Social Network Aided Multicast Delivery Scheme For Human Contact-Based Networks

Mooi Choo Chuah
Lehigh University
Bethlehem, PA 18015
(01)610-758-4061
chuah@cse.lehigh.edu

ABSTRACT

Mobile devices carried by people form dynamic networks. Understanding the social structures within the human mobility traces captured from the mobile devices help us to design more efficient message dissemination schemes. People who are in multiple communities are good message carriers. Thus, the ability to identify the different communities efficiently from the various communication traces e.g. contact traces from users' mobile devices is important. In this paper, using some human mobility traces from the real world, we first identify nodes that can play connector roles using some social network metrics e.g. position role centrality. Then, we investigate the usefulness of utilizing the connector nodes information in the design of multicast delivery schemes in human contact-based networks. Our results indicate that using such information can achieve similar delivery performance as the multi-copy epidemic scheme but at a much smaller communication cost.

Categories and Subject Descriptors

C2.2 [**Network Protocols**]: Routing protocols

General Terms

Algorithms, Performance, Design, Experimentation.

Keywords

Social Network Analysis, Multicast Delivery Scheme, Human Contact Based Networks

1. INTRODUCTION

In recent years, many powerful yet small wireless devices e.g. PDAs have emerged in the market. Many users carry such mobile devices and use them to access information anywhere anytime. These mobile devices can be used in peer-to-peer mode to disseminate messages based on human mobility. Pocket Switched Network (PSN) [1] is one such example. Several researchers have attempted to design new forwarding schemes for such human-based networks e.g.[1],[2]. However, most of these efforts are only focusing on delivering messages from one person to another. It is interesting to explore how one can design multicast delivery scheme where messages need to reach multiple recipients. Providing a multicast delivery feature in DTN is very useful e.g. a message needs to be sent to members of a emergency response team, software updates need to be pushed to some mobile devices in a battlefield scenario. Several DTN multicast delivery schemes [15],[17] have been designed in the past but none of the researchers have evaluated their schemes using human contact-based traces. None of these designed schemes utilize social-network based metrics. In human contact-based networks, there tend to be communities of nodes and some nodes are more likely to belong to different communities at different times. We refer to nodes that can communicate with different communities of nodes as the connector nodes. Further, it is likely that nodes do not form a connected graph all the time. Thus, identifying such connector nodes that can bridge between various partitioned groups of nodes is very important for such nodes can be used as forwarding agents across different communities. In summary, we believe that a multicast scheme that is designed using the connector node and the community information can deliver multicast messages more efficiently than those schemes that do not make use of such information.

In this paper, we first present the techniques we use to identify communities of nodes and connectors (i.e. nodes that can reach multiple communities and many nodes). Then, we design a social network aided multicast delivery (SNAMD) scheme that makes use of such information to efficiently deliver multicast messages. Via simulations, we compare the delivery performance of our designed scheme with a multicopy epidemic scheme as well as the BubbleRap scheme which is designed for unicast message delivery [2]. We use two human contact-based traces in our simulation studies. One came from the Haggle Project [1],[2] and the other from the MIT Reality Mining Project [10]. Our results indicate that the SNAMD scheme we design can achieve comparable performance as the multicopy epidemic scheme but at only 1/3 to 1/2 of the communication cost of the multicopy scheme.

The rest of the paper is organized as follows: In Section 2, we discuss related work, including some existing DTN unicast and multicast forwarding schemes which are designed for disruption tolerant networks. In Section 3, we discuss a community extraction scheme for contact traces of mobile devices and some metrics which are often used to identify connector nodes in such traces. We also discuss an enhanced scheme for determining key role nodes for multicast delivery. In Section 4, we describe our social network aided multicast delivery (SNAMD) scheme. In Section 5, we present our simulation set up, and discuss the simulation results we obtain. We conclude by discussing some future work in Section 6.

Simplex'09, July 1, 2009, Venice, Italy.

2. Related Work

Researchers in [1],[2],[12] have studied the characteristics of the human mobility traces in terms of the devices' intercontact time, the contact durations etc. In addition, based on such characteristics, some researchers have proposed more efficient forwarding schemes [1],[12] for such networks. The Bubble-Rap scheme presented in [2] is the most relevant to our work. It is designed for unicast message delivery. The scheme assumes that each node has a global ranking (i.e. global centrality) across the whole system and a local ranking within its local community. If a node has a message destined to another node, this node first bubbles the message up the hierarchical ranking tree using the global ranking, until it reaches a node which is in the same community as the destination node. Then, the local ranking system is used to relay the message until the destination is reached or the message expires. However, not much has been done in studying the community structures of these traces that can help in designing more efficient one-to-many, and many-to-many message dissemination strategies.

