]\). In this subsection, we propose the relay selection algorithm based on the network formation game described above. Relay selection consists of \( M \) rounds, with \( M \) being the maximum number of hops on the relay paths. In the \( m \)-th round, \( m \in \{1, 2, \cdots, M\} \), the \( m \)-th hop relay links for those destinations, which have not been reached, are formed.

For the first round, i.e., \( m = 1 \), the information source searches within its transmission range for the destinations. For each \( i \in V_D^in \), \( I_s \) forms the relay path \( L_i = \{ (I_s, i) \} \) and transmits the information to \( i \) directly. Then, the following steps are performed to form the first-hop relay paths for those destinations in \( V_D^out \):

...
**Step 1**: $I_s$ broadcasts an employment information, including the set $V_{D_out}^s$ and the value of $\alpha$, to those users belong to $N(I_s)$.

**Step 2**: Each user $j \in N(I_s)$ feedbacks a relationship list $\Omega_j = \{\omega_j|i \in V\}$ to $I_s$.

**Step 3**: After receiving the replies, $I_s$ selects a relay $j^*_s \in N(I_s)$ for each destination $i \in V_{D_out}^s$ according to the following rule:

$$j_s^* = \arg \max_{j \in N(I_s)} U^s_i(j)$$  \hfill (11)

Where $U^s_i(j) = B^s_i(j) - C^s_i(j)$ can be calculated according to (5), (6), and (7).

Then, $I_s$ constructs an employment list $E_j$ for each $j \in N(I_s)$.

$$E_j = \{i|j_s^* = j\}$$  \hfill (12)

We should notice that $E_j = \emptyset$, if user $j \in N(I_s)$ is not selected as a relay for any $i \in V_{D_out}^s$. For any user $j \in N(I_s)$ with nonempty $E_j$, $I_s$ sends the employment list to it.

**Step 4**: If user $j$ receives a nonempty employment list $E_j$ from $I_s$, it accepts to act as relays for a subset of destinations $V^*_j \subseteq E_j$, and feedbacks $V^*_j$ to $I_s$. $V^*_j$ is determined according to the following rule:

$$V^*_j = \{\forall i \in V^*_j|U^*_i(j) > 0\}$$  \hfill (13)

Where $U^*_i(j) = B^*_i(j) - C^*_i(j)$ is the payoff of user $j$ for acting as a relay to destination $i$. $C^*_i(j)$ can be calculated according to equation (8) and $B^*_i(j) = \alpha$ in the first round since $j$ is the only relay for destination pair $(I_s, i)$ now.

**Step 5**: After receiving the feedbacks, $I_s$ constructs the relay links for any $i \in V^*_j$ as $L_i = \{(I_s, j), (j, i)\}$, and modifies the set of destinations need to be relayed as $V_{D_out}^s = V_{D_out}^s \setminus \bigcup V^*_j$. Then, it performs Step 1 to Step 5, until $V_{D_out}^s = \emptyset$. It should be noticed that the relays which have rejected the employment of $I_s$ will not be considered again.

After the first round is finished, the first hop relays of all the destinations have been employed (if necessary). Denote the set of all the first hop relays by $R_1$. The information source $I_s$ sends the information to the relays in $R_1$. Once the relays in $R_1$ have accepted the employment, they have the responsibilities to transmit the information to the destinations. Otherwise, they will be punished by a tough punishment mechanism.

Then, our relay selection algorithm selects the second-hop relays in the second round. In this round, each first hop relay $j \in R_1$, which has the responsibility to transmit the information to the destinations in $V^*_j$, can be considered as a virtual information source $I_{s,j} = j$ with a set of destinations $V^*_j$. Each virtual information source $I_{s,j}$ just needs to perform Step 1-Step 5 as in the first round. The only difference we should notice is that the declared value of $\alpha$ in Step 1 is changed in the second round. Since relay $I_{s,j}$ has accepted the employment of $I_s$, it has the responsibility to relay the information to destination $i \in V^*_j$. When $i$ is out of the transmission range of $j$, $j$ is motivated to employ a second hop relay for help. However, $j$ is a selfish user who has to consider its own benefits. Therefore, $j$ would not give all the wages it earns from $I_s$ to the second hop relay. In our work, we assume that $j$ declares a wage budget $\alpha_2 = \frac{\alpha}{2}$ in Step 1. Therefore, the benefit earned by a second hop relay $j_2$ will be $B^*_i(j_2) = \alpha_2 = \frac{\alpha}{2}$ in Step 4.

