A Social Behavior Based Interest-Message Dissemination Approach in Delay Tolerant Networks

Tzu-Chieh Tsai(✉), Ho-Hsiang Chan, Chien Chun Han, and Po-Chi Chen

Computer Science Department, National Chengchi University, Taipei, Taiwan
ttsai@cs.nccu.edu.tw,
{100753503,103753009}@nccu.edu.tw, hanjord@gmail.com

Abstract. Compared with 3G, 4G and Wi-Fi, Delay-Tolerant Networking (DTN) can only have intermittent chance to transmit messages. Without a clear end-to-end path, routing a message in DTN to the destination is difficult. But in some particular case, it could be an advantage. People around the world have their personal habit and it will be projected on their social life. Therefore we use the social behavior as a foundation feature of our routing algorithm. We propose two new kinds of routing algorithms with our own trace file. On one hand, birds of a feather flock together, so people who have similar interests tend to go to the same places. In case of this, we combining the personal interests and the trace file to different buildings where each node locates, we propose the building-based routing algorithm. On the other hand, we think people who have similar interests hang out together more often, so we use the social relationship as a feature and propose social-based routing algorithm. In the end, we compare our algorithms with Epidemic, MaxProp and PRoPHET routing algorithms. The result shows that our algorithms outperform the others.

Keywords: Delay-Tolerant network · DTN · Campus environment · Personal information · Personal interest · Social relationship

1 Introduction

1.1 Background and Motivation

In these years, smart phone has become more and more important in people's lives. Because smart phone is a very powerful device, it can be useful in many situations like sending and receiving e-mails, communicating with friends, acting as a digital calendar to remind us special days, checking daily weather, and virtual wallet. We can see everyone carry a smart phone wherever they go. Smart phone has become part of our lives.

The issue we focus on is plenty of advertising spam we might receive every time we open our email box. Maybe there are some messages we are interested in. However, we don't have time to check the spam one by one to pick what we want to read, so we ignore them in most cases. If messages can be transferred to people who are interested in them, it surely can reduce the overhead and make better performance. Transferring

R. Doss et al. (Eds.): FNSS 2016, CCIS 670, pp. 62–80, 2016.
DOI: 10.1007/978-3-319-48021-3_5

the messages through the Internet is not always the best way. First, transferring messages through the Internet is limited to its ability to access the Internet. Second, besides subscribing to specific channel s and receiving arbitrary spam, we can't get the messages we want. We want to push the messages to wherever the users are despite the ability to access the Internet. To overcome this problem, we think DTN (Delay-Tolerant Networking) is a good choice.

1.2 Delay-Tolerant Network (DTN)

Delay-Tolerant Networking is a dynamic wireless network. Every node may move freely and be organized depending on their social relationship. DTN can provide interoperable communication in challenging environments, which is defined as the network is not always in connect or the network has no end-to-end path. It is an approach of computer network architecture that can use the strategy of store, carry, and forward to transmit messages in the disconnected network environment. In Fig. 1, when the node is in the environment without network connection, it may convey messages to nearby nodes by using the short distance transmit technique such as Bluetooth or Wi-Fi direct. When the relay node receives the message, it can carry the message until meeting the next proper node to help transmit and forward the message. Via this approach, messages can travel around the environment and be transmitted to the destination. The connection between two nodes in the DTN environments can only keep for just few seconds, so it needs to find appropriate node to help transmit messages in limited time.

Fig. 1. Store, carry, and forward in DTN

There are two main reasons why we propose a routing algorithm based on DTN. First, we think that people have a routine trace everyday, just like most people have to go to work, and students have to go to school. We always get up around the same time in the morning, do the same chores, and most important of all, we commute to our destination in almost the same route and in regular time. It reveals that we might meet the same stranger every day, but we don't even notice. This stranger would be a terrific node in our routing algorithm. Because we can keep meeting this person everyday, we can update the messages with people we meet. Thus, we can know which person is closer to the destination that messages should be transferred to. Second, the Internet only provides us an end-to-end way to transfer the messages. Under this condition, if we want to transmit a message, we have to know where the destination is first. But in some cases, especially in advertising messages, we do not know where all the destinations are when the message is created. If we use the Internet as communication model, there are only three scenarios: (1) Enterprise, which creates the advertising messages, can only send these messages to people who have registered before. (2) People can only transfer messages to their friends. (3) The enterprise can spread the messages randomly, which will cost a lot. But in DTN, we can transfer the message through the node with store, carry and forward strategy. In the previous work [1], which is published in IEEE magazine, we can see that the research of using the interest as a feature in their routing protocol has an excellent performance. So we want to take it a step further and continue this research. In this paper, we suppose people who have similar interests tend to go to the same places. For example, people who like sports or exercise will go to the gym or sports field. Moreover, when they are shopping or doing something else, they are more likely to do things related to sports. People who are interested in art will go to see art exhibitions, and people who like reading will go to the library or bookstores. In this paper, we use this as a feature on our routing algorithm.

In tradition, there are several ways to route the message in DTN environment such as Epidemic, MaxProp [2], and PRoPHET [3]. The routing of Epidemic is by the way of transferring messages to every node where the carriers meet. The overhead of Epidemic is extremely high, but it has a better performance. We want to reduce the overhead as much as we can and prevent the performance from dropping too much. The routing of MaxProp concerns with people they have met before and the sequence of messages to be sent. MaxProp does not keep an eye on which direction the message should be sent to. In our routing algorithm, it will calculate the probability of which relay node gets a higher chance to arrive at the destination. PRoPHET focuses on two nodes meeting each other and which one gets a path that can transfer messages to the destination in a relatively higher probability. However, PRoPHET only works well at unicast case, and it saves all the probability of meeting all the other nodes, which would cost a lot of memory in a giant trace file. In our routing algorithm, we only saves data of the node that has met other nodes before, and it surely helps us reduce the usage of memory.