Multiple DTN multicast schemes have also been designed [15],[17]. However, many of these schemes either use a variant of the Dynamic Source Routing (DSR) route discovery process or are based on node density. The authors do not evaluate their schemes using human contact-based traces and do not make use of social network related information in their schemes. In addition, their approaches cannot work in human contact-based traces because their route discovery process does not work well in such sparse networks.

3. Community Detection and Connector Identification Schemes

3.1 Community Detection Scheme

Many centralized community detection methods have been proposed for social network data in the literature. Recent review papers in this area include [3],[4]. Such centralized methods are useful for offline data analysis on collected mobility traces to explore structures in the data. However, for self-organizing networks, it will be useful to have a distributed community detection scheme which allows each mobile device to detect its own local community. Here, we give a quick summary of a community detection scheme we use in our work. It is the K-clique distributed community detection scheme presented in [2].

To understand how K-clique works, we need the following definitions [8]:

Familiar set: each node, say *a,* keeps a list of nodes they have encountered with the corresponding cumulative contact durations. When the cumulative contact duration with a node, say *b,* exceeds a certain threshold, T_{th} , then node b is included in node *a*'s familiar set, *F*.

Local Community: a node's local community, denoted by *C*, contains all the nodes in its familiar set (its direct neighbors) and also the nodes that are selected by the community detection algorithm.

The K-clique scheme works as follows [8]:

1. Each node, v_o, maintains the following information: a list of nodes it encounters (or communicates with), and the contact duration (or number of emails exchanged), its familiar set, F_o , its local community, C_o detected so far. And a local approximation of the familiar sets of all vertices in its local community C_0, *denoted as FoC(v).*
2. Initialization: $C_o = \{v_o\}$, $F_o = \Phi$ (empty set), $FoC_o = \Phi$
3. When v_o encounters another node v_i, they exchange local information, i.e., v_o will acquire C_i, F_i and FoC_j from v_i . Each local approximation of familiar set in FoC_0 is merged with the corresponding versions just obtained from FoC_j .
4. If v_i is not in F_o , v_o updates the total contact duration (or total number of emails exchanged) of v_i until v_i falls out of contact and meanwhile the algorithm forks and proceeds to Step 5. When the total contact duration (or total number of email exchanges) has exceeded a certain threshold, v_o will insert v_i in F_o and C_o .
5. If v_i is not in C_o , then add v_i to C_o if it satisfies the following criteria: $| F_i \cap C_o | > (k-1)$
6. If v_i is added to C_o in the previous steps, the aggressive variants of the algorithm behaves as follows: If $| F_j \cap C_o | > (k-1)$, then, the two communities are merged. If this criteria is satisfied, FoC_0 is also updated to include F_j .

In our experiment, λ is set to 0.6 and k is set to 4.

3.2 Identifying Connectors

Social networking researchers often use the term "connectors" (defined by Gladwell [5]) to refer to individuals who are acquainted with many individuals from different circles, and can be the bridging nodes between disjoint or weakly-linked social groups. Different social-network based centrality metrics can be used to identify nodes that play key roles in their networks. Below, we discuss two that we consider in our work:

1. Position Role Centrality, $\mathbf{C_{pr}}$ **[7]**, is computed based on the network efficiency concept. C_{pr} is computed by subtracting from the current network efficiency the new network efficiency obtained by removing the ego (the node being analyzed) from the current network as shown in the following equation

$$C_{pr} = E(G) - E(G\text{-}v_i) \quad \text{Eqn (1)}$$

where E(G) is the efficiency of the network represented by graph G, and (G-v_i) is the graph obtained by removing v_i (the node for which we are currently computing Position Role Centrality) from graph G.

Network Efficiency is computed as follows: $E(G) = \frac{1}{N*(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$, where *N* is the number of nodes in the network graph and d_{ij} is the length of the shortest path between nodes i and j.

As connectors are individuals who hold the network together, they are defined herein as being the nodes with values of **Position Role Centrality** exceeding a certain threshold. The removal of connectors would affect greatly the information flow within the network.

