How Small Labels create Big Improvements

Pan Hui
Computer Laboratory
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD United Kingdom
pan.hui@cl.cam.ac.uk

Jon Crowcroft
Computer Laboratory
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD United Kingdom
jon.crowcroft@cl.cam.ac.uk

ABSTRACT

It is widely believed that identifying communities in an ad hoc mobile communications system, such as a pocket switched network, can reduce the amount of traffic created when forwarding messages, but there has not been any empirical evidence available to support this assumption to date. In this paper, we show through use of real experimental human mobility data, how using a small label, identifying users according to their affiliation, can bring a large improvement in forwarding performance, in term of both delivery ratio and cost.

1. INTRODUCTION

Pocket Switched Networks (PSN) [7] represent one particular intermittent communication paradigm for mobile radio devices. With the rapid increase in the number and diversity of mobile devices today, we believe that PSN is likely to be a popular communication architecture in the very near future. Traditional naive multiple-copy-multiple-hop flooding schemes have been empirically shown to work well in dense environments such as academic conferences, and provide fair performance in sparser settings, such as city-wide communications in term of delivery ratio and delay [2]. However, in terms of delivery cost, the naive approach is far from satisfactory, as it creates a lot of unwanted traffic as a side-effect of the delivery scheme, and the overhead rapidly becomes unacceptable in a contentious, vulnerable and resource scarce mobile network.

In the research community, it has been a widely held belief that identifying community information about recipients can help select suitable forwarders, and reduce the delivery cost compared to naive "oblivious" flooding. This is a very reasonable intuition, since people in the same community are likely to meet regularly, and hence be appropriate forwarders for messages destined for other members of their community. However, to date as far as we are aware, there has been no experimental evaluation of this belief, and no-one knows whether it is valid or not.

We created a human mobility experiment during IEEE Infocom 2006, with the participants labelled according to their academic affiliations. After collecting 4 days of data during the conference period, we replayed traces using an emulator, and we discovered that a small label indicating affiliation can indeed effectively reduce the delivery cost, without trading off much against delivery ratio. The intuition that simply identifying community can improve message delivery turns out to be true even during a conference where the people from different sub-communities tend to mix together.

2. RELATED WORK

Pocket Switched Network is a sub-class of Delay Tolerant Networking (DTN) and Mobile Ad Hoc Networks. DTN explores networking in the presence of networks with challenged networking conditions, such as links which are often disconnected, or are subject to long delays. To date, activities in the DTN area have addressed various communication environments where standard Internet protocols would be difficult to use or would provide very poor performance, e.g. networking using buses following predictable routes [5], interplanetary networking [1], interfacing with sensor networks, and using mobile nodes to bridge data between remote village networks and the Internet [6]. Pocket Switched Network is focus on mobile human applications.

Forwarding strategies under intermittently connected mobile Ad hoc networks have been explored by a number of research groups. Chaintreau et al. [3] presented an analytical foundation on the impact of human mobility on the design of opportunistic forwarding algorithms based on six real human mobility traces from four different research groups. Lindgren et al. [9] considered the community concept for control flooding. They have the assumption that nodes mainly remain inside their community and sometimes visit others. To route a message to a destination, a node may transfer that message to a node that belongs to the same community as the destination. Their work provide a well theoretical hypothesis for community based routing, but there is not yet any empirical evaluation. Musolesi et al. [10] proposed community based mobility model for mobile ad hoc research. Hui et al. [8] did a similar measurement of human mobility in conference environment but community issues haven't been touched. In this paper, we are looking at the problem on an experimental approach, the empirical result could be helpful for both modelling and theoretical work of other research groups.

3. EXPERIMENTAL SETUP

The device used to collect the contact opportunity data and mobility statistics in this experiment is the Intel iMote. This is a small platform designed for embedded operation, comprising an ARM processor, Bluetooth radio and flash RAM. We packaged these devices in a dental floss box, due to its ideal size, low weight, and hard plastic shell.

Eighty of these boxes were distributed to attendees at the IEEE Infocom conference in Barcelona in April 2006. The participants are specially selected so that thirty-four out of eighty form four subgroups according to academic affiliations. We intended to have all the participants divided into subgroups, but due to the modest participating level we could only have four subgroups. Out of these four groups, there were two groups from institutes in Paris with size of four and ten respectively, and there is one group from Switzerland of five people, and another, larger group of 15 people from the local organization in Barcelona. All the participants were asked to keep the iMotes with them for as much of their day as possible during the conference period.