1.2.1 Cosine Similarity

In our routing algorithm, we have to decide whether people interest the advertisement message or not. For simplicity without losing of generality, we first choose cosine similarity as our indication for social interest relation. It is simple and quite easy to perform. We bring out the cosine similarity to help us determine who wants to know the message. The formula is showing below at Formula (1)

$$Cosine\ Similarity = \frac{\vec{A}, \vec{A_X}\ (D)}{\left\|\vec{A}\right\| \cdot \left\|\vec{A_X}\ (D)\right\|} = \frac{\sum_{i=1}^{n} A_i \times A_X(D)_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2 \times \sum_{i=1}^{n} (A_X(D)_i)^2}}, i < n \quad (1)$$

In our trace file, we collect five different interests. But we only use two of the interests as the input. Because the rest three interest columns do not show the difference unfortunately. So, we only take two different interests as input column to calculate cosine similarity. The interest column will be assigned when a message is created. And all of the nodes in our simulator have their own interests. We can use both the interest column of the message and the interest column of the node to determine whether the node is interested or not. And the detail will be described in Sect. 4.

We propose a new DTN routing algorithm, based on the assumption that people who have daily routine and who have similar interests flock together. Then use the cosine similarity to see who is interested in the message. Finally, we compare our algorithm with the classic routing algorithm, and the result shows that we have a better performance.

2 Related Work

In the paper [4], the author thinks that social-based routing and location-based routing are slightly different. Due to the difficulty of collecting the real trace data, we can use the social data as a substitution. If we can collect both social-based and location-based data, we can compare them with each other.

In such a variety of DTN routing researches, we focus on two types of research (1) Collecting human real movement data. (2) How to use social data to send data to the destination quickly with less resource in DTN environment. Three trace data we will discuss below do not consider the condition of what the real society is like. There are different kinds of people with different interests in the real society, and this is what makes a diverse society. People may go to different places or do different things depending on their jobs and interests. So if we can use a more realistic trace file that we can regard it as a tiny real society. Different kinds of nodes are moving freely in the emulator, which is closer to the real society.

2.1 Social Trace Data

2.1.1 Reality Mining: MIT [5]

This experiment was carried out by MIT. The researcher gives 100 NOKIA's smart phone to 100 students, and the experiment duration is 9 months. Students who participated in the experiment were asked to use smart phones to communicate with other students by Bluetooth, and their trace, contact time, and communicate time were recorded. Via this experiment, we can analyze and predict social activities' relation with the subjects to know its next movement and social relations. The disadvantage of the experiment was that 75 students were from MIT Media Laboratory, and the other 25 students were from MIT Sloan business school. We think that the composition of participants can't be a miniature of the real society (Fig. 2).

2.1.2 Cambridge [6]

This experiment was carried out by Cambridge computer lab. The researcher used the equipment named iMote to collect the real trace data. In Cambridge05, the experiment separated students into freshman and sophomore, and it also included graduate and doctoral students (Fig. 3).

During the 11 days, 54 students used iMote equipment with Bluetooth technique, which helps to measure and record the main active area, the other students they contact,

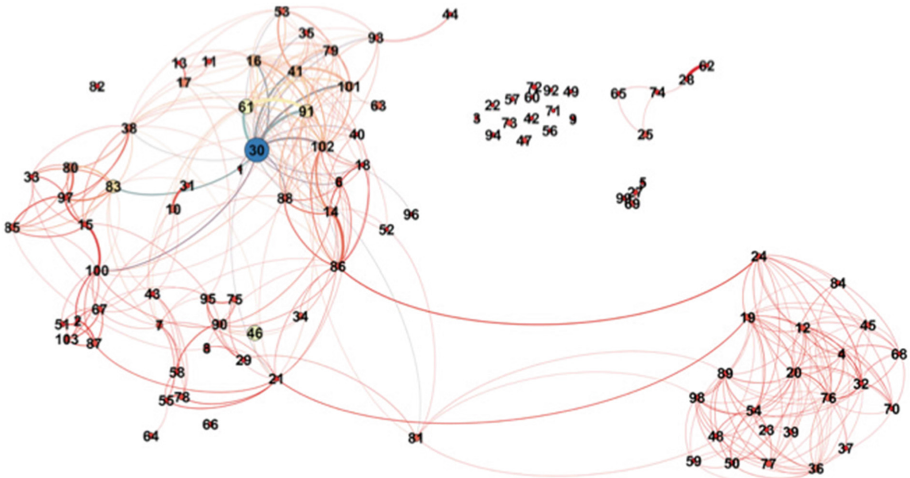
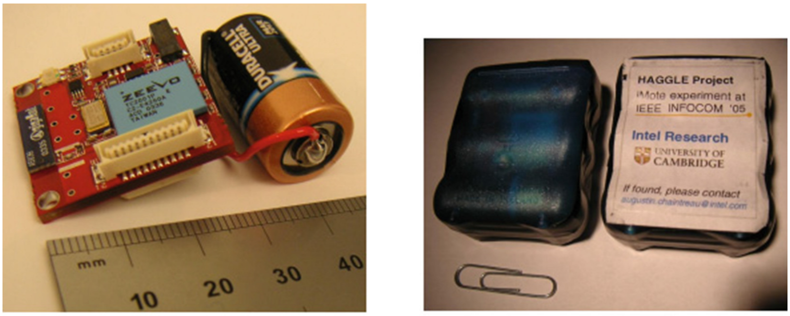


Fig. 2. One view of the network created by MIT reality mining dataset



(a) iMote with battery

(b) iMote package

Fig. 3. iMote for the experiment

and the length of communicate time of each student. In the future, this data can be used in social relationship experiment.