Similarly, the $m$-th hop relays could be selected independently with the selection of the $(m - 1)$-th hop relays. Here, the $(m - 1)$-th relays are considered as the
virtual information sources, the wage budget $\alpha_m$ is declared to be $\alpha_m = \frac{2\alpha}{\alpha + \beta}$ in Step 1, and the benefit earned by a $m$-th hop relay $j_m$ is $B'_m(j_m) = \frac{\alpha}{\alpha + \beta}$ in Step 4. Finally, the relay selection algorithm ends when all the destinations of the information source $I_s$ are reached.

5.3. Stability analysis. Performing our relay selection algorithm, a relay path is formed for each destination. The totally $|V_D|$ relay paths construct a network structure, denoted by $g$. In this subsection, we prove the stability of the formed network structure.

According to [9], there are two kinds of stability, that is, pairwise stability and strong stability, in the network formation game. The definitions of these two kinds of stability are given as follows:

**Definition 5.1.** A network structure $g$ is **pairwise stable** with respective to the payoff function $U$ if

1. for any edge $(i, j) \in g$, $U_i(g) \geq U_i(g - (i, j))$ and $U_j(g) \geq U_j(g - (i, j))$, and
2. for any edge $(i, j) \notin g$, if $U_i(g + (i, j)) > U_i(g)$ then $U_j(g + (i, j)) < U_j(g)$

Here, $U_i(g)$ denotes the payoff earned by user $i$ in network structure $g$.

**Definition 5.2.** A network structure $g$ is **strong stable** with respective to the payoff function $U$ if for any other network structure $g'$, which is obtained via deviation from $g$, there exists at least one player (user) $i$ such that $U_i(g') < U_i(g)$.

It could be found from the definitions that pairwise stability only considers deviations on a single link at a time, while strong stability allows for more complicated deviations on groups of links. In this paper, we ignore the collusion among users and only consider the deviations on a single link at a time. Based on this assumption, we prove that the proposed relay selection algorithm is pairwise stable as follows.

**Theorem 5.3.** The network structure $g$ formed by the relay selection algorithm is pairwise stable.

**Proof.** First, we prove the first condition of pairwise stability. According to the description in Section 3, we use edge $(i_{m-1}, i_m) \in g$ to denote the relay link between the $(m-1)$-th hop relay and the $m$-th hop relay on relay path $L_i$. Then, we consider four cases.

1. **Case 1.** $i_{m-1} = I_s$, $i_m = i$. Edge $(i_{m-1}, i_m)$ indicates that destination $i_m$ is within the transmission range of the information source $i_{m-1} = I_s$. Obviously, it is against common sense to break the direct transmission link when $i_m$ is reachable by $i_{m-1}$. Theoretically, user $i_{m-1}$ earns a payoff of 1 when edge $(i_{m-1}, i_m)$ exists (refer to the first row of equation (7)). If edge $(i_{m-1}, i_m)$ is removed, $i_{m-1}$ should find another user $j \neq i_m$ to relay the message. According to the second row of (7), $i_{m-1}$ will earns a payoff of $-\infty$ in such a case. Therefore, $U_{i_{m-1}}(g) > U_{i_{m-1}}(g - (i_{m-1}, i_m))$. As for the destination $i_m$, it earns the payoff of 1, i.e., getting information from $i_{m-1}$, if edge $(i_{m-1}, i_m)$ exists, and earns nothing if edge $(i_{m-1}, i_m)$ is removed. Thus, $U_{i_m}(g) > U_{i_m}(g - (i_{m-1}, i_m))$.

2. **Case 2.** $i_{m-1} = I_s$, $i_m \neq i$. Edge $(i_{m-1}, i_m) = (I_s, i_m)$ indicates that user $i_m$ is a relay selected by the information source. According to (7) and (9), $i_{m-1}$ and $i_m$ earn a payoff of $U_{i_{m-1}}(g) = U_{i_{m-1}}(g) = B'_i(i_m) - C'_i(i_m)$ and $U_{i_m}(g) = U_{i_m}(g) = B'_i(i_m) - C'_i(i_m)$, respectively. It can be seen from (13) that $U_{i_m}(g) = U_{i_m}(i_m) > 0$. If edge $(i_{m-1}, i_m)$ is removed, user $i_m$ can not earn the payoff $U'_{i_m}(i_m)$ anymore. In other words, we can say that user $i_m$ earns a zero payoff now. Therefore, $U_{i_m}(g) >$
As for the information source $i_{m-1} = I_s$, a new relay $i'_m$ should be selected to replace $i_m$ if edge $(i_{m-1}, i_m)$ is deleted. According to (11), $U_i^s(i_m) > U_i^s(i'_m)$. Therefore, $U_{i_{m-1}}(g) > U_{i_{m-1}}(g - (i_{m-1}, i_m))$.

