Yuan PY, Ma HD, Duan PR. Impact of strangers on opportunistic routing performance. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 28(3): 574–582 May 2013. DOI 10.1007/s11390-013-1357-x

Impact of Strangers on Opportunistic Routing Performance

Pei-Yan Yuan (袁培燕), *Member*, *ACM*, Hua-Dong Ma* (马华东), *Member*, *CCF*, *ACM*, *IEEE*
and Peng-Rui Duan (段鹏瑞)

Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China

E-mail: peiyan@htu.cn; {mhd, dpr}@bupt.edu.cn

Received May 3, 2012; revised January 18, 2013.

Abstract Routing is one of the challenging tasks in Delay Tolerant Networks (DTNs), due to the lack of global knowledge and sporadic contacts between nodes. Most existing studies take a greedy scheme in data forwarding process, i.e., only nodes with higher utility values than current carriers can be selected as relays. They lack an in-depth investigation on the main features of the optimal paths in Epidemic. These features are vital to any forwarding scheme that tends to make a trade-off between packet delivery delay and cost. This is mainly because Epidemic provides an upper bound on cost and a lower bound on delivery delay. Therefore, a deep understanding of these features is useful to make informed forwarding decisions. In this paper, we try to explore these features by observing the roles of different social relationships in the optimal paths through a set of real datasets. These datasets provide evidence that strangers have two sides in data forwarding process, and that the importance of strangers shows a decreasing trend along the forwarding paths. Using this heuristic knowledge, we propose STRON, a distributed and lightweight forwarding scheme. The distributed feature makes it very suitable for opportunistic scenarios and the low communication and computation features make it easy to be integrated with state-of-the-art work. The trace-driven simulations obviously confirm its effectiveness, especially in terms of packet delivery delay and cost.

Keywords stranger, forwarding mechanism, social relationship, Delay Tolerant Network

1 Introduction

One of the main characteristics of Delay Tolerant Networks (DTNs) is that an end-to-end path between the source and destination rarely (if ever) exists at any moment, which makes routing very challenging in DTNs[1-2]. In this work, we focus on the influence of strangers on routing performance within a pure darkness environment, where the mobility of nodes cannot be acquired in advance and each node depends only on itself to locally estimate the forwarding metric to destination.

Obviously, the Epidemic scheme[3] is a potential solution to deliver messages under the above scenario because it tries to send each message over all possible paths (i.e., multiple copies) in the network. Thus, the message will be successfully received so long as one of the copies reaches the destination. Whereas, the immoderate spraying will incur a high price of system resources such as the splurge on energy and buffer space and the rapid consumption of available bandwidth. Considering the limited computation and storage ability of mobile sensing devices, this scheme is obviously not desirable for opportunistic scenarios.

These deficiencies of Epidemic have motivated researchers to develop other novel routing algorithms. For these algorithms, the main issue is which forwarding scheme can achieve the best trade-off between the packet delivery delay and cost. Most of the proposed algorithms that control message copies try to infer the delivery probability of nodes to the destination by making use of two kinds of contexts: physical contact information of nodes (e.g., the number of contacts[4], contact frequency/locations[5-7]) and social metrics in the network (e.g., community[8], centrality[9-10] and similarity[11]). Such heuristic knowledge helps the algorithms to make informed forwarding decisions on which nodes to relay. However, few studies focus on the main

Regular Paper

This work is supported by the National Basic Research 973 Program of China under Grant No. 2011CB302701, the National Science Fund for Distinguished Young Scholars of China under Grant No. 60925010, the National Natural Science Foundation of China under Grant Nos. 61133015, 61003280, and 61272517, the Funds for Creative Research Groups of China under Grant No. 61121001, the Program for Changjiang Scholars and Innovative Research Team in University of China under Grant No. IRT1049, and the Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20120005130002.

*Corresponding Author

features of the optimal paths in Epidemic. Note that these features are vital to any forwarding scheme that tends to make the trade-off. This is mainly because Epidemic provides an upper bound on cost and a lower bound on delivery delay. Therefore, a deep understanding of these features is useful to make informed forwarding decisions.

To verify this, we classify nodes into four categories: strangers, familiar strangers, friends, and community partners. We observe their roles in Epidemic and greedy schemes through a set of real datasets. These datasets provide the following evidences. 1) Strangers have two sides in data forwarding process. On the one hand, there exist a lot of strangers in the shortest paths of Epidemic. This means that strangers play a positive role in data forwarding process. On the other hand, the strangers have speeded up the dissemination process by infecting many other nodes, which increases the routing cost as well. 2) The greedy scheme reduces the cost by replacing part of strangers which were initially present in the shortest paths of Epidemic with familiar strangers or community partners (refer to Subsection 3.1). As a result, it increases the delivery delay. 3) The importance of strangers shows a decreasing trend along the forwarding paths.