2. Betweenness Centrality, $C_B(v)$, is computed as follows [6]: $C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where σ_{st} is the number of the shortest paths from s to t, σ_{ss} = 1 and $\sigma_{st}(v)$ is the number of shortest paths from s to t that include node v. If the nodes with high betweenness values are removed, then a significant number of shortest paths would become unavailable.

There are similarities and differences between the two centrality measures. Position Role Centrality uses the length of the shortest path with and without the considered node while betweenness centrality considers whether there are alternate shortest paths that do not go through the considered node. All shortest paths between two nodes need to be computed before the betweenness centrality can be evaluated. Thus, for large networks, using the position role centrality may be more efficient in terms of computations.

The above definitions treat all nodes to be the same. However, nodes may belong to different communities. Thus, in this work, we use the following method to identify the connectors: we first identify the communities each node belongs to. Then, we extract the nodes that each node encounters, and identify the different communities that they belong to. Thus, we can determine the number of communities each node can encounter. Next, we identify the top 10 nodes that have the highest positional role or betweenness centrality values. Our analysis of the MIT/Infocom 2005 traces indicates that the top 10 lists produced by the above two centrality measures are similar. From these 10 nodes, we identify a subset that will act as the connectors. One should choose a sufficient number of connector nodes that can cover all available alternative paths to the multicast receivers.

4. Social Network Aided Multicast Delivery (SNAMD) Scheme

In this section, we investigate how to use the information related to the connectors and communities to improve the performance of multicast delivery in human-contact based DTNs. In Fig 1, we illustrate how a delivery scheme that uses social network related information improves delivery performance. At time t_0, *S* sends a copy of a multicast message to n_1. At time t_2, n_1 meets a node C_1 that connects communities CL_1 and CL_2. One receiver R_3 is in CL_2. Thus, n_1 forwards a copy of the message to C_1 which eventually delivers the message to R_3 via other nodes within CL_2. n_1 also forwards a copy of the message to n_2 which has a higher delivery predictability to two receivers R_1 and R_2. At time t_1, *S* meets another connector node C_2 that can reach nodes in CL_3 and one of the receivers R_4 is in CL_3. Note that C_1 (C_2) may not belong to CL_2 (CL_3) based on the output of the community detection algorithm discussed in Section 3.1.

Our SNAMD scheme is motivated by the following observations: a node should forward a copy to an encountered node if that encountered node is either in the same community as any of the destination nodes or have higher chances of reaching any of the destination nodes.

Our multicast scheme assumes that (a) each node has an estimate of its contact rate with other nodes in the network (this can be measured from past history), (b) each node knows other nodes' communities, and (c) the system can determine the nodes that are connectors i.e. such nodes can meet many nodes within its own communities and nodes from other communities. The distributed community detection algorithm discussed in Section 3 can be used by each node to determine if it belongs to the same community as another node and beacons exchanged between nodes can be used to disseminate the information regarding the community of other nodes. The connector nodes are identified as follows: we use the position role centrality metric to identify the top 10 nodes. Then, we determine the different communities each of these 10 nodes can encounter and manually select a subset of these nodes to be the connector nodes. For example, nodes 39,57,32,86,14,17 are chosen as connectors for the MIT trace. From our evaluations, this is the optimal choice since adding additional connector nodes result in higher data transmission cost without further improving the delivery performance.

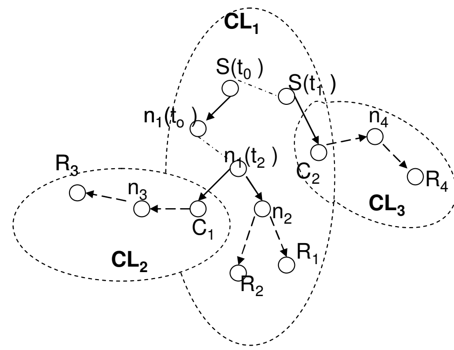


Fig 1: Social Network Aided Multicast Delivery Scheme

The metric used in BubbleRap to decide if node *a* needs to forward a copy to node *b* is (a) the global rank if the two nodes come from different communities and (b) local rank if the two nodes are from the same cluster. However, in our SNAMD schemes, we use (a) the delivery predictability [18] as the metric when the two nodes belong to the same cluster or (b) a node's estimated contact rate to a destination if the two nodes are from different communities.