The iMotes were configured to perform a Bluetooth baseband layer inquiry discovering the MAC addresses of other Bluetooth nodes in range, with the inquiry mode enabled for five seconds. Despite the Bluetooth specification recommending that inquiry last for ten seconds, preliminary experiments showed that five seconds is sufficient to consistently discover all nearby devices, while halving the power- expensive inquiry phase. Between inquiry periods, the iMotes were placed in a sleep mode in which they respond to enquiries but are not otherwise active, for a duration of 120 seconds plus or minus twelve seconds in a uniform random distribution. The randomness was added to the sleep interval in order to avoid a situation were iMotes timers were in sync, since two iMotes performing inquiry simultaneously cannot see each other.

4. LABELLING BEFORE THROWING OUT

The forwarding scheme we used here is called the "labelling strategy". We imagine that each node has a label telling others about its affiliation/group, just like the name badge in the conference. The strategy chosen is exclusively to forward messages to destinations, or to next-hop nodes belonging to the same group (same label) as the destinations. The assumption we made here is that people from the same affiliation tends to meet more often than people outside the affiliation and hence can be good forwarders to relay messages to the other members in the same affiliation/with the same label. This strategy required very little information about each individual and is believed to be very easy to implemented in the real life, by just tapping a mobile device and write down the affiliation of the owner, that is what we usually required to input when we have a new PDA.

5. ANALYSIS INTER-CONTACT TIMES

The assumption we made for the whole work is people from the same group would tend to interact more often then with people from different group. We want to look from our data set to see whether this assumption is valid before doing any further performance investigation. Inter-contact times distribution is a good indication for this relationship. For a given pair of nodes A and B, the time-line can be divided into two regions, contact times and inter-contact times. The contact times are when A and B are in range of one another, and could therefore have sent data if they had wished to. Inter-contact times are the times between the contact times, so the Inter-contact times distribution simply indicates the frequency of interaction. Previous work [8] has shown that inter-contact time follows a power-law distribution, the bigger that value of the power coefficient, the more frequently the nodes pair interact. In this work, we extend to look at inter-contact distribution for all the nodes inside a group and also the inter-contact distribution between two groups. We believe the power-law coefficient of these inter-group inter-contact time distribution, if they are following power law, indicate the closeness of two groups.

Figure 1(a) shows a typical inter-contact time distribution for a pair of nodes in one community, and a pair of nodes from different communities. We can see that the intra-community pair has higher power law coefficient than the inter-community pair; that is, nodes-pair in the same community tend to meet more often. Figure 1(b) also shows the inter-any-contact time distribution for a node with all the other nodes in its community, as well as for all other nodes in another community; we can also observe a variation in the power law coefficient. To avoid the bias causing by a single node, we also calculate the inter-any-contact time distribution for all the nodes within a same group and also the same distribution with nodes all outside the group, the results are shown in Figure 2. We can see a significant deeper slope for the intra-community inter-any-contact distribution. This provides empirical evidence that people from the same organization tend to meet more often than people from different organizations. This provides good hints to identify forwarders for messages delivery.