(3) **Case 3.** $i_{m-1} \neq I_s$, $i_m \neq i$. As described in Section 5.2, $i_{m-1}$ can be considered as a virtual information source in this case. Similar to the analysis in Case 2, we find that $U_{i_{m-1}}(g) > U_{i_{m-1}}(g - (i_{m-1}, i_m)), U_{i_m}(g) > U_{i_m}(g - (i_{m-1}, i_m))$.

(4) **Case 4.** $i_{m-1} \neq I_s$, $i_m = i$. In this case, $i_{m-1}$ can also be considered as a virtual information source. Similar to the analysis in Case 1, we have $U_{i_{m-1}}(g) > U_{i_{m-1}}(g - (i_{m-1}, i_m)), U_{i_m}(g) > U_{i_m}(g - (i_{m-1}, i_m))$.

Summarizing these four cases, we conclude that $g$ satisfies the first condition of pairwise stability.

Then, we prove the second condition of pairwise stability. For any edge $(p, q) \notin g$, there are two cases.

1. **Case 1:** $\forall i \in V_D, p \notin L_i, q \notin L_i$. This case indicates the situation where neither $p$ or $q$ belongs to any relay path. Forming such an edge is meaningless in the relay selection process. Thus, we do not consider this case in our work.

2. **Case 2:** There is a pair of $i$ and $m \in \{1, 2, \cdots, M\}$ such that $p = i_{m-1}$. If $q = i_m$, edge $(i_{m-1}, i_m)$ already exists in the network structure $g$. Otherwise, adding edge $(p, q)$ means replace relay $i_m$ with another user $q$. In other words, edge $(i_{m-1}, i_m)$ should be deleted before adding edge $(p, q)$. According to the proof of the first condition, deleting edge $(i_{m-1}, i_m)$ will lead to a decreased payoff for user $i_{m-1}$. Similarly, the payoff of user $i_m$ will be decreased when there is a pair of $i$ and $m \in \{1, 2, \cdots, M\}$ such that $q = i_m$ and $p \neq i_{m-1}$. Therefore, the network structure $g$ formed by our relay selection algorithm also satisfies the second condition of pairwise stability.

Based on all the above analysis, we draw the conclusion that the network structure $g$ formed by our relay selection algorithm is pairwise stable. \qed

## 6. Simulation

In this section, we compare the performance of the proposed relay selection algorithm with the random relay selection algorithm (Rand), the relationship based relay selection algorithm (Relation), and the leakage probability based relay selection algorithm (Leakage), in MATLAB, using both synthetic trace and real-world trace. In our simulation, we consider a time-slotted system with a set of users, denoted by $N = \{1, 2, \cdots, n\}$. For each formed relay path, we define the transmission latency as the number of slots cost for reaching the destination, and define the maximum leakage probability as the maximum leakage probability of the users on the relay path. The relay selection algorithms are conducted for 3000 times, the average latency (A-Latency) and the average maximum leakage probability (A-MLP) are used to evaluate the performance of the relay selection algorithms. Our relay selection algorithm is denoted by ESRS for short in the simulation study. The three schemes for comparison are briefly introduced as follows,

- **Rand:** the next hop relay is randomly selected.
- **Relation:** the next hop relay is selected as the user having the maximum social relationship with the destination.
- **Leakage:** the next hop relay is selected as the user with the minimum leakage probability.

### 6.1. Simulation study using synthetic $\zeta_{i\tau}$

As defined in Section 4.1, $\zeta_{i\tau}$ is the probability for user $i$ moving towards zone $A_{i\tau}$. Different users have different
probabilities for moving towards each zone. As we know, many human activities in social networks can be modeled by the power law distribution [16, 14, 28, 7, 27]. Therefore, for each user, we model the probability that it moves towards different zones as a power law distribution. That is, a user moves to several zones (such as the home, the working place, and the favorite restaurants) with high probability and rarely moves to other zones. We assume there are 100 zones \{A_1, A_2, \cdots, A_{100}\}
in a $100m \times 100m$ area $\mathcal{A}$, with each zone being a $10m \times 10m$ square. 100 users move within the whole area. For each user $i$, the probability function of choosing different moving directions is modeled as a power law distribution: $\zeta_i = \tau^{-k_r}$, with $\tau \in \{1, 2, \cdots, 100\}$ being the index of a zone, $k_r$ being the exponent of the distribution, and $\sum \tau^{-k_r}$ being the normalization factor.