2.1.3 Infocom05, 06 [7]

This experiment was held in students and professors who attended Infocom conference in 2005 and 2006. In the beginning, every participant was given an equipment named iMote. Because there were lots of different topics in the conference, every participant would go to listen to the topic they were interested in. Thus, we can know every participant's interests and who they communicate with.

In 4 days, 98 people participated in this experiment. Through the experiment, we can know each participant's professional specialty and whether they communicated with other people who have the same research domain., we can use participants' communicate time to conjecture their social relationship.

After reviewing previous research, we think we have to select the participants in order to make the trace data more similar to the real world. One of the most important things is to pick who can enroll in our experiment. All the details are described in Sect. 3.

In the trace file above, the participants are comprised of one or two particular group. Which will lead to the trace file is not general. And the trace file will be limited. What we want is a trace file that is a miniature of real society. So we will keep an eye on this while we are picking the participants to involve our experiment. And the detail will be described in Sect. 3.

2.2 Social-Based in Delay-Tolerant Network

2.2.1 Social-Aware Data Diffusion in Delay Tolerant MANETs [8]

This research proposes a routing algorithm based on different interests of each node. If two nodes have similar interests, which means the similarity of interests has exceeded the threshold, then we define them as friends; otherwise, the two nodes are defined as strangers. Therefore, when two nodes meet, they will exchange interest list and data list. When their interests are similar, they are friends to each other, and they will exchange data that the carrier likes. On the other hand, if they are strangers to each other, they will diffuse data that they are not interested in the message, just like the state shown below (Fig. 4).

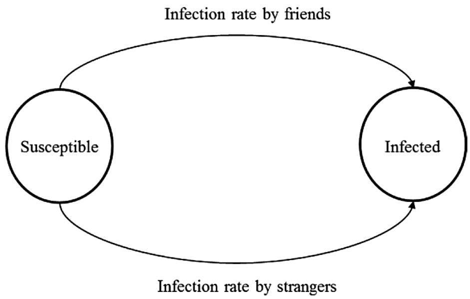


Fig. 4. Node infected by friend or stranger

Even if the two nodes are strangers, it doesn't mean that their friends or people they might meet are not interested in the message. So in our routing algorithm, we give a chance to case like that. We log every node that one node has met before, and we give it a try if the log file reveals they have a high chance to meet each other again and the data also interests them. For example, node A and node B go to work at the same time in the morning, and they take the same transportation. However, they are strangers, so they may not talk to each other. It is a good time to exchange the advertising messages they got while they are on the bus. During the time they are on the bus, they can know the interests of each other, and the interests of people who they usually meet. It helps us know whether other people are good relay nodes or not.

2.2.2 Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETS [9]

It is difficult to diffuse messages in spare MANETs. All the nodes can move freely. So to find out the most efficient path is the key to this research. Previous research has conducted many theories to discover the best way to route the messages. To overcome this issue, there is Centrality, which can reveal whether the node is connected to the neighbor nodes. In other words, the node is able to know whether there is a path to the destination. They propose a new routing protocol named SimBet. SimBet is based on the betweenness Centrality and Similarity of nodes, and it chooses the intermediate node to help carry the messages. But the disadvantage is that all the messages tend to gather at some active nodes. Nodes, which are in relatively static state, can only transfer the messages and are not able to get the messages they want. In our routing algorithm, we do not want a few people to carry most of the messages. We think it might lead to information starving for some nodes. We consider not only the connection between nodes but also the landmark where the node has gone before. A node can be the relay node if this node will meet some other nodes which are interested in the message. Everyone can be the relay node depending on their movement and social relationship, and we think it is a better way to route the messages.

3 NCCU Trace Data

All the participants in trace data we mentioned above are limited to some specific group, either students of particular college or participants in particular conference. For example, the participants in MIT trace data were composed of Media Laboratory and business school students. In the Cambridge trace data, only computer laboratory students were enrolled. Furthermore, the Infocom trace data was collected during the conference, so most of the participants were related to the conference. We think these three trace data above can't reveal how the real social network works. In real social network, people are not supposed to do the same thing or the same work all the time, and they have different interests. We assume that where people usually go depends on their jobs and interests, and this issue is what we concentrate on. Because we don't have enough participants that can represent the real society to implement the experiment, we limit the environment to our campus. We recruit participators according to the ratio of different college. Also, the participators of the same college are from different departments for diversity. This is close to reality instead of participators with similar background like previous ones. Then, referring to [10], we not only record the trace file, but also record the self-declared interests of participators. In the end, because we don't have enough participants that can represent the real society to implement the experiment, we limit the environment to our campus. Participators are students in National ChengChi University.

3.1 Form (Selecting Participants)

When we were building our own trace file, we selected participants first. In order to find the suitable participants, we asked everyone who enrolled in to fill out the form, which asked some basic profile information and the most important thing, participants' college and interests. Figure 5 shows a part of the real data that we obtained.

ID	College	Sports	Reading	Social	Arts	Service
A	4	0.75	0.75	1	1	0.75
B	0	1	0.5	0.5	0	0.5
C	2	0.5	1	0.75	0.75	0.75

Fig. 5. Form list

3.1.1 College

All the participants should not come from one specific group, and they should not be all strangers to each other, either. Two participants were assigned to a group, and they must know each other or we would not accept them. We also cared about what college they are from. The college quantity ratio depended on the real college quantity ratio of total students in NCCU. If the quantity of one college was about to surpass the quantity we wanted, we would not accept the coming group, either. In this case, participants in our trace data were distributed to different colleges. Some people knew each other, and others didn't just like the real society.