In summary, our SNAMD scheme works as follows: when a node (say *a*) meets with another node (say *b*), node *a* first determines if node *b* is one of the destination nodes. If so, it will deliver the message. Otherwise, it checks to see if it has more than one copy. If it has more than one copy, it will check to see if node *b* is a connector node. If so, node *a* sends a copy of the multicast message to node *b*. Otherwise, node *a* checks to see if node *b* belongs to the same cluster as itself. If node *b* is from a different cluster, then node *a* determines if any of the multicast recipients are in the same cluster as node *b*. If so, node *a* will send a copy to node *b*. Otherwise, node *a* will only forward a copy to node *b* if node *b* has a higher delivery predictability to at least x% (currently set to 40%) of the multicast recipients than node *a*. We refer to this scheme as the Social Network Aided Multicast Delivery (SNAMDV1) scheme. Another variation of the scheme referred to as the SNAMDV2 scheme works as follows; instead of using the delivery predictability, we use the estimated contact rate of node *b* with any destination node as the metric. If node *b* has a higher contact rate to any of the destination nodes than node *a*, then node *a* will send a copy of the message to node *b*. Details of our schemes can be found in our technical report [16].

5. Simulation Studies

5.1 Simulation Set Up

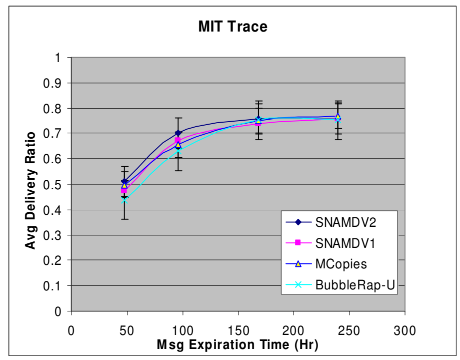
We implemented our schemes in a Java simulator that the authors developed in [2]. We compare these two schemes with the following two schemes: (a) the MCopies scheme where each node will send 4 copies of the message to any 4 nodes it encounters (before the message expires) and each of the relay nodes that receives a copy can in turn spread 4 copies to any 4 nodes it encounters, (b) the BubbleRap-U scheme which is a social-network based unicast DTN delivery scheme described in [2]. The MCopies scheme is an enhanced version of the MCP scheme evaluated in [2] for multicast delivery. The enhancement we made is as follows: any intermediate node that receives a multicast message will spread 4 copies as in the MCP scheme. If a multicast message reaches one of the multicast receivers and its copy value still exceeds one, then that receiver is allowed to spray to other nodes it encounters such that the message can reach other multicast receivers. This scheme is pretty much an epidemic flooding scheme which will have high delivery ratio given sufficiently long message expiration time but incur high overhead. We include the results for the BubbleRap-U scheme for the following reason: some may argue that there is no need to design a multicast delivery scheme for sparsely connected networks. One can replicate as many packets as there are multicast receivers for each multicast message, and then use the BubbleRap-U scheme to deliver these packets individually. We want to show that such a delivery mechanism performs poorer than our SNAMD scheme. The multicast version of the BubbleRap (denoted as BubbleRap-M) scheme is similar to our SNAMDV1 scheme except that no connector is used to improve the delivery performance.

The metrics we used to compare these four schemes are: (i) delivery ratio (DR) which is the total number of messages (y) that reach any multicast recipient divided by the total number of messages sent to each multicast recipient (i.e. DR=y/(Nm*NR) in each run where Nm is the total unique messages sent, and NR is the number of multicast receivers), (ii) average message delivery latency which is the average delay for all the delivered messages, (iii) the total cost which is the total number of data transmissions incurred during the simulations.