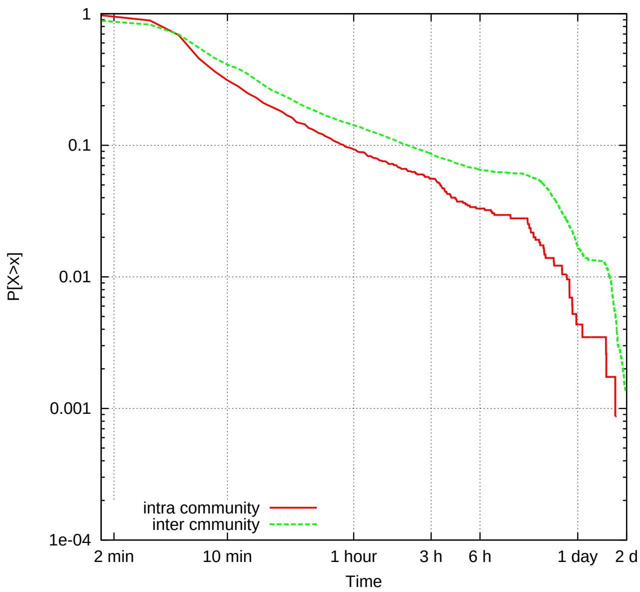


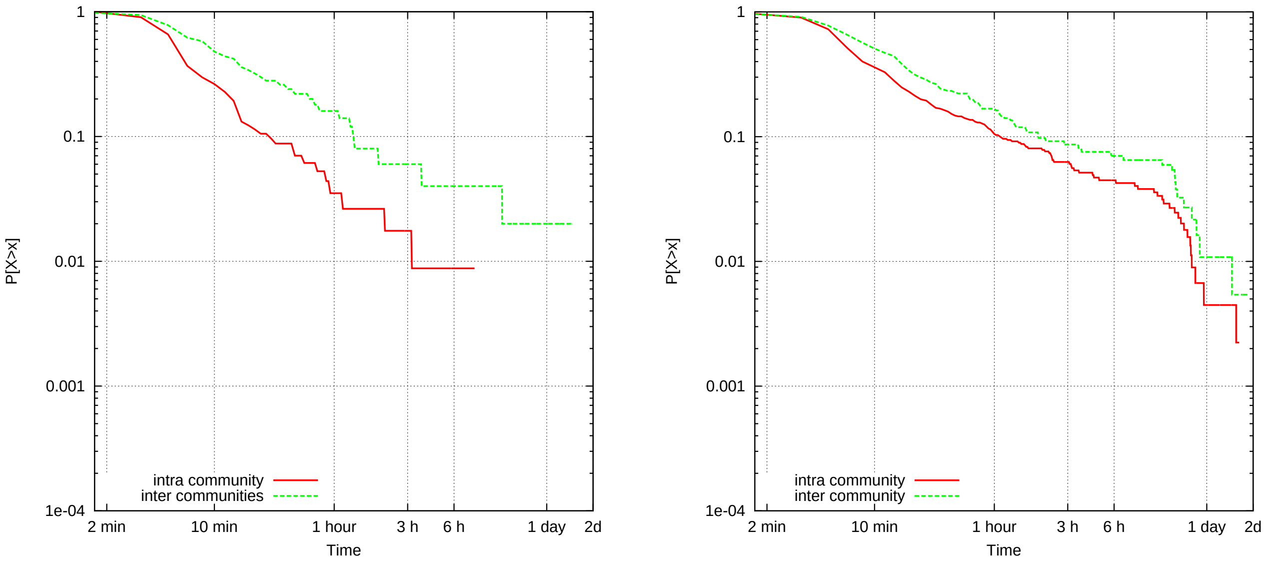
Figure 2: The inter-any-contact time distributions for all nodes inside a same group and also with all other nodes outside the group.

Here we also want to introduce the concept of friendship communities[4]. The University of Cambridge's Computer Laboratory researchers may be a friendly community towards Intel Research Cambridge staff, since these two groups of people use the same building and have a number of collaborations. Hence people from one group may be good forwarders for people in the corresponding friendship group. In our experimental data, we happen to have two groups from Paris, and we want to look at whether they have closer relationship when compared to other groups, based on the inter-contact time distribution. As shown in Figure 3, we can see that within the same group, the power law coefficient is the largest, and next largest with the nodes in a friendship group, and lowest for a normal group. Although the different is not very significant, but we can still observe it. Later, we will also look at how this friendship community can be used explicitly to help forward messages.

6. EVALUATION METHODOLOGY

6.1 HaggleSim Emulator

In order to evaluate different forwarding algorithms, we developed an emulator called *HaggleSim*, which can replay the mobility traces we collected and emulate different forwarding strategies on every contact event. This emulator is completely driven by contact



(a) inter-contact (b) inter-any-contact

Figure 1: Comparison of inter-contact and inter-any-contact time for intra and inter community nodes