3.1.2 Interest

In our algorithm, we had to determine where the destination of each message was. So we asked the participants what kind of interest they are. The interests were divided into five types, which were sports, reading, social, art, and service. The score was limited from 0 to 1, and each scale was at least 0.25. On the other side, when a message was created, it would be assigned to these five interest types accordingly. But when we calculating the average of these interest columns, we find out that the social, arts and service column don't show the difference. So we can only use sports and reading column as an input to calculate the cosine similarity as we mentioned earlier at Formula (1) to define whether the nodes are interested in the messages or not. If yes, the node will be one of the destinations of the message.

3.2 Trace Data

The most important part of trace data is the movement of each node. In order to get the real trace data, using smart phones was the best way to collect the information we needed. In traditional work, they created a device to record the user location, but the users may forget to bring the device with them. In reality, carrying another device only to send or receive messages is difficult to implement. In these days, people carry their own smart phone wherever they go, even at the places they are not allowed to use it. In conclusion, we think the smart phone is the best device to implement the DTN routing algorithm. For this reason, we designed an Android app and installed it in each participants' smart phone. We ran a background service to record GPS position every 10 min. If we scan too often, the battery of the smart phone will dry out fast. In our previous experiment, if we scanned every 5 min, the battery couldn't hold on for a day. To continue record the user position, we stipulated a scan every 10 min.

After all the trace data were collected, we ignored the trace data that were not on the campus and normalized the trace data. In addition, we want the trace data to move smoothly but not to disappear in one place and appear in another place suddenly. We had to normalize the trace data and keep the trace track continuous. Finally, when these works were done, our trace data were about to be used. There were 115 available data in our trace data in total, and the experiment lasted for two weeks, from 17th Dec to 31st Dec in 2014. All the trace data can be downloaded as soon as the paper is published [11].

4 Routing Approach

Most traditional routing scenarios only have one single destination, but our goal is to deliver messages to various destinations and reduce the overhead. However, using DTN to transmit a personal message to another person would not be as suitable as the advertising message. Thus, if a company wants to spread an advertising message to as many customers as possible, DTN is a suitable option.

4.1 Environment Definition

First, because the trace data we obtained were from the campus, we limited the simulation environment to the campus, too. When an advertising message is created, we don't quite know which destination we should send it to. All we can do is to find out the people who might be interested, and push the message to the right person. But in reality, we won't know whether a node is interested in the message or not until it bumps into other nodes and exchange metadata. The metadata includes the information about the contact node and other's interests, what nodes the contact node has met, and also what messages the contact node has got so far. When the receiver receives the metadata, it will calculate the cosine similarity between the interest type of the message and the interests of other nodes. If the result surpasses the threshold, then the message will be delivered. The entire process is termed "contact."

4.2 Routing Strategy

The message forwarding flow is shown in Fig. 6. Whenever a message is created, it will be assigned to one node randomly. For instance, we assume the message is assigned to node A. Node A may meet node B along the way to its destination. The time node A and node B spend on exchanging data with each other is called the total "contact time". When the connection is interrupted or out of the connection area, the data transferring stops. During the contact time, node A will try all kinds of messages

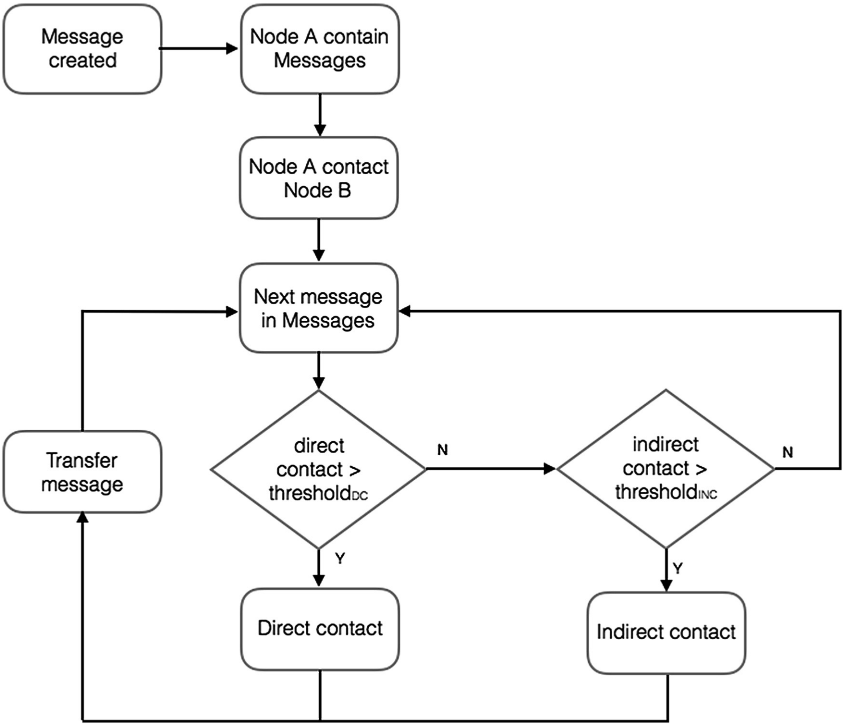


Fig. 6. Message forwarding flow

in the transferring waiting queue one by one. In next paragraph, there are two kinds of contact in our algorithm: direct contact and indirect contact.