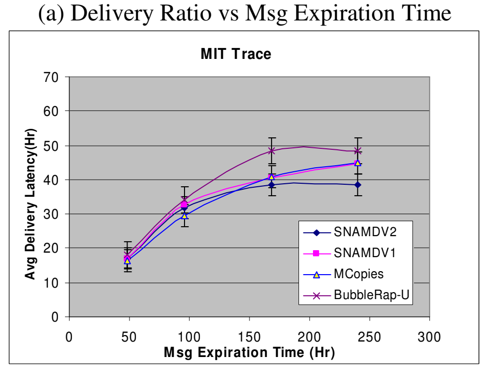
We use both the MIT reality trace and the Infocom 2005 trace. For each simulation run, we use two multicast sessions with each session having five receivers. For the MIT trace, we divide the active nodes into two groups, one group with nodes that have more contacts with other nodes and another group with nodes that have fewer contacts with other nodes. Then, we generate two scenarios. In the first scenario, all multicast recipients are chosen from the first group which have more contacts with other nodes and hence we expect better delivery performance. In the second scenario, all multicast recipients are chosen from nodes in the second group. Hence, we expect the delivery performance to be poorer in Scenario 2. For the Infocom 2005 trace, we merely have one scenario: two multicast sessions with five receivers in each session. The five receivers for each session are randomly chosen from the 77 more active nodes. We create 200 messages in each simulation run. Each message is generated at a time uniformly distributed within the duration of the trace and is chosen either to be from Session 1 or Session 2. In our experiment, we vary the message expiration time from 1 day to 10 days for the MIT trace and from 4 hrs to 18 hours for the Infocom 2005 trace. Each data point is the average of six runs. We replace each multicast message with five unicast messages when we evaluate the BubbleRap-U scheme.

5.2 Simulation Results

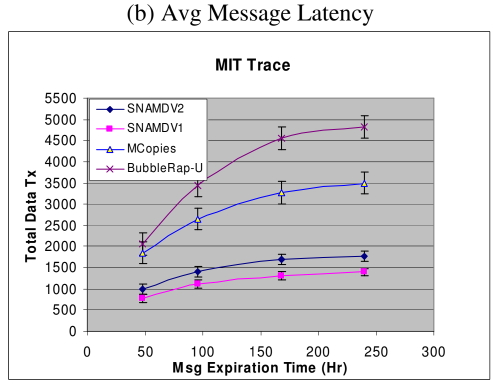
Figs 2 (a)-(c) plot the results for the delivery ratio, the average message latency and the total cost respectively for Scenario 1 (Easy sessions) using the MIT reality trace. It shows that all four schemes achieve similar delivery ratio. The SNAMDV2 achieves slightly better delivery ratio and a lower average delivery latency. The clear distinction is in the transmission cost. Both SNAMDV1 and SNAMDV2 schemes can achieve similar delivery ratio and message latency performance with much smaller transmission cost. Compared to the BubbleRap-U and the MCopies schemes, the two social network aided schemes incur only either 1/3 or 1/2 of the original transmission costs.



(a) Delivery Ratio vs Msg Expiration Time



(b) Avg Message Latency



(c) Total Data Transmission

Fig 2: Delivery Performance

(MIT Reality Trace- Scenario 1)

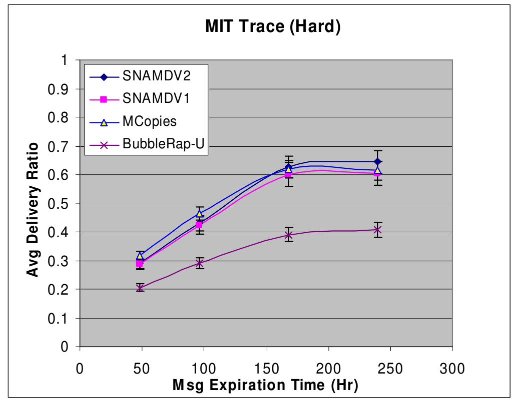
Figs 3(a)-(c) plot the results for the delivery ratio, the average message latency and the total cost respectively for Scenario 2 (Harder to Deliver Sessions) using the MIT reality trace. It is interesting to note that for Scenario 2, the delivery ratio of the BubleRap-U scheme is significantly lower than the MCopies and the two social network aided multicast delivery schemes we designed.

There are two reasons why the two social network aided multicast schemes perform better than the MCopies scheme in terms of the transmission cost. In the MCopies scheme, a node that has a copy of the message sprays a copy to any node that it encounters while the two social network aided schemes only spray a copy when the encountered node has a better chance of meeting the destination (either based on the probability predictability metric or the estimated node contact rate metric). Our schemes work better than the BubbleRap-U scheme because we make use of the delivery predictability metric to decide whether or not a node will forward a copy to an encountered node. The local rank values for many nodes in [2] are the same, and hence nodes within the same community end up not forwarding a copy to one another. In addition, our use of the estimated node contact rate in SNAMDV2 is more destination-aware and hence the two schemes achieve better delivery performance than the Bubble-Rap-U scheme. As for the total transmissions reported in Fig 3(c), it may seem like there is a contradiction that Bubble-Rap-U results in more transmissions. This is explained as follows: in many cases, the next hop forwarding node is the same and the Bubble-Rap-U approach uses as many transmissions as there are receivers for each multicast message while the multicast approach only transmits once. We also have results with random sources sending to either the first or the 2^{nd} set of five receivers. The same trend (i.e. SNAMDV2 scheme achieves the best delivery ratio at a lowest transmission cost) is observed. Due to space limitation, such results are not included here but interested readers can find them in [16].