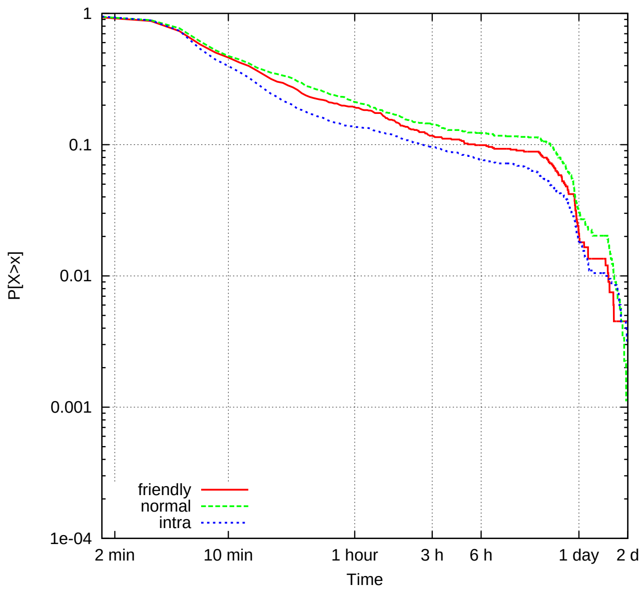


Figure 3: The inter-any-contact time distributions for all nodes inside a same group, with a friendship group and a normal group.

events. The original trace files are divided into discrete sequential contact events, and fed into the emulator as inputs. The event granularity depends on our choice of balance between replay speed, and the degree of accuracy desired. In all the simulations in this work, we divided the traces into discrete contact events with an granularity of 100 seconds. We analyze the successful delivery rate, the delivery cost, the delay distribution, the hop distribution of all the successful delivery, and the popularity of a node as relay based on the log files produced by the emulator.

Table 1 shows a snapshot of our emulation source file. Our simulator reads the file line by line, treating each line as a discrete encounter event, and makes a forwarding decision on this encounter based on the forwarding algorithm under study. As we see from the source file, some events happen at the same time stamp, and should be treated as simultaneous. Just reading the file line by line artificially imposes an order on events, so instead we keep contacts in a buffer while reading line ahead: if the next event happens at the same time, and there would be an exchange of messages between the nodes referenced in the line due to the forwarding law, we will re-read the contacts in the buffer, and apply the same forwarding law to the newly exchanged messages. This removes the artificial ordering.

Node	Node	Time Stamp
6	24	14991
25	11	14991
25	240	14991
24	240	14991
6	240	14991
35	511	14991
24	11	14991
6	24	15001
25	11	15001
25	240	15001

Table 1: Snapshot of Simulation Source File

6.2 Simulation Parameters

There are three parameters we used in our simulation to achieve control flooding in MCP strategy.

- *Number of copies*: The maximum number of duplicates of each message created at each node.
- *Number of hops*: The maximum number of hops, counted from the source, that a message copy can travel before reaching the destination; this is similar to TTL value in the Internet.
- *Time TTL*: The maximum time a message can stay in the system after its creation. This is to prevent expired messages from further circulation.

6.3 Performance Metrics

For all the simulations we have conducted for this work, we have measured the following metrics:

- *Delivery ratio*: The proportion of messages that have been delivered out of the total of messages created.
- *Half-life delivery time TTL*: This is the time TTL value that would allow half of the messages created to be delivered; in other words it is equivalent to the delay time that half of the created messages experienced. It measures how fast and efficient a forwarding strategy for messages delivery.
- *Hop-distribution for deliveries*: The distribution of the number of hops needed for all the deliveries. This metric gives some ideas of how a forwarding strategy picks forwarders. In the LABEL strategy, it reveal the social distance between sources and destinations.
- *Delivery cost*: For the cost, we measure the total number of medium accesses; that is the total number of messages (include duplicates) transmitted across the air. To normalize this, we divide it by the total number of unique messages created.

6.4 Simulation Scenario

In order to study exclusively the effect of community on forwarding, we created the following scenario: all the seventy-seven nodes [1] create 1000 messages, destined only to the thirty-four nodes belonging to the four groups; the message creation times are uniformly distributed throughout the experimental duration.

In order to compare the performance of the labelling strategy with a naive strategy, we run an emulation of the multiple-copy-multiple-hop (*MCP*) strategy. To ensure that the performance improvement is not due to randomly limited number of forwarders, for every round of simulation, we created four random groups of same group sizes as the original groups but with nodes randomly selected from all the seventy-seven nodes. We refer to the labelling strategy and the control experiment as *LABEL* and *CONTROL* respectively, in our analysis. To achieve statistical fairness, we run the emulation 20 times with different traffic patterns.

7. RESULTS AND ANALYSIS

In this section, we compare the performance of four strategies, MCP, LABEL, CONTROL, and one further approach, which we call *WAIT*; this entails waiting until the source of a message has direct contact with the destination. In this simulation we used 4-copy-4-hop for the MCP scheme which has been shown by experiences to be the best naive scheme in this kind of conference scenario in term of delivery and cost.