4.2.1 Direct Contact

Direct contact is the most common way to transfer the messages. When two nodes meet each other, they will exchange metadata about their interests. And the node will calculate and compare the interest type of the message with the interests of another node it meets. For instance, node A encounters node B, and they exchange the metadata. Node A gets the interests of node B, and it will calculate the cosine similarity (Cos) between the interest of node B and the interest type of message (I_m). Node A will compare every message that node A has and node B doesn't have with the interest of node B (I_{N_B}). Formula (2) is shown below.

$$Cos(I_{N_B}, I_m) > Thres\_DC \tag{2}$$

If the result of cosine similarity is greater than the *Thres_DC*, the message will be put in queue and ready to be transferred. The transferring time will last as long as the contact time of two nodes or until all the messages in queue are delivered.

4.2.2 Indirect Contact

In addition to direct contact, we expect the contact node can be a relay node, which can carry the messages to other nodes that are interested in the message. If we transfer these messages randomly, it will lead to high overhead and doesn't make our performance better. So we propose two kinds of indirect contact routing algorithms, which are base on building and social relationship.

Building Based Indirect Routing. We use the historical data of which building the students went to forecast the future. First, students of different colleges go to different buildings. For example, students of Commerce College have a higher chance to meet one another than any other student of other colleges because they usually go to the classrooms in the Commerce Building. Besides, students of Accounting Department have a much higher chance to meet students of the same department than students of Statistics Department in Commerce College. Second, we assume that students who have similar interests gather in the same building.

In Fig. 7, the students of Language College go to the library more often compared with the students of Science College and Law College. The reason might be that some ancient documents only have print editions, so most of them have to go to the library to get these books. However, the students of Science College can search the information on the Internet.

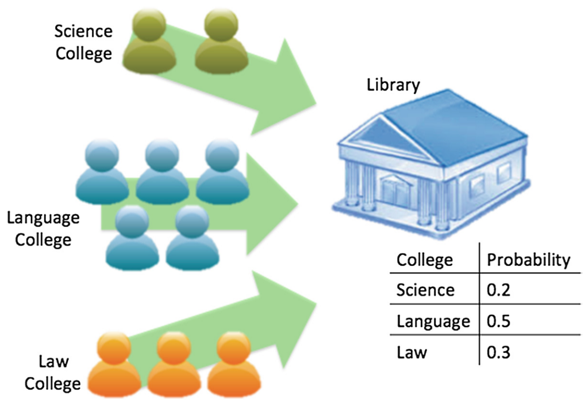


Fig. 7. The counting of different college of students go to library

In order to prove our assumption, we calculate all the probability that students from different colleges go to each building on the campus. We can calculate the interests, which can be defined just like the form we ask the participants to fill out. Suppose that the interest of the building can be counted and defined according to what kinds of people have come before. Figure 8 displays the counting process. We first initialize all the interest of each building to 0. Second, whenever someone goes to the building, we add the interest of that person to the interest of the building. After the entire trace file is checked, we normalize the interest of the building. Finally, we can define the interest of the building.

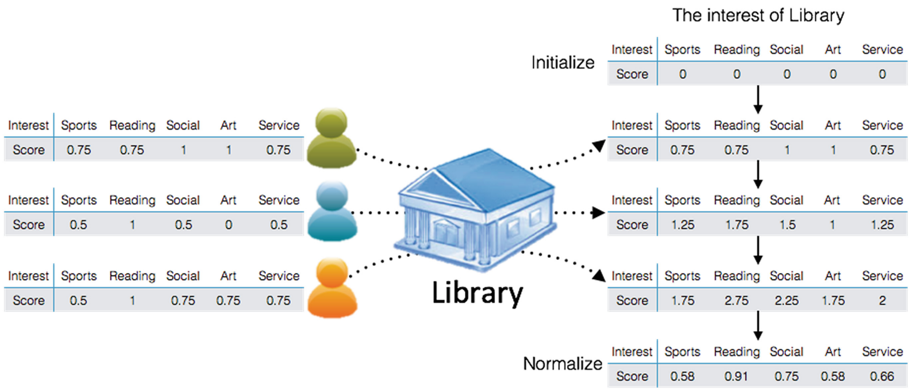


Fig. 8. The counting of the interest of library

When we finish collecting the entire trace file, we calculate how many times students of specific college go to specific building. We think there is logical evidence that students of the same college or the same department have a higher chance to go to the same building.

In our trace file there are 8 different colleges in total 115 nodes ($\mathcal{S}$). And there are we define 17 buildings ($\mathcal{B}$) to calculate the probability of college to each building ($Prob_{College(N_i),b_k}$) and try to verify our proposal. The calculating algorithm will be described below.

Algorithm1: Calculating probability of which building do students in college go
$Input: every\ node\ i(N_i) \in \mathcal{S}, trace\ file(TF), every\ building\ b_k \in \mathcal{B}$
$Output: Prob_{College(N_i),b_k}$
$Set\ every\ Prob_{College(N_i),b_k}\ to\ 0$
$For\ every\ N_i \in \mathcal{S}$
  $For\ every\ time\ j(T_j)\ in\ TF$
    $if\ location\ (N_i, T_j)\ in\ b_k$
      $Prob_{College(N_i),b_k}(College(N_i), b_k) += 1$
  $end\ for$
$end\ for$
$Normalize\ Prob_{College(N_i),b_k}\ to\ between\ 0\ to\ 1, for\ every\ b_k$

Algorithm2: Calculating algorithm of the interests of building
$Input: every\ node\ i(N_i) \in \mathcal{S}, Interest\ of\ N_i(I_{N_i}), trace\ file(TF), Interest\ of\ building\ b_k(I_{b_k})$
$Output: I_{b_k}$
$Set\ every\ I_{b_k}\ to\ 0, b_k \in \mathcal{B}$
$For\ every\ N_i \in \mathcal{S}$
  $For\ every\ time\ j(T_j)\ in\ TF$
    $if\ location\ (N_i, T_j)\ in\ b_k$
      $I_{b_k} = I_{b_k} + I$
  $end\ for$
$end\ for$
$Normalize\ I_{b_k}\ to\ between\ 0\ to\ 1, for\ every\ b_k$