Motivated by these observations, we try to improve the performance of greedy scheme by employing a few strangers. We propose a stranger-oriented forwarding algorithm, called STRON, to achieve this goal. It mainly includes the following two features. 1) Distributed feature: STRON does not need any global knowledge. Each node only records the number of contacts and the intra-contact time between itself and any other node. 2) Lightweight feature: STRON uses the mean of the number of contacts and that of the intra-contact time as thresholds to identify social relationships. It does not need to store additional information unlike the traditional community-detection mechanism does[8,12]. This feature makes it easy to be integrated with the state-of-the-art work. We summarize our main contributions as follows:

• We explore the main features of the optimized paths in Epidemic and try to integrate them into the opportunistic forwarding scheme.

• We observe that strangers have two sides in data forwarding process and the importance of strangers shows a decreasing trend along the forwarding path. Using these heuristic knowledge, we develop a distributed routing algorithm, called STRON, to improve the data forwarding performance.

• We conduct extensive simulations to compare our routing metric with the state-of-the-art work through real DTNs traces. The simulation results show that our algorithm largely improves the performance, especially in terms of cost.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 shows the preliminary work. We explore the main features of the optimized path in Section 4. Section 5 discusses how to integrate the main features into STRON, followed by a performance evaluation in Section 6. Finally, we conclude our paper and discuss some future research areas in Section 7.

2 Related Work

Routing messages through intermittently connected network is challenging. In the past few years, researchers have proposed several strategies to solve this issue. According to the contexts they exploited, these solutions can be classified into the following categories.

2.1 Routing with Extra Nodes

Several projects try to deliver messages by the help of extra nodes which are called data MULEs or message ferries[13-17].

The authors of [13] first utilized mobile MULEs to collect data for sparse sensor networks. The MULEs have large storage capacities and renewable power, which make them have the ability to buffer data for a relatively long time and thus can be used in a delay tolerant environment. The data MULEs scheme provides opportunistic forwarding between static sensor nodes and the mobile MULEs, whereas, they neglect the influence of the mobility of MULEs on routing performance. On the contrary, the message ferries scheme (e.g., [14-16]) goes further in exploiting special mobile nodes which are called ferries to deliver messages by assuming control or influence over ferries movements. For example, the authors of [14] proposed the idea of exploiting controlled mobility of extra nodes to facilitate message transmission in disconnected mobile Ad hoc networks. Zhao, Ammar, and Zegura[15] designed two kinds of no-random movements to forward data in DTNs. The first is node initiated mobility, where ferries move around the deployed region according to conventional routes. The second is ferry initiated mobility, where a ferry will adjust its trajectory to meet up the node when receiving a request from that node. They also evaluated the trade-off between the incurred cost of extra ferries and the improved performance in [16]. Besides, the authors of [17] further relaxed the assumptions used in [15] and [16]. The ferries can navigate themselves intelligently only based on partial observations and statistical information of nodes mobility. Recently, the authors of [7] employed the static nodes placed in the system "hot region" to relay messages: if the message enters the "hot region", the static node sprays one replica of the message to any other nodes it encounters, otherwise, the message is sprayed in a binary way[18].

2.2 Routing with Periodic Mobility of Nodes

In some particular scenarios (e.g., bus transportation system[19] and interplanetary internet[20]), the mobility patterns of nodes have periodicity, which motivates researchers to design periodic information based routing scheme[21-23]. Most of them used a modified Dijkstra algorithm to compute shortest paths between sources and destinations. The routing table was designed based on intermediate nodes along those paths. That is, each node has a global view on network structures. For instance, Merugu *et al.*[21] delivered messages over a space-time routing table, which was derived from the mobility of nodes and carried by each node. Jain *et al.*[22] computed the shortest paths between transceivers by utilizing the periodicity of nodes movements. Besides, the authors of [23] proposed a source routing in DTNs. They exploited the expected minimum delay (EMD) as forwarding metric and applied the Markov decision process to derive the EMD of messages at particular moments.

2.3 Routing with Partial Observations

Sometimes, it is difficult or impractical to acquire global information of the network. This is mainly because of the problems such as time-varying topology, privacy protection or selfishness of nodes. In these scenarios, different local contexts can be exploited to improve routing performance. Based on the contexts they use, they can be further classified into two subclasses.

Physical Contact Metric. It mainly includes the number of contacts, contact frequency and contact location. For example, MaxProp[4] updates the utility values between two nodes by a method called incremental ageing. When two nodes encounter, their utility values are incremented by 1 and then all other values are re-normalized. Using this method, MaxProp shows a better performance than protocols that have proactive knowledge. CAR (context aware routing) was proposed in [24], which exploits the context information such as the changing rate of neighbors of a node and its current energy level to estimate the delivery probability. In addition, the authors of [25] proposed PER, a prediction and relay algorithm for DTNs, which considers the time of a contact. Similarly, Lindgren *et al.*[5] presented PROPHET, in which the transitive property and an aging constant are both considered to try to accurately predict the probability of future encounters. Recent work[26] proposed an optimized probabilistic method (called OP in this paper) based on the brief contact. Furthermore, Leguay *et al.*[6] presented MobySpace, a high-dimensional Euclidean space constructed by the past motion patters of nodes.