The probability that a user launches a malicious attack is also modeled by a power law distribution. Specifically, we define $p_i = C_l i^{-k_l}$, with $i \in N$ being the user index, $N$ being the set of users, $k_l$ being the exponent, and $C_l$ being a constant indicating the maximum value of $p_i$, respectively. The effect of $k_r$ on $\zeta_i$, the effects of $C_l$ and $k_l$ on $p_i$ can be shown by Fig. 2.

In our simulation, we set the parameters according to Table 1 and we get the simulation results shown in Table 2.

**Table 2. Simulation Results**

|          | ESRS | Relation | Leakage | Rand |
|----------|------|----------|---------|------|
| A-Latency| 16.2 | 15.4     | 17.9    | 30.4 |
| A-MLP    | 0.38 | 0.63     | 0.35    | 0.72 |

6.2. **Simulation study using real-world meeting probability.** We take the meeting probability obtained from real trace [17], which is conducted at University of St Andrews, as the social relationship between users. To the best of our knowledge, there is no available trace indicating the probability that a user launches a malicious attack. In the real trace simulation, we change the exponent $k_l$ of the power law distribution $p_i = C_l i^{-k_l}$ from 1 to 3 to investigate the performance of the schemes under different attack probabilities.

The simulation results using the real-world trace are given in Fig. 3. As shown in Fig. 3(a), Relation always gets the best A-Latency performance as it selects the relays based solely on the relationship (i.e. the encounter rate) among the users. The A-Latency performance of Rand and Leakage are random since the encounter rate is not considered for relay selection. As for the relay selection algorithm, its A-Latency performance approaches that of Relation with the increase of $k_l$. The reason is explained as follows. According to Fig. 2, the probabilities of users launching malicious attacks decrease with the increasing of $k_l$. In this case, the relationship among the users play a vital role for relay selection. Therefore, our relay selection algorithm will find the relays with higher encounter rates with the destination and the A-Latency performance will approach that of Relation. The comparison of the A-MLP performance is given in Fig. 3(b). It can be seen that the relay selection algorithm maintains a similar A-LMP performance with that of Leakage which selects the relays based solely on the leakage probabilities of the users.

7. **Conclusion.** This paper tackles the challenge of designing an efficient and secure relay selection algorithm for MSNs. We propose a novel mobility model for MSN users considering both the randomness and the sociality of the movements. Through analysing of the mobility model, the social relationship among users is predicted for helping find relay paths with low transmission latency. We propose a network formation game based relay selection algorithm by defining the payoff.
functions of the users and designing the game evolving rules. The stability of the network structure formed by our relay selection algorithm is proved and the performance of the algorithm is compared with other algorithms through extensive simulation study. The simulation results using both synthetic trace and real-world trace show that our algorithm outperforms other algorithms by enabling lower latency and lower information leakage probability communication in MSNs.

Acknowledgments. The authors would like to thank the support from the National Natural Science Foundation of China (61602062, 61702056), the Natural Science Foundation of Jiangsu Province (BK20160410), the Educational Commission of JiangSu Province (17KJB520001), the Provincial Key Laboratory for Computer Information Processing Technology, Soochow University (KJS1521, KJS1522).

REFERENCES

[1] J. A. Bazerque and G. B. Giannakis, Distributed spectrum sensing for cognitive radio networks by exploiting sparsity, *IEEE Transactions on Signal Processing*, 58 (2010), 1847–1862.
[2] Z. Cai, Z. He, X. Guan and Y. Li, Collective data-sanitization for preventing sensitive information inference attacks in social networks, *IEEE Transactions on Dependable and Secure Computing*, 15 (2018), 577–590.
[3] Z. Cai and X. Zheng, A private and efficient mechanism for data uploading in smart cyber-physical systems, *IEEE Transactions on Network Science and Engineering*.
[4] W. Cheng, D. Wu, X. Cheng and D. Chen, Routing for information leakage reduction in multi-channel multi-hop ad-hoc social networks, *Lecture Notes in Computer Science*, 7405 (2012), 31–42.
[5] R. Ciomaru, C. Dobre and V. Cristea, Sprint: Social prediction-based opportunistic routing, in *IEEE WoWMoM*, 2013, 1–7.
[6] V. Erramilli, M. Crovella, A. Chaintreau and C. Diot, Delegation forwarding, in *Proc. MobiHoc*, 2008, 251–260.
[7] Z. He, Z. Cai and J. Yu, Latent-data privacy preserving with customized data utility for social network data, *IEEE Transactions on Vehicular Technology*, 67 (2018), 665–673.
[8] Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun and Y. Li, Cost-efficient strategies for restraining rumor spreading in mobile social networks, *IEEE Transactions on Vehicular Technology*, 66 (2017), 2789–2800.
[9] M. O. Jackson, A survey of models of network formation: Stability and efficiency, *Cambridge University Press*, (2010), 11–57.
[10] T. Jing, J. Zhou, H. Liu and Z. Zhang, Soroute: a reliable and effective social-based routing in cognitive radio ad hoc networks, EURASIP Journal on Wireless Communications and Networking, 2014 (2014), 200–214.