After we do both Algorithms 1 and 2, we can use them to implement our Building-Based Indirect Routing. For instance, when node A encounters node B, node A carries one message M, but node B doesn't want message M. Then A will check the interest of every building and the probability of node B going to each building. If the calculating result surpasses the threshold ($Thres\_meet$), node A will still transfer the message M to node B. The formula is shown in Formula (3).

$$\sum\nolimits_{b_k \in \mathcal{B}} Cos(I_M, I_{b_k}) * Prob_{College(N_B),b_k} > Thres_{meet} \tag{3}$$

Social Based Indirect Routing. This is another indirect routing algorithm we propose. We think every one has a routine schedule in a period of time. In most cases, the period is a week. Because most people have to go to work or school on weekdays, it leads to the result that we will do mostly the same task like what we did seven days ago. In our campus scenario, students have to follow their own schedule and go to class accordingly. So if we can record where they were last week, we may forecast where they will go in the near future. To achieve our goal, we define a new interest (*IL*) that records what kind of people the node met seven days ago. The new interest of node A last week is defined as IL_{N_A,D_1}. The calculating algorithm of the new interest is described below.

Algorithm3: Calculating the new interest of node (*IL*)
$Input: every\ node\ i(N_i) \in \mathcal{S}, Interest\ of\ N_i(I_{N_i}), trace\ file(TF)$
$Output: IL$
$Set\ every\ IL_{N_i,D_d}\ to\ 0, N_i \in S_N, D_d\ between\ Monday\ to\ Sunday, 1 \le d \le 14$
$For\ every\ N_i \in \mathcal{S}$
$\quad For\ every\ time\ j(T_j)\ in\ TF$
$\quad\quad if\ N_i\ meet\ N_x\ on\ D_d, N_x \in \mathcal{S}$
$\quad\quad\quad IL_{N_i,D_d} = IL_{N_i,D_d} + I_{N_x}$
$\quad\quad\quad IL_{N_x,D_d} = IL_{N_x,D_d} + I_{N_i}$
$\quad end\ for$
$end\ for$
$For\ every\ N_i \in \mathcal{S}$
$Normalize\ IL_{N_i,D_d}\ to\ between\ 0\ to\ 1, for\ every\ 1 \le d \le 14$

For example, node B encountered node C, D and E last Monday. So the new interest of node B (IL_{N_B,D_1}) will be the mean of the interests of node C, D and E, just like what we do when counting the interest of the building (Fig. 9).

When node A which has a message M meet node B next Monday, node B is not interested in the message M. Node A will check the IL_{N_B,D_1} value. If the cosine similarity between the message M and the new interest IL_{N_B,D_1} surpass the threshold (*Thres_Cos*). Node A will still transfer the message M to node B. Then node B will be

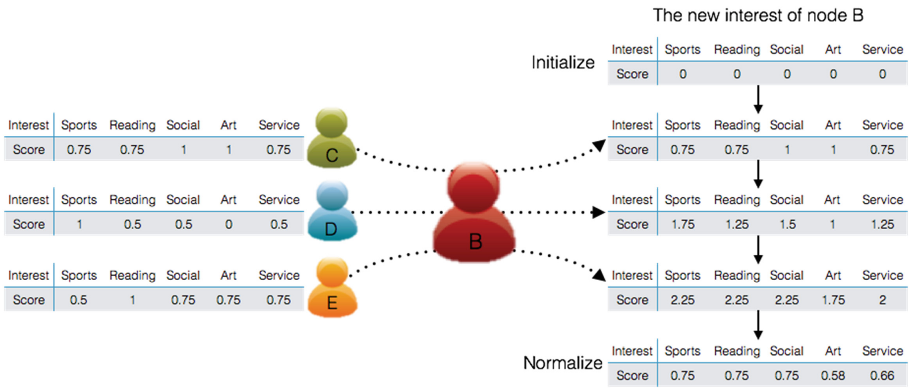


Fig. 9. The counting of the new interest of node B

a carrier of message M and help spread the message M. The formula is displayed below.

$$Cos\left(I_M, IL_{N_A,D_d}\right) > Thres_{Cos} \tag{4}$$

When we simulate our trace file, we only have two weeks of trace data. So, if we wt to simulate the first week, we have to use the IL_{N_i,D_d} of the second week. In other words, we have to use the $IL_{N_A,D_{d+7}}$ in our simulator if we want to simulate the first week. On the other hand, when we want to simulate the trace file of the second week, we use $IL_{N_A,D_{d-7}}$.

5 Simulation Settings

5.1 Simulation Environment

In our simulation, we use ONE (Opportunistic Network Environment simulator) [12] (shown as Fig. 10) and the map of NCCU (Nation Cheng-Chi University) surrounding area to validate our approach. All nodes in the simulation represent the student of this college, and they walk around according to their class schedule or for some purpose.

5.2 Simulation Setting

The simulation setting is shown as Table 1. The map area is 3764 m x 3420 m, which is the main active area of NCCU (Fig. 11), and the simulation time is from 12 a.m. to 12 a. m. of the next day. This is about 172800 s, it is equivalent to two days. The reason why we choose this time slot is that some students are at the school during the day and others live in the dorms at night. The node data transmission rate is 250KBps, and the transmission rage is 10 m. The message size of data is 500 KB $\sim$ 1 MB, the node buffer size is 100 MB, and the message's TTL is 1080 min, which is equivalent to 18 h.