In Figure 4(a) we see that, as expected, LABEL has a delivery ratio between MCP and WAIT, and the trend is for it to approach closer to the performance of MCP, as the allowed time TTL of the messages increases. In term of cost, in Figure 4(b) we can see that MCP costs much more than LABEL, especially when the time TTL is increased up to 1 day, where MCP has less than a 10% improvement over LABEL, but it has around 6 times higher cost. Of course, WAIT has the lowest cost: since it is during a conference, it doesn't need to wait too long time to meet the destination, hence the delivery ratio is not too low.Figure 5 shows the number of other nodes meet directly by each node during the experimental period. It shows that almost every node has the chances to meet most of the other nodes during that period. However, if we look at the half-life delivery value, we can see that the half-life delivery is 3 hours for MCP, 9 hours for LABEL and around 1 day for WAIT. In other words, if you can tolerate a 1 day delay, you could use the WAIT strategy, otherwise LABEL would perform the best in term of delivery ratio, delay and cost.

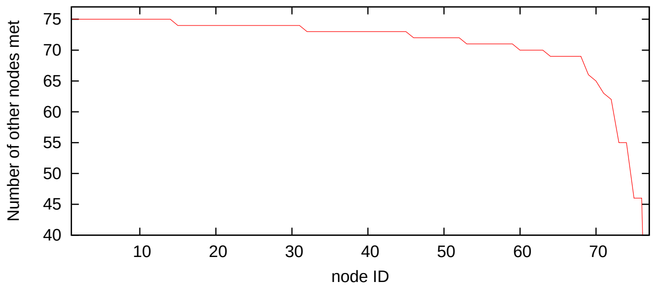


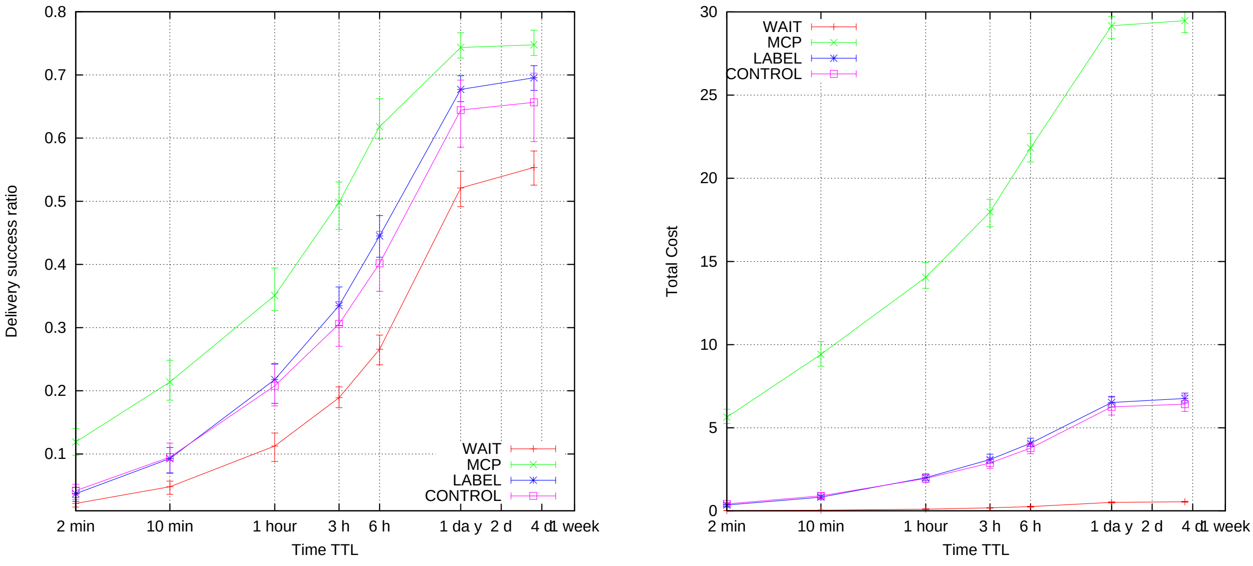
Figure 5: The number of other nodes met directly by each node during the experimental period.