Social Metric. With the popularization of smart hand-held devices, human mobility has been shown to have a big impact on the network performance. Considering this fact, researchers have recently focused on the social metrics underlying the network structure and exploited them to make smart forwarding decisions. For instance, the authors of [11] presented SimBet, which exploits neighbor's adjacency matrix to compute the centrality and similarity of nodes and then utilizes these social attributes to predict the best relay to the final destination. The adjacency matrixes should be swapped and updated each time two nodes have a contact. Similarly, Bubble and PeopleRank[9-10] forward messages to the popular nodes in the network.

3 Preliminaries

The goal of this paper is to develop a distributed forwarding algorithm for opportunistic scenarios based on the social relationships extracted from real human traces. In this section, we briefly introduce the taxonomy of social relationships and then present the greedy scheme.

3.1 Social Relationships

Using the taxonomy proposed in [27], we classify nodes into four categories:

• strangers: with short duration and low number of contacts.

• familiar stranger: with short duration and high number of contacts.

• friend: with long duration and low number of contacts.

• community partners: with long duration and high number of contacts.

According to [28], in this paper, we use the mean as threshold to identify such social relationships, for both the number of contacts and the intra-contact time. Mathematically, let random variables X_i and Y_i denote the number of contacts and the intra-contact time between node i and other nodes, respectively. Let $x_i(j)$ and $y_i(j)$ denote those between node i and any node j. Let $E(X_i)$ denote the mean of X_i and $E(Y_i)$ denote the mean of Y_i. Let N denote the set of nodes in the network, we have

$$E(X_i) = \frac{\sum\limits_{i,k\in N, i\neq k} x_i(k)}{\|N\|}, \tag{1}$$

$$E(Y_i) = \frac{\sum\limits_{i,k \in N, i \neq k} y_i(k)}{\|N\|}. \quad (2)$$

We call node j is a stranger to node i if $x_i(j)$ is smaller than $E(X_i)$ and $y_i(j)$ is smaller than $E(Y_i)$. The notations used in this paper are listed in Table 1.

Table 1. Notation Summary

Notation	Explanation
N	Set of nodes
$\|N\|$	Number of nodes
i, j	Two randomly chosen nodes
$x_i(j)$	Number of contacts between node i and node j
$y_i(j)$	Intra-contact time between node i and node j
m_d	Destination node of message m
$D_s(i, j)$	Degree of strangeness between two nodes
p_{ij}	Shortest path from i to j
$\mathcal{S}_{ij}$	Number of strangers in p_{ij}
$\mathcal{R}_{ij}$	Number of relays in p_{ij}

3.2 Greedy Scheme

In the past few years, researchers have proposed a large number of routing metrics in DTNs. Although they exploited different kinds of contexts (e.g., similarity[11], intra-contact time[5], and virtual community[8]), most of them take a greedy forwarding mechanism. That is, when two nodes have a contact, a node with a lower utility value to the destination will forward messages to nodes with higher utility values (for ease of presentation, we here use the term "$u_i(j)$" as a general representation for the utility function as shown in Algorithm 1). When node i meets node j, for any message m that i carries, if its destination m_d is node j, node i delivers it to node j and removes it from i's buffer. Otherwise, if node j does not hold this message, the two nodes swap their own utility value. If $u_i(m_d)$ is smaller than $u_j(m_d)$, node i forwards m to node j, where nodes i, j and $m_d \in N$.

Algorithm 1. Greedy Mechanism

Upon meeting up node j **do**
for any message m in i's buffer **do**
  if $m_d == j$ **then**
    deliverMsg(m), *remove*(m)
  else if $m \notin j$ **then**
      $i \leftarrow u_j(m_d)$
      if $u_i(m_d) < u_j(m_d)$ **then**
        forwardingMsg(m)
      end if
    end if
  end if
end for

4 Exploring the Main Features of the Shortest Paths in Epidemic and Greedy Scheme

This section explores the main features of the shortest paths in Epidemic and greedy scheme through two real datasets called KAIST[29-30] and PMTR[26]. Thirty-four volunteers carried the GPS devices (GPS 60CSx) from 2006-09-26 to 2007-10-03 and altogether 92 daily traces were gathered in KAIST. Each individual trace consists of a sequence of three-tuples (timestamp, X-coordinate, Y-coordinate), where a tuple denotes a stay point recorded every 30 seconds. In PMTR, 44 people and 5 fixed positions were chosen and 11 895 contact events were reported in 19 days. Table 2 summaries the main attributes of the two datasets.

Table 2. Dataset Statistics

	KAIST	PMTR
Radio range (m)	250	10
Number of nodes	34	49
Number of contacts in total	25 535	11 895
Radius (m)	18 650	3 500
Year	2006	2008

These datasets provide the following evidences.

4.1 Two Roles of Strangers in Data Forwarding Process

In this subsection, we analyze the social relationships between a relay and the final destination. We evaluate their importance in data forwarding process by computing their occurrence frequencies in the optimized paths and in the greedy paths, respectively.