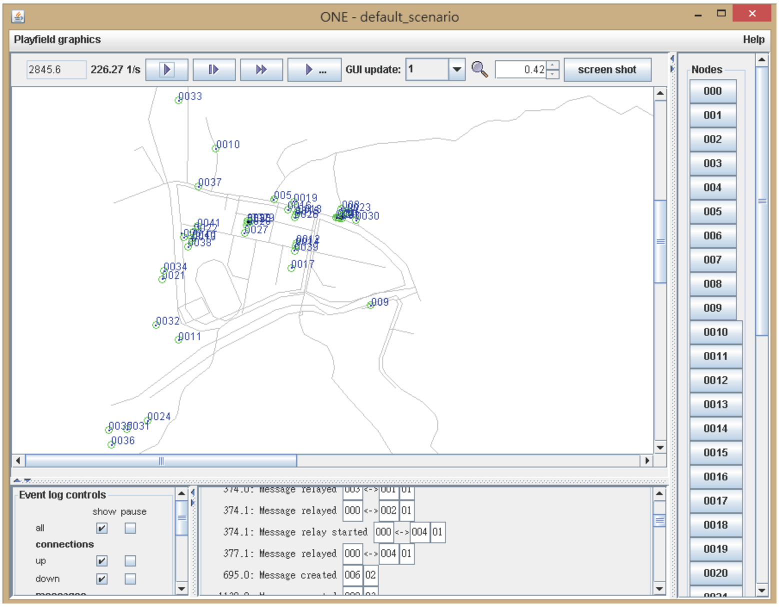


Fig. 10. One simulator

Table 1. Simulation settings

Area	6750*5100 m
Simulation time	172800 s
Data rate	250 KBps
Radio range	10 m
Message size	500 KB $\sim$ 1 MB
Buffer size	100 MB
Total message created	62
Time to live	1080 min

5.3 Simulation Results

We simulated our trace file with both Building-based and Social-based routing algorithms, and we used the setting that we have described above. We evaluated the performance by checking the deliver rate and the overhead. If we can get higher deliver rate with lower overhead, it means that we get better performance. The deliver rate is what we cared about most. The results are as follows.

5.3.1 Delivery Ratio

Formula (5) is the way we counted the deliver rate. Because we had various destinations for one message (m_i), we had to know how many destinations the message would go to ($DestinationNum$) and whether these destinations received the message

Fig. 11. NCCU surrounding area

$(DestinationRelayed)$. The total message number $(TotalMessageNum)$ created in one day was 62 messages. After we got the 2 parameters above, we could calculate the mean deliver rate of messages (M) per day. Finally, we could get the deliver rate.

$$Delivery\ Ratio = \frac{\sum_{m_i}^{M} \frac{DestinationRelayed}{DestinationNum}}{TotalMessageNum}, \forall m_i \in M \tag{5}$$

Figure 12 displays the result that our routing algorithms have a better performance than MaxProp and PRoPHET, but they still have something to improve to reach the Epidemic.

In traditional simulation environment, the nodes can only be influenced by some particular factors like the nodes are in particular group, so there is a key routing feature that can be used. Some of the simulation results could reach the performance of Epidemic. But in reality, there are too many factors to affect people's behavior. For example, people may catch a cold, so they have to go to the hospital or clinic instead of working space. Furthermore, students may skip the class and go out to get some fun. Besides, we may run into somebody we know and then go to a coffee shop, which is not in our plan. So many factors can have an impact on our lives. It is hard to forecast the future precisely, and this is why we cannot beat the Epidemic routing algorithm.

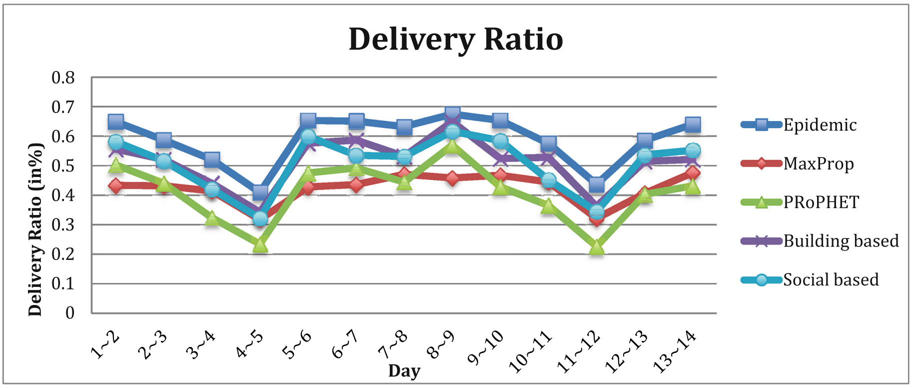


Fig. 12. Delivery ratio

5.3.2 Overhead

In order to have a better understanding, we calculated overhead with overhead copies and overhead ratio. The calculating formula is shown below. And we separate into two parts of overhead.

$$Overhead\ Copies = \frac{\sum_{m_i}^{M} Relayed - DestinationRelayed}{TotalMessageNum}, \forall m_i \in M \tag{6}$$

$$Overhead\ Ratio = \frac{\sum_{m_i}^{M} \frac{Relayed - DestinationRelayed}{DestinationRelayed}}{TotalMessageNum}, \forall m_i \in M \tag{7}$$

In Figs. 13 and 14, we can see that we have lower overhead than Epidemic, and we have a little bit lower overheads than MaxProp and PRoPHET relatively. Combined with Fig. 12 proves that the Building-Based and Social-Based routing algorithms have a better performance than the three traditional routing algorithms.