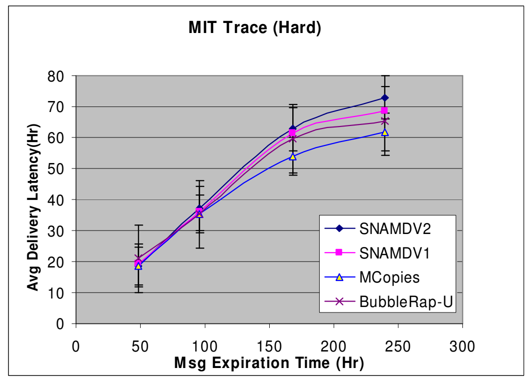
Figs 4(a)-(c) plot the results for the delivery ratio, the average message latency and the total cost respectively for the Infocom 2005 trace. We use the K-clique clustering information that the authors in [2] provided. The results show that our schemes achieve better delivery ratios than the BubbleRap-U scheme. The transmission cost for SNAMDV2 scheme is lower than BubbleRap-U but the transmission cost of the SNAMDV1scheme is still higher than BubbleRap-U scheme. However, all these 3 schemes cost less than the MCopies-M scheme.

6. Concluding Remarks

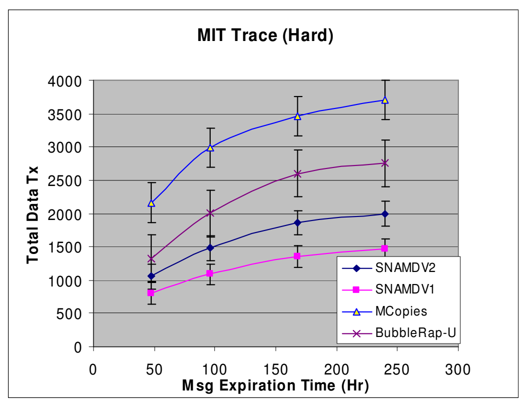
In this paper, we have designed new multicast delivery schemes that make use of the social network related information for human contact-based networks. Specifically, our schemes forward a copy of the message to connectors (nodes that can meet different communities) and let each node make use of either the delivery predictability or the estimated node contact rate in its forwarding decision. Simulation results show that our schemes can achieve better performance since (a) they are more destination aware, and (b) the connectors can meet many nodes as well as nodes from different communities. We hope to conduct more extensive evaluations using more scenarios and other traces e.g. NUS trace [19]. In addition, the connectors are identified via offline analysis of the traces. Identifying a proper subset of connectors for other traces is left for future work. In this work, we assume that the clustering information is static. In future, we intend to investigate the impact of dynamic clustering on the delivery performance of our SNAMD scheme.



(a) Delivery Ratio vs Msg Expiration Time



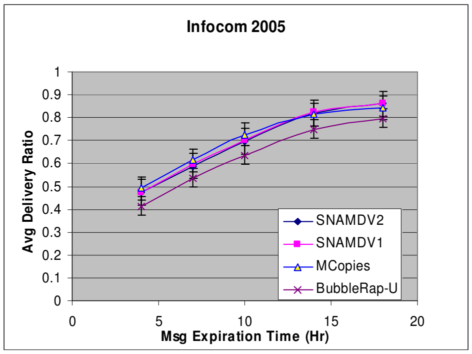
(b) Avg Msg Latency vs Msg Expiration Time