Definition 1 (Optimized Path). *We call a path an optimized path if it is one of the shortest paths achieved by Epidemic.*

Definition 2 (Greedy Path). *We call a path a greedy path if it is one of the shortest paths achieved by greedy scheme.*

For each optimized path p (greedy path g), we count the number of each kind of relays and then average the results over all paths. For instance, let $\mathcal{O}$ denote the set of optimized paths. Let p_{ij} denote the shortest path from node i to node j ($p_{ij} \in \mathcal{O}$), $\mathcal{S}_{ij}$ denote the number of strangers which participate in p_{ij} and $\mathcal{R}_{ij}$ denote that of total relays. We use the ratio $\mathcal{R}_s$ to estimate the importance of the strangers along the optimal paths, and we have

$$\mathcal{R}_s = \sum_{\forall i \in N} \sum_{\forall j \in N, j \neq i} \frac{\mathcal{S}_{ij}}{\mathcal{R}_{ij}}. \quad (3)$$

Fig.1 shows the roles of the relationships in the optimal paths and greedy paths, respectively (where we use the term "Greedy1" to denote the greedy scheme based on the number of contacts and the term "Greedy2" to denote that with the intra-contact time). It is clear to see that strangers play a big role in Epidemic, which shows the power of crowd[31-32]. Interestingly, we also observe that familiar strangers (FStrangers) and community partners (Community) dominate the greedy paths. This phenomenon is in accordance with the nature of the greedy scheme, where only nodes with higher utility values to the destination can be selected as relays. We conjecture that this phenomenon is also the reason why the greedy scheme can make a better trade-off between the cost and the packet delivery delay.

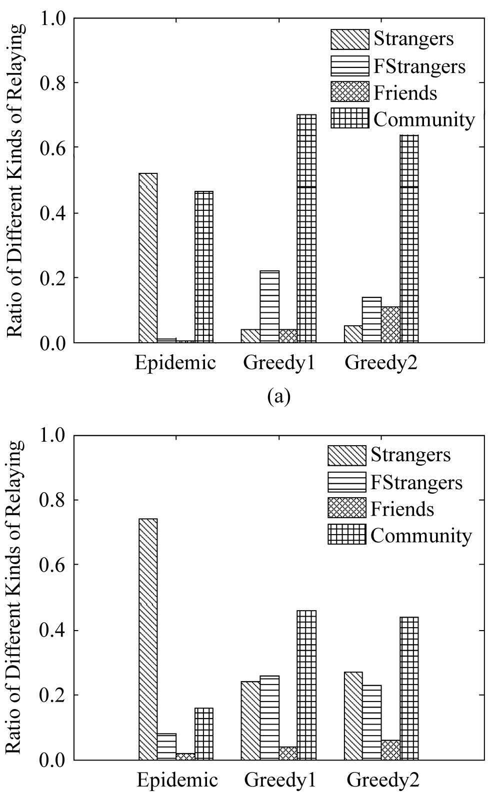


Fig.1. Roles of different social relationships in data forwarding process. (a) KAIST. (b) PMTR.

We present the two performance metrics in Table 3 and Table 4 respectively. We notice that the greedy scheme reduces the routing cost but increases the delivery delay at the same time. This is mainly because that the greedy scheme replaces part of the strangers which were initially present in the optimal paths with part of familiar strangers and community partners. For example, at KAIST, the ratio of strangers is over 50% in the optimal paths, whereas, the ratio of strangers almost decreases by 90% in the greedy paths. In the meantime, the number of familiar strangers and that of community partners increase by 20%.

Table 3. Statistics About Infected Ratio (Cost)

	Epidemic (%)	Greedy1 (%)	Greedy2 (%)
KAIST	97	27	31
PMTR	70	25	28

Table 4. Statistics About Mean Delivery Delay

	Epidemic	Greedy1	Greedy2
KAIST (s)	310.0	1 550.0	1 493.0
PMTR (h)	89.5	97.4	107.9

For Epidemic, it achieves the optimal delivery delay by infecting most of the nodes. We think the reason is that strangers bring messages to different parts of the network, which on the one hand increases the probability to meet the destination. On the other hand, it infects the nodes located in these regions. Hence, we should control the number of strangers to refrain the infection process. We next analyze the changes of the social relationship along the optimal paths.

4.2 Importance of Strangers' Decreasing Trend Along the Optimal Paths

This subsection explores the changes of social relationships at each step along the optimal paths as shown in Fig.2. When moving toward the final destination, the

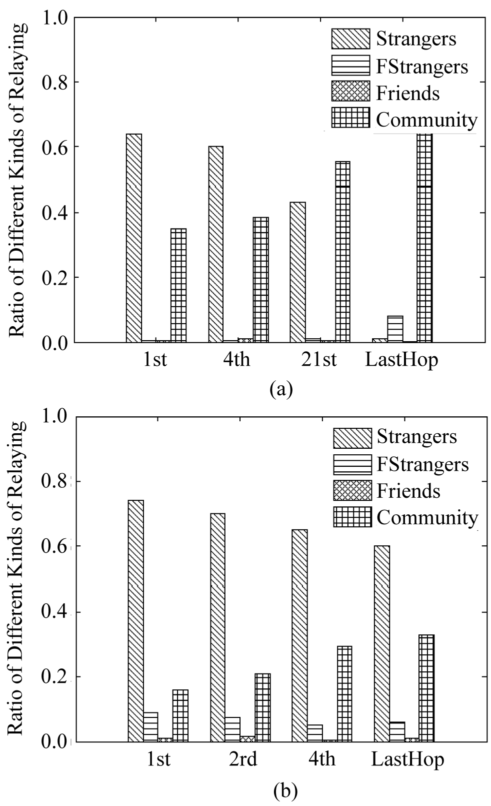


Fig.2. Changes of social relationships along the optimal paths. (a) KAIST. (b) PMTR.