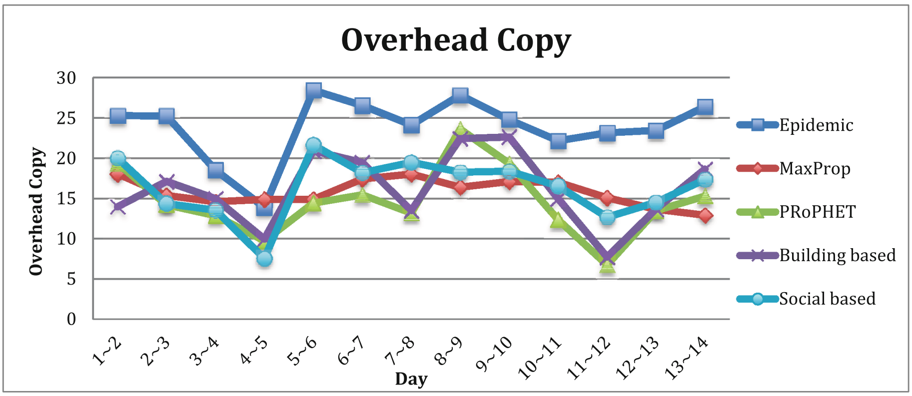


Fig. 13. Overhead copy

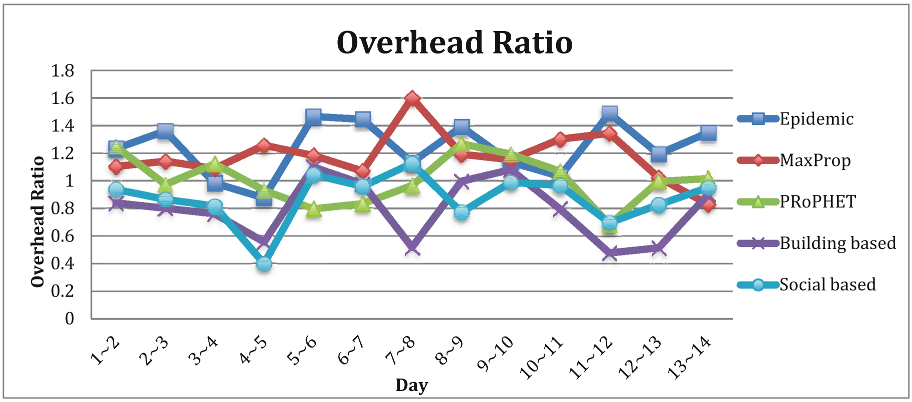


Fig. 14. Overhead ratio

6 Conclusion and Future Work

In this paper, we propose Building-Based and Social-Based routing algorithms. We use the trace file and the interests of people, which were collected in our experiment. Nodes in our experiment can not only be a destination node, but also be a relay node. It is helpful to spread messages. When two nodes meet, they will exchange metadata. We check whether the nodes are destinations and whether they are good relay nodes to make sure that the message transferring is efficient. Finally, we evaluate our routing algorithms with other algorithms, and the result shows that our algorithms have a better performance. In the future, we will consider using the real social relationship between each node. We think if two nodes know each other, they will have a higher probability to meet again. If this hypothesis is true, it will be a good feature to check whether the node is a good relay node or not.

References

1. Tsai, T.-C., Chan, H.-H.: NCCU trace: social-network-aware mobility trace. Commun. Mag. IEEE **53**, 144–149 (2015)
2. Burgess, J., Gallagher, B., Jensen, D., Levine, B.N.: MaxProp: routing for vehicle-based Disruption-Tolerant Networks. In: Proceedings of IEEE Infocom, April 2006
3. Lindgren, A., Doria, A., Schelen, O.: Probabilistic routing in intermittently connected networks. In: Dini, P., Lorenz, P., de Souza, J.N. (eds.) Service Assurance with Partial and Intermittent Resources (SAPIR 2004). LNCS, vol. 3126, pp. 239–254. Springer, New York (2004)
4. Zhu, K., Li, W., Fu, X.: Rethinking routing information in mobile social networks: Location-based or social-based? Elsevier Computer Communications (to appear), 2014
5. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. Pers. Ubiquit. Comput. **10**(4), 255–268 (2006)

6. Hui, P.: People are the network: experimental design and evaluation of social-based forwarding algorithms, Ph.D. Dissertation, UCAM-CL-TR-713. University of Cambridge, Computer Laboratory (2008)
7. Srinivasan, V., Motani, M., Ooi, W.T.: Analysis and implications of student contact patterns derived from campus schedules. In: Proceedings of ACM MobiCom, Los Angeles, CA, pp. 86–97 (2006)
8. Zhang, Y., Gao, W., Cao, G., Porta, T.L., Krishnamachari, B., Iyengar, A.: Social-aware data diffusion in delay tolerant MANET. In: Thai, M.T., Pardalos, P.M. (eds.) Handbook of Optimization in Complex Networks: Communication and Social Networks. LNCS, vol. 58, pp. 457–481. Springer, New York (2010)
9. Daly, E., Haahr, M.: Social network analysis for routing in disconnected delay-tolerant MANETs. In: Proceedings of ACM MobiHoc (2007)
10. Socievole, A., De Rango, F., Caputo, A.: Wireless contacts, Facebook friendships and interests: analysis of a multi-layer social network in an academic environment. In: 2014 IFIP Wireless Days (WD), IEEE (2014)
11. https://github.com/NCCU-MCLAB/NCCU-Trace-Data
12. Keränen, A., Ott, J., Kärkkäinen, T.: The ONE simulator for DTN protocol evaluation. In: Proceedings of the 2nd International Conference on Simulation Tools and Techniques, March 2009