The random generated groups scenario CONTROL is only 5% worse in term of average delivery than LABEL in many cases, but it has a larger confidence interval than LABEL, and the lower bounds are usually around 10% lower than the average value of the LABEL strategy. This means that a badly generated combination of groups would affect the delivery quite significantly. Considering that 34 out of the 78 nodes have a community relationship, it is not difficult for the random groups generated to consist of members which belong to real affiliation groups. Furthermore, considering that during a conference, people from different research groups often mix together, this kind of performance is reasonable.

Figure 6 shows further the improvement of LABEL's performance compared to MCP in term of delivery against cost. Each point represent a different time TTL value, and we can see that as we vary this, the delivery almost varies linearly with the cost. The anomaly after the cost is equal to 22 on the MCP case, where the slope slightly decreases, is because the system is going into saturation, and further increase in cost brings slower increase in the delivery. Clearly LABEL has a much deeper slope than MCP: this means that this strategy is much more cost effective.

In Figure 7, we also look at the hops distribution, which is the distribution of the number of hops required for all the delivered messages. In this case, we set the time TTL to 3 hours without lose of generality. The x-axis shows the number of hops and the y-axis shows the probability for a message to be delivered with at least

[1] Because of hardware problem, three out of eighty did not yield any data.



(a) Delivery

(b) Cost

Figure 4: Comparison of delivery ratio and cost of different strategies

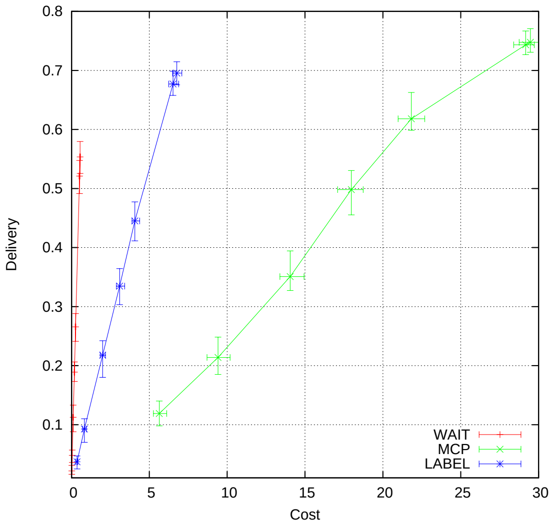


Figure 6: The Delivery-Cost Graph for MCP, LABEL and CONTROL strategies.

that number of hops. For this value of time TTL, an maximum of 50% of the total messages created can be delivered(the y-value at 1-hop). For MCP, half of the deliveries are on using 4 hops (the y-value at 4-hop minus the y-value at 5-hop), that is because the MCP sends out messages on a blind first-come-first-send approach, the 4th hops will help to reach the most people. But instead, for the LABEL, the delivery are almost the same for 1 hop to 4 hops with a slightly bigger on the 3 hop cases. The direct contact case (1 hop) is only help to deliver less than 10% of the messages, much of the delivery is relied on the intermediate social relays.

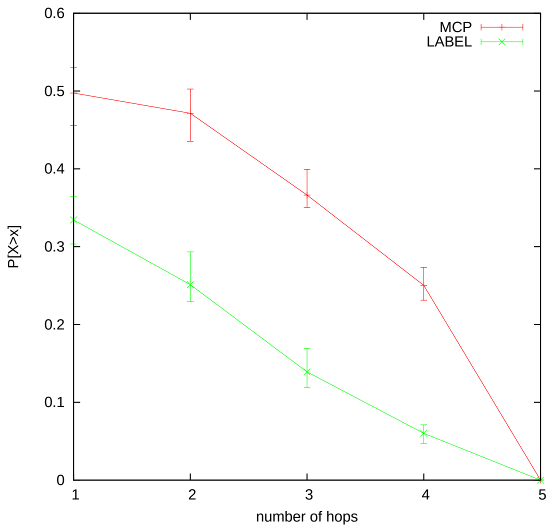


Figure 7: The hops distribution for all the deliveries in MCP and LABEL.

In order to ascertain whether the friendship group concept is helpful for message delivery, we ran another series of simulations. In these, members of different friendship groups are allowed to act as relays for each other. We assigned the two groups from Paris to be friendship-groups of each other, and they help to relay messages. The result we expect to see is that the use of friendship groups can help to improve the delivery ratio, without too much increase in the delivery cost. A controlled experiment is also done, by using a random group chosen as a friendship group, rather than one with a known affiliation.