relays become more and more familiar with the destination. In general, the importance of strangers shows a decreasing trend along the optimal paths, though there exist differences between different scenarios. For example, at PMTR, the ratio of strangers is still over 50% at the last hop, while the number of strangers almost reduces to 0 at KAIST.

5 Implementing Main Features of Stranger into STRON

In this section, we discuss how to design STRON. It mainly includes two components : 1) the degree of strangeness of nodes; 2) adjusting the number of strangers.

5.1 Degree of Strangeness of Nodes

Motivated by above observations, we integrate the degree of strangeness between two nodes into STRON. Note that the degree of strangeness and similarity between two nodes are complementary. Mathematically, let $Sim(i,j)$ denote the similarity between node i and node j, $D_s(i,j)$ denote the degree of strangeness between them, we have $D_s(i,j) = 1 - Sim(i,j)$. Fig.3 portrays the behavior of $D_s(i,j)$ at different contact times when using (4), where two nodes are randomly chosen as partners.

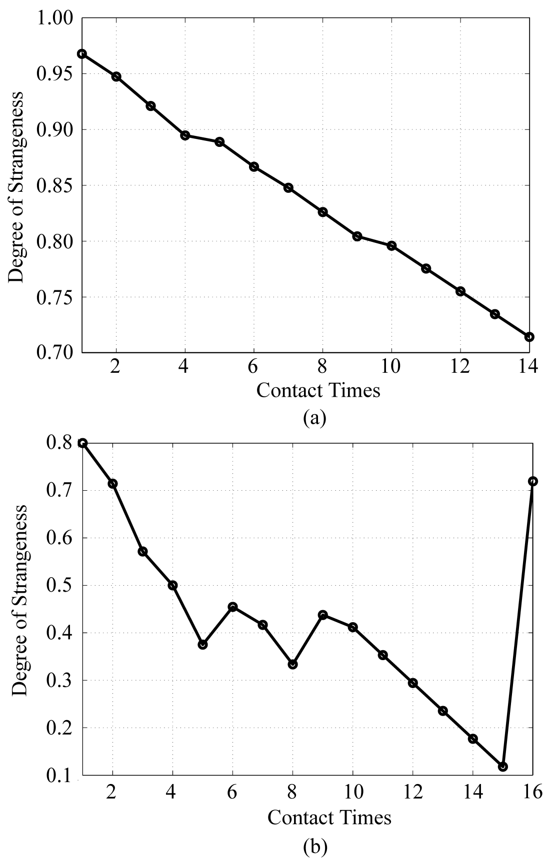


Fig.3. Degree of strangeness under different contact times. (a) KAIST (node-pair(1, 77)). (b) PMTR (node-pair(1, 11)).

In general, we use the min-max function to measure the similarity between two nodes. Take $x_i(j)$ as a sample, we hold

$$Sim(i,j) = \frac{x_i(j) - \min(X_i)}{\max(X_i) - \min(X_i)}, \quad \forall i \in N. \quad (4)$$

By taking both $Sim(i,j)$ and $D_s(i,j)$ into account, we give the following new forwarding metric.

$$U(i,j) = \alpha D_s(i,j) + (1-\alpha) Sim(i,j). \quad (5)$$

Based on the claim of Subsection 4.2, we empirically set $\alpha = H^{-C}$, where H is the current number of hops of the message and C is a system parameter ($0 \leqslant C \leqslant 1$, in Section 6, we show how the system parameter C impacts the performance of STRON through trace-driven emulations). By using this novel metric, STRON can conveniently help nodes to make smart forwarding decisions. For example, when node i meets node j, the message m is forwarded to the node j if $U(i, m_d) < U(j, m_d)$, and vice versa.

5.2 Adjusting the Number of Strangers

To further improve the networking performance, when strangers or friends meet familiar strangers or community partners of the destination, the former needs to forward messages to the latter. This is mainly because the relays become more and more familiar with the destination along the forwarding paths (please refer to the above section). In this situation, we have a chance to adjust the number of strangers. After a stranger forwards messages to a community partner of the destination, it deletes the messages from its buffer with a probability H^{-C}, based on the claim that the importance of strangers shows a decreasing trend when moving toward the destination. Though this mechanism is very simple, our trace-driven simulation results show that it significantly improves the routing performance. Algorithm 2 summarizes the above process.

5.3 Possible Issues

In this paper, though we take the min-max function as a solution to evaluate the similarity between nodes, STRON can conveniently incorporate other methodologies, such as the cosine angle separation, Euclidean distance, and Pearson correlation measures[33]. We think all of them should be esteemed, whereas, since we mainly focus on the roles of social relationships in the data forwarding process[34], we here do not discuss them.