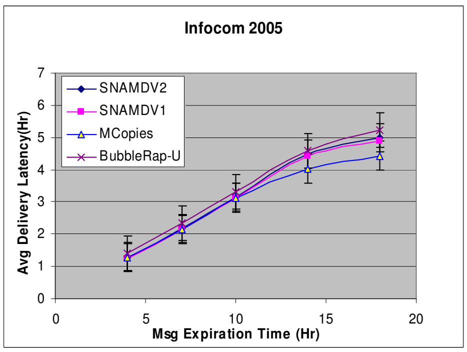
(c) Total Data Transmissions

Fig3: Delivery Perforamnce

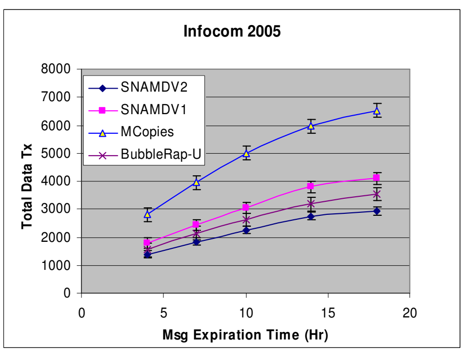
Scenario 2 (MIT Reality Trace)



(a) Delivery Ratio vs Msg Expiration Time



(b) Avg Msg Latency vs Msg Expiration Time



(c) Total Data Transmissions

Fig 4: Delivery Performance (Infocom 2005)

7. ACKNOWLEDGMENTS

The author thanks Dr. Pan Hui for sharing his Java simulator and the MIT Reality Mining and Haggle traces in a useful format. She also wishes to thank Nishanth Sastry and anonymous reviewers for their useful feedbacks on an earlier version of the paper.

8. REFERENCES

[1] A. Chaintreau et al, “Impact of Human Mobility on Opportunistic Forwarding Algorithms”, Proceedings of IEEE Infocom, April, 2006

[2] P. Hui, J. Crowcroft, E. Yoneki, “Bubble Rap: Social –Based Forwarding in Delay Tolerant Networks”, Proceedings of ACM Mobihoc, pp241-250, 2008

[3] M. Newan, “Detection Community Structure in Networks”, European Physic Journal B 38:321-330, 2004

[4] L. Danon et al, “Comparing Community Structure Identification”, J. Stat. Mech, 2005, P09008

[5] M. Gladwell, “The Tipping Point: How Little Things Can Make a Big Difference”, 2000

[6] U. Brandes, “A Faster Algorithm for Betweenness Centrality”, Journal of Mathematical Sociology, Vol 25, pp 163-177, 2001

[7] D. Hicks, M. Nasrullah, “Detecting Key Players in 11-M Terrorist Network: A Case Study”, 3rd IEEE National Conference on Availability, Reliability, and Security, pp 1254-1259, 2008

[8] P. Hui et al, “Distributed Community Detection in Delay Tolerant Networks”, Proceedings of ACM Sigcomm Workshop, MobiArch, 2007

[9] J. Baumes et al, “Dynamics of bridging and bonding in social groups, a multiagent model”, Proceedings of 3rd Conference of the North American Association for Computational Social and Orgnizational Science (NAACSOS 05), Notre-Dame, Indiana, June, 2005

[10] N. Eagle, A. Pentland, “Reality mining: sensing complex social systems”, Personal and Ubiquitous Computing, Vol 10(4):255-268, May, 2006

[11] M. Chuah, A. Coman, “Identifying Connectors and Communities: Understanding Their Impacts on the performance of a DTN publish/subscribe system”, CSE Technical Report, April, 2009

[12] P. Hui, A. Chaintreau, et al, “Pocket Switched Networks and Human Mobility in Conference Environments”, Proceedings of ACM WDTN, 2005

[13] UCINet:A Social Network Analysis tool www.analytictech.com/downloaduc6.htm

[14] M. Goldberg et al, “Communication Dynamics of Blog Networks”, The 2nd SNA-KDD Workshop, Aug, 2008

[15] Q. Ye et al, “Performance Comparison of different multicast routing strategies for disruption tolerant networks”, to appear at Computer Communications, 2009

[16] M. Chuah, “Social Network Aided Multicast Delivery Scheme for Human Contact Based Networks”, CSE Technical Report, Lehigh University, May, 2009

[17] W. Zhao, M. Ammar, E. Zegura, “Multicasting in Delay Tolerant Networks: Semantic Models and Routing Algorithms”, ACM Workshop in Challenged Networks, Aug, 2005

[18] A. Lingren et al, “Probabilistic Routing in Intermittently Connected Networks”, Proceedings of Workshop on Service Assurance with Partial and Intermittent Resources, Aug, 2004

[19] V. Srinivasan, M. Motani, W.T. Ooi, “Analysis and Implications of Student Contact Patterns Derived from Campus Schedules”, Proceedings of ACM Mobicom, 2006