[11] S. K. Kim, J. H. Yoon, J. Y. Lee, G. Y. Jang and S. B. Yang, A cooperative forwarding scheme for social preference-based selfishness in mobile social networks, Wireless Networks, 22 (2016), 537–552.

[12] W. Li, X. Cheng, T. Jing and X. Xing, Cooperative multi-hop relaying via network formation games in cognitive radio networks, in IEEE INFOCOM, 2013, 971–979.

[13] Y. Liang, Z. Cai, Q. Han and Y. Li, Location privacy leakage through sensory data, Security and Communication Networks, 2017 (2017), Article ID 7576307, 12 pages.

[14] J. Lu, Z. Cai, X. Wang, L. Zhang, P. Li and Z. He, User social activity-based routing for cognitive radio networks, Personal and Ubiquitous Computing, 22 (2018), 471–487.

[15] W. Moreira, P. Mendes and S. Sargento, Opportunistic routing based on daily routines, in IEEE WoWMoM, 2012, 1–6.

[16] L. Muchnik, S. Pei, L. C. Parra, S. D. S. Reis, J. S. Andrade Jr., S. Havlin and H. A. Makse, Origins of power-law degree distribution in the heterogeneity of human activity in social networks, Scientific Reports, 3.

[17] I. Parris and F. Ben Abdesslem, Crawdad trace/social network analysis/st_andrews/locshare/2010/sta1, Downloaded from http://crawdad.org//download/st_andrews/locshare/locshare-StA1.tar.gz, Nov. 2010.

[18] T. Spyropoulos, K. Psounis and C. Raghavendra, Efficient routing in intermittently connected mobile networks: The multiple-copy case, IEEE/ACM Transactions on Networking, 16 (2008), 77–90.

[19] A. Vahdat and D. Becker, Epidemic Routing for Partially-Connected ad Hoc Networks, Technical report, Duke University, 2000.

[20] J. Wang, Z. Cai, Y. Li, D. Yang, J. Li and H. Gao, Protecting query privacy with differentially private k-anonymity in location-based services, Personal and Ubiquitous Computing, 1–17.

[21] S. Wang, M. Liu, X. Cheng, Z. Li, J. Huang and B. Chen, Hero-a home based routing in pocket switched networks, in Proc. Wireless Algorithms, Systems, and Applications, 2012, 20–30.

[22] J. Wu and Y. Wang, Social feature-based multi-path routing in delay tolerant networks, in Proc. IEEE INFOCOM, 2012, 1368–1376.

[23] J. Wu, M. Xiao and L. Huang, Homing spread: Community home-based multi-copy routing in mobile social networks, in Proc. IEEE INFOCOM, 2013, 2319–2327.

[24] X. Xing, T. Jing, W. Zhou, X. Cheng, Y. Huo and H. Liu, Routing in user-centric networks, IEEE Communications Magazine, 52 (2014), 44–51.

[25] X. Zheng, Z. Cai, J. Li and H. Gao, Location-privacy-aware review publication mechanism for local business service systems, in IEEE INFOCOM, 2017.

[26] X. Zheng, Z. Cai and Y. Li, Data linkage in smart iot systems: A consideration from privacy perspective, IEEE Communications Magazine.

[27] X. Zheng, Z. Cai, J. Yu, C. Wang and Y. Li, Follow but no track: Privacy preserved profile publishing in cyber-physical social systems, IEEE Internet of Things Journal, 4 (2017), 1868–1878.

[28] X. Zheng, G. Luo and Z. Cai, A fair mechanism for private data publication in online social networks, IEEE Transactions on Network Science and Engineering.

Received July 2018; revised August 2018.

E-mail address: xing@cslg.edu.cn
E-mail address: gfsun@cslg.edu.cn
E-mail address: jinyong@cslg.edu.cn
E-mail address: yigetangwenyi@outlook.com
E-mail address: cheng@gwu.edu