Algorithm 2. STRON

Upon meeting up node j **do**
for any message m in i's buffer **do**
  if $m_d == j$ **then**
    deliverMsg(m), *remove*(m)
  else if $m \notin j$ **then**
    if $U_i(m_d) < U_j(m_d)$) or (a stranger or friend meets a familiar stranger or community partner of the destination) **then**
      forwardingMsg(m)
    end if
    if a stranger meets a community partner of the destination **then**
      remove(m) with a probability H^{-C}
    end if
   end if
  end if
end for

6 Performance Evaluation and Analysis

This section illustrates the performance gain of STRON by comparing with the state-of-the-art algorithms based on the two real datasets. We compare five forwarding schemes: 1) Epidemic. Epidemic achieves a lower bound on mean delivery delay and an upper bound on routing cost, hence, we use it as a baseline to compare other algorithms; 2) Greedy scheme based on the number of contacts (Greedy1); 3) Greedy scheme with the intra-contact time (Greedy2); 4) OP[26]: the optimal probabilistic method. The optimization means that each node knows the exact forwarding probability when meeting another node and the forwarding probability is acquired by an off-line method in advance; 5) STRON: our newly proposed algorithm.

In each scenario, each source sends one message to a randomly chosen destination and altogether 1 200 messages are generated. The communication range of nodes under KAIST is set to 250 m, a typical value of WiFi. The simulation results are the average over 20 runs for $C = 0.4$ in KAIST and $C = 0.8$ in PMTR. The evaluation metrics include 1) the relative delivery delay: the delivery delay achieved by each algorithm over that achieved by Epidemic; 2) routing cost: the ratio of the number of nodes infected by a message over the total number of nodes in the network.

Fig.4 demonstrates our results. The first observation indicates that STRON shows a competitive result on delivery delay as shown in Fig.4(a). Compared with the second best algorithm (OP), there does not exist noticeable difference between STRON and OP. On average, STRON only increases the delay by 4% at KAIST. Compared with the two greedy schemes, STRON reduces the delay almost by 50%. Note that it seems that different scenarios have different impacts on delivery delay. For example, at PMTR, the five algorithms almost achieve the same performance. We conjecture that this is mainly because PMTR is a very sparse scenario (please refer to Table 2, Section 4). Hence, the long inter-contact time between nodes dominates the delay.

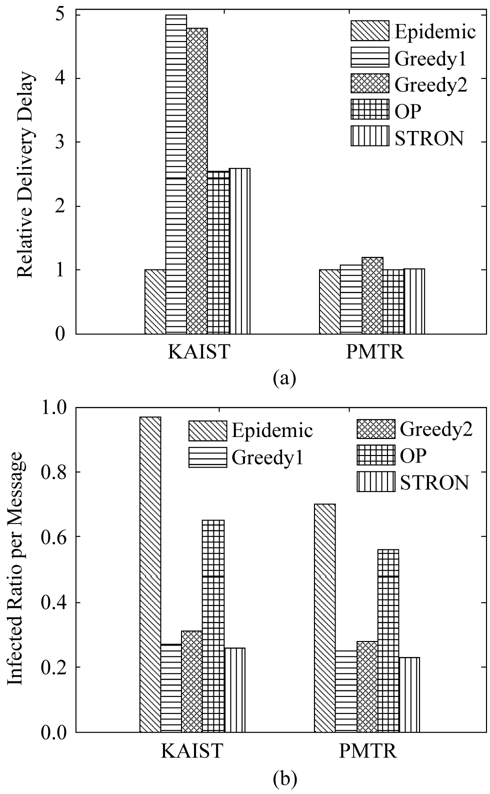


Fig.4. Relative delivery delay and cost under KAIST and PMTR. (a) Relative delivery delay. (b) Cost.

The second observation reveals that STRON achieves the best performance metric in routing cost as depicted in Fig.4(b). For instance, only 25% nodes are infected by STRON during the forwarding process in contrast with 65% of OP and the huge ratio 97% of Epidemic. Thus, STRON helps considerably in reducing up to 2.5× and 3.9× overhead in OP and Epidemic at KAIST. The similar situation also happens at PMTR. On the other hand, though the two greedy schemes have competitive performance in cost, they however slow down the dissemination speed as shown in Fig.4(a).

We now evaluate the influence of the system parameter C on the performance metrics of STRON. We increase the value of C to observe its impact on the delivery delay and routing cost. Fig.5 shows the results. As we can see from it, we can alleviate the routing cost by increasing the value of C, while we achieve these improvements at the cost of longer delivery delay, though there is a negligible increase at PMTR scenario. It seems to be better if we can define a more fine-grained selection strategy for C. Whereas, since different human contact traces show different mobility patterns, a general definition for C is impossible.