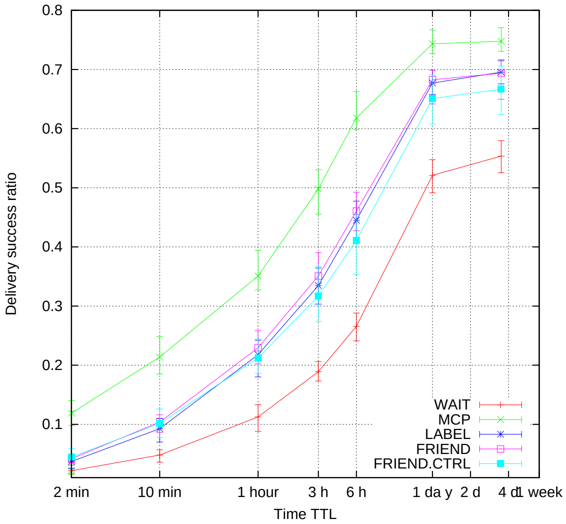


Figure 8: Comparison of delivery ratio on different strategies, with friendship groups.

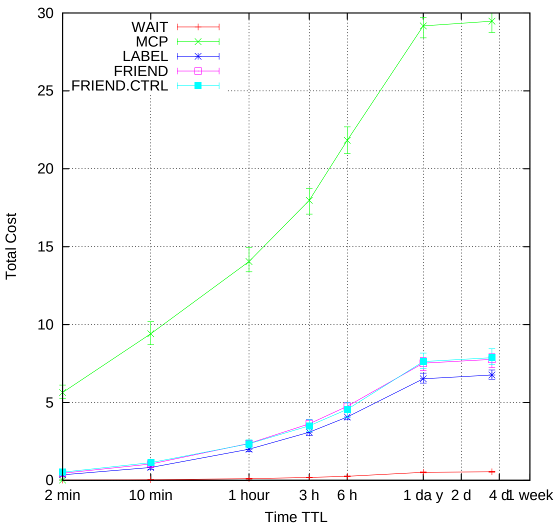


Figure 9: Comparison of delivery cost on different strategies, with friendship groups.

We can see from Figure 8 and Figure 9 that the friendship group did indeed help to improve delivery, with only slightly increased cost, Also, as expected, the randomly generated friendship group just increased the cost, without any improvement in delivery. We know it is difficult to study group and friendship group behaviour in a conference, since the people often mixing promiscuously: that is one of the purpose of a conference; but we can still make some observations on the correlation we see. We further believe that the techniques and metrics we have developed here can be used for research on friendship groups with more easily specified boundaries.

8. CONCLUSIONS

The addition of even a very small amount state information, in our experiments just an affiliation label used to choose nodes preferentially, is shown to bring significant improvement in forwarding performance over oblivious or naive forwarding algorithms in pocket switched networks.

We need to embed this kind of state information in our future designs for forwarding strategies, and we have shown that labels that identify community provide a good start.

9. REFERENCES

[1] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss. Delay-tolerant networking: an approach to interplanetary internet. *IEEE Communications Magazine*, 41(6):128–136, 2003.

[2] A. Chaintreau, P. Hui, J. Crowcroft, and C. Diot. An empirical analysis of controlled flooding algorithms in pocket switched networks. Technical Report UCAM-CL-TR, University of Cambridge.

[3] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. INFOCOM*, April 2006.

[4] R. Dunbar. Grooming, gossip and the evolution of language, 1996.

[5] K. Fall. A delay-tolerant network architecture for challenged internets. In *Proc. SIGCOMM*, 2003.

[6] R. Fletcher, A. Hasson, and A. Pentland. Daknet: Rethinking connectivity in developing nations. *IEEE Computer*, 37(1):78–83, 2004.

[7] P. Hui, A. Chaintreau, R. Gass, J. Scott, J. Crowcroft, and C. Diot. Pocket switched networking: Challenges, feasibility, and implementation issues. In *Proc. WAC*, 2005.

[8] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *Proc. WDTN*, 2005.

[9] A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in intermittently connected networks. In *Proc. SAPIR*, 2004.

[10] M. Musolesi and C. Mascolo. A Community based Mobility Model for Ad Hoc Network Research. In *Proceedings of the 2nd ACM/SIGMOBILE International Workshop on Multi-hop Ad Hoc Networks: from theory to reality (REALMAN'06)*. ACM Press, May 2006.