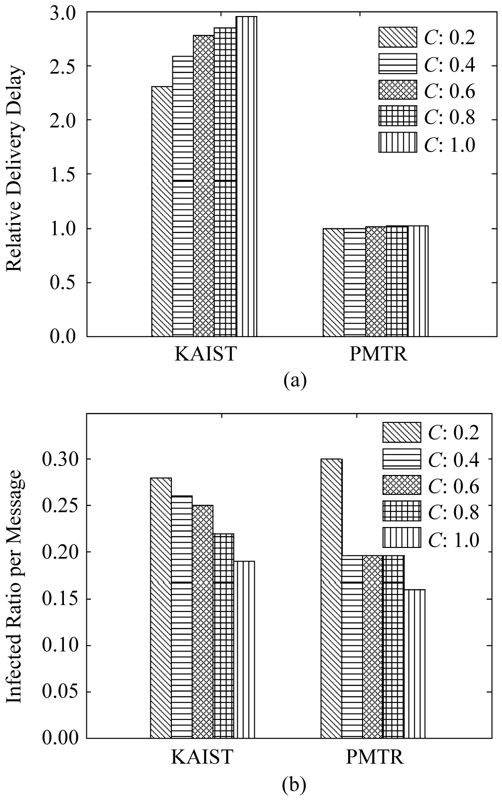


Fig.5. Performance metrics of STRON with different values of system parameter C. (a) Relative delivery delay. (b) Cost.

In summary, STRON achieves a striking improvement in routing cost while keeping mean delivery delay sufficiently close to those by OP under different scenarios. These experimental results clearly verify the effectiveness of STRON.

7 Conclusions

In this paper, we studied DTNs routing within a more challenging scenario. We explored the influence of strangers on routing performance. We observed the roles of social relationships in Epidemic and greedy scheme through two real datasets. We noticed that strangers have two sides in data forwarding process, and that the importance of strangers shows a decreasing trend along the forwarding path. Based on these observations, we developed STRON, a simple but efficient forwarding scheme, to improve the routing performance. The distributed feature makes it very suitable for opportunistic scenarios and the low communication and computation features make it easy to be integrated with state-of-the-art work. The trace-driven simulations obviously confirm its effectiveness, especially in terms of packet delivery delay and cost.

The significant topics for future work include the methods to compute the similarity between nodes and the taxonomy to identify the social relationships.

Acknowledgment We wish to thank the anonymous reviewers for their valuable comments and the CRAWDAD archive project for making the DTNs traces available to research community.

References

[1] Fall K, Farrell S. DTN: An architectural retrospective. *IEEE Journal on Selected Areas in Communications*, 2008, 26(5): 828-836.

[2] Conti M, Kumar M. Opportunities in opportunistic computing. *Computer*, 2010, 43(1): 42-50.

[3] Vahdat A, Becker D. Epidemic routing for partially connected ad hoc networks. Technical Report CS-200006, Duke University, 2000.

[4] Burgess J, Gallagher B, Jensen D, Levine B N. MaxProp: Routing for vehicle-based disruption-tolerant networks. In *Proc. the 25th INFOCOM*, April 2006, pp.1-11.

[5] Lindgren A, Doria A, Schelen O. Probabilistic routing in intermittently connected networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2003, 7(3): 19-20.

[6] Leguay J, Friedman T, Conan V. Evaluating mobility pattern space routing for DTNs. In *Proc. the 25th INFOCOM*, April 2006, pp.1-10.

[7] Yuan P, Ma H. Hug: Human gathering point based routing for opportunistic networks. In *Proc. WCNC*, March 2012, pp.3024-3029.

[8] Nguyen N, Dinh T, Tokala S. Thai M. Overlapping communities in dynamic networks: Their detection and mobile applications. In *Proc. the 17th MOBICOM*, Sept. 2011, pp.85-96.

[9] Hui P, Crowcroft J, Yoneki E. BUBBLE rap: Social-based forwarding in delay tolerant networks. In *Proc. the 9th MobiHoc*, May 2008, pp.241-250.

[10] Mtibaa A, May M, Diot C, Ammar M. PeopleRank: Social opportunistic forwarding. In *Proc. the 29th INFOCOM*, March 2010, pp.111-115.

[11] Daly E, Haahr M. Social network analysis for routing in disconnected delay-tolerant MANETs. In *Proc. the 8th MobiHoc*, Sept. 2007, pp.32-40.

[12] Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 2005, 435(7043): 814-818.

[13] Shah R, Roy S, Jain S *et al.* Data MULEs: Modeling and analysis of a three-tier architecture for sparse sensor networks. *Ad Hoc Networks*, 2003, 1(2/3): 215-233.

[14] Li Q, Rus D. Sending messages to mobile users in disconnected ad-hoc wireless networks. In *Proc. the 6th MobiCom*, Aug. 2000, pp.44-55.

[15] Zhao W, Ammar M, Zegura E. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *Proc. the 5th MobiHoc*, Sept. 2004, pp.187-198.

[16] Zhao W, Ammar M, Zegura E. Controlling the mobility of multiple data transport ferries in a delay-tolerant network.

In *Proc. the 24th INFOCOM*, March 2005, pp.1407-1418.

[17] He T, Lee K, Swami A. Flying in the dark: Controlling autonomous data ferries with partial observations. In *Proc. the 11th MobiHoc*, Sept. 2010, pp.141-150.

[18] Spyropoulos T, Psounis K, Raghavendra C. Efficient routing in intermittently connected mobile networks: The multiple-copy case. *Transactions on Networking*, 2008, 16(1): 77-90.

[19] Qin R, Li Z, Wang Y *et al.* An integrated network of roadside sensors and vehicles for driving safety: Concept, design and experiments. In *Proc. the 8th PerCom*, March 29-April 2, 2010, pp.79-87.

[20] Ivancic W, Wood L, Hdiday P *et al.* Experience with delay-tolerant networking from orbit. In *Proc. the 4th ASMS*, Aug. 2008, pp.173-178.

[21] Merugu S, Ammar M, Zegura E. Routing in space and time in networks with predictable mobility. Technical Report GIT-CC-04-7, Georgia Institute of Technology, 2004.

[22] Jain S, Fall K, Patra S. Routing in a delay tolerant network. In *Proc. the ACM SIGCOMM*, Aug. 30-Sept. 3, 2004, pp.145-158.

[23] Liu C, Wu J. Routing in a cyclic mobispace. In *Proc. the 9th MobiHoc*, May 2008, pp.351-360.

[24] Musolesi M, Hailes S, Mascolo C. Adaptive routing for intermittently connected mobile ad hoc networks. In *Proc. the 6th WoWMoM*, June 2005, pp.183-189.

[25] Yuan Q, Cardei I, Wu J. Predict and relay: An efficient routing in disruption-tolerant networks. In *Proc. the 10th MobiHoc*, May 2009, pp.95-104.

[26] Gaito S, Pagani E, Rossi G. Strangers help friends to communicate in opportunistic networks. *Computer Networks*, 2011, 55(2): 374-385.

[27] Yoneki E, Hui P, Crowcroft J. Visualizing community detection in opportunistic networks. In *Proc. the 2nd CHANTS*, Sept. 2007, pp.93-96.

[28] Hui P. People are the network: Experimental design and evaluation of social-based forwarding algorithms. Technical Report UCAM-CL-TR-713, University of Cambridge, March 2008.

[29] Rhee I, Shin M, Hong S *et al.* On the levy-walk nature of human mobility. In *Proc. the 27th INFOCOM*, April 2008, pp.924-932.

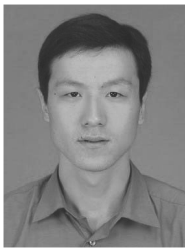
[30] Lee K, Hong S, Kim S *et al.* SLAW: A mobility model for human walks. In *Proc. the 28th INFOCOM*, April 2009, pp.855-863.

[31] Zyba G, Voelker G, Ioannidis S *et al.* Dissemination in opportunistic mobile ad-hoc networks: The power of the crowd. In *Proc. the 30th INFOCOM*, April 2011, pp.1179-1187.

[32] Pietiläinen A, Diot C. Dissemination in opportunistic social networks: The role of temporal communities. In *Proc. the 13th MobiHoc*, June 2012, pp.165-174.

[33] Sarwar B, Karypis G, Konstan J *et al.* Application of dimensionality reduction in recommender system — A case study. In *Proc. WEBKDD*, Aug. 2000, pp.1-12.

[34] Zhang Y, Yu T. Mining trust relationships from online social networks. *Journal of Computer Science and Technology*, 2012, 27(3): 492-505.



Pei-Yan Yuan received the B.S. and M.S. degrees in computer science from Henan Normal University, in 2001 and Wuhan University of Technology, China, in 2007, respectively. He is currently a Ph.D. student in the School of Computer Science, Beijing University of Posts and Telecommunications. His research interests include networking and protocol engineering, mobile ad hoc networks, sensor networks, delay-tolerant networks, social networks, quality of service, performance evaluation, etc. He is a member of ACM.

Hua-Dong Ma received the B.S. degree in mathematics from Henan Normal University in 1984, the M.S. degree in computer science from Shenyang Institute of Computing Technology, Chinese Academy of Science in 1990, and the Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Science, Beijing, in 1995. He is currently a professor and director of Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, executive dean of School of Computer Science, University of Posts and Telecommunications, China. He visited UNU/IIST as a research fellow in 1998 and 1999, respectively. From 1999 to 2000, he held a visiting position in the Department of Electrical Engineering and Computer Science, the University of Michigan, USA. He was a visiting professor at University of Texas at Arlington from July to September 2004, and a visiting professor at Hong Kong University of Science and Technology from Dec. 2006 to Feb. 2007. His current research focuses on multimedia system and networking, Internet of things, and sensor networks. He has published more than 100 papers and four books on these fields. He is member of the CCF, ACM, and IEEE.

Peng-Rui Duan received the B.S. and M.S. degrees in computer science from Beijing University of Posts and Telecommunications, China in 1996 and 2002, respectively. He is currently a Ph.D. student in the School of Computer Science, Beijing University of Posts and Telecommunications, China. His research interests include networking and protocol engineering, mobile ad hoc networks, multimedia sensor networks, delay-tolerant networks, social networks, etc.