Social-Aware Stateless Forwarding in Pocket Switched Networks

Alessandro Mei
Computer Science Department
Sapienza University, Rome, Italy

Giacomo Morabito
DIIT
University of Catania, Italy

Paolo Santi
IIT
CNR, Pisa, Italy

Julinda Stefa
Computer Science Department
Sapienza University, Rome, Italy

Abstract—In this paper we describe SANE, the first forwarding mechanism that combines the advantages of both *social-aware* and *stateless* approaches in pocket switched network routing. SANE is based on the observation—that we validate on real-world traces—that individuals with similar interests tend to meet more often. In our approach, individuals (network members) are characterized by their *interest profile*, a compact representation of their interests. Through extensive experiments, we show the superiority of social-aware, stateless forwarding over existing stateful, social-aware and stateless, social-oblivious forwarding. An important byproduct of our interest-based approach is that it easily enables innovative routing primitives, such as interest-casting. An interest-casting protocol is also described, and extensively evaluated through experiments based on both real-world and synthetic mobility traces.

I. INTRODUCTION

The vision of a near future in which a multitude of hand-held devices establish direct wireless communication links in an opportunistic fashion has recently attracted the attention of the research community. Indeed, powerful hand-held devices are becoming increasingly popular and that these devices are typically endowed with wireless technologies allowing direct communication between them (e.g., Bluetooth). The above vision has motivated researchers to focus on a specific type of delay tolerant network, called *opportunistic* or *pocket switched network* [8], in which nodes are *individuals* carrying hand-held devices and links appear and disappear as these individuals move and get in physical proximity. In particular, several authors have tried to use the social ties between individuals to optimize the performance of unicast communications [3], [6], [10], as well as multicasting [5], and publish-subscribe mechanisms [1], [2], [9].

Social-aware routing protocols have been shown to have superior performance to social-oblivious routing protocols such as, e.g., BinarySW [14]. This performance improvement comes at the expense of storing a significant amount of state information (e.g., history of past encounters, portion of the "social network" graph, etc.) at the local memory of the nodes. In other words, a common feature of the social-aware routing approaches introduced so far is that they heavily build upon a notion of *state*. Our goal in this paper is to design a

suite of *social-aware*, *stateless* routing protocols that combine the advantages of social-aware forwarding with the negligible extra storage requirements typical of stateless protocols.

We start by giving a model for representing user interests and their similarity. In our approach, the collection of interest are represented in an m-dimensional *interest space*, and the individuals of the network are characterized by their *interest profile*, an m-dimensional vector corresponding to a point in the interest space. The forwarding strategy is then driven by a measure of similarity between interest profiles that, at least indirectly, expresses strength of social ties between the corresponding individuals. Through extensive experiments based on both real-world and synthetic mobility traces, we show the superiority of our proposed social-aware, stateless routing approach over existing stateful, social-aware as well as stateless, social-oblivious routing approaches.

II. RELATED WORK AND CONTRIBUTION

The idea of exploiting information regarding social ties between network nodes in PSNs is not new. For instance, in [3] the authors use the notions of "ego-centric betweenness" and "social similarity" to improve end-to-end routing performance. In [6], the authors propose to use a social "centrality" metric to achieve the same purpose. In [10], the authors use a "social similarity" metric locally computed from the history of past encounters to route messages within the network. Recently, a social-based approach based on a notion of "ego-centric betwenness" has been proposed also to optimize multicast performance [5].

The above protocols have shown how the social structure of a PSN can be successfully exploited to improve traditional, social oblivious approaches. However, existing social-aware approaches heavily build upon the ability of storing a large amount of information at the nodes (typically, to keep trace of past encounters), i.e., their are *stateful* approaches. This fact has important implications for what concerns i) scalability and ii) effects of memory size on routing performance. As for i), we observe that relying on a rich state (in some cases, $O(n^2)$ storage capacity is required at the nodes, where n is the number of network nodes) might impose severe limits to the ability of these approaches to scale up to networks of even medium size. As for ii), to the best of our knowledge, the effect of limited memory size on social-aware routing performance has not been investigated so far.

III. Interest Space and Profiles

We assume each individual in the network can be represented through her *interest profile*, a compact representation of her interests within the *interest space*. We represent the interest space as an m-dimensional unit cube $\mathcal{C} = [0,1]^m$, where m is the total number of interests in the network. Interests are intended in a very broad sense, they might represent degree of interest in a certain topic (e.g., cinema, literature, etc.), the fact that an individual belongs to a certain physical or virtual community (e.g., living in a certain neighborhood, member of a Facebook interest group, etc.), and so on.

Given the above definition of interest space, it is natural to represent the interest profile of an individual A with an m-dimensional vector reporting, for each possible interest dimension, A's degree of interest in the particular topic/community (either a real number or a binary value). To express similarity between individual interests, and thus quantitatively measure "homophily"—degree of interest similarity [11]—we use the well-known *cosine similarity* metric [4]:

Definition 1: Given two m-dimensional vectors A and B, the cosine similarity metric, denoted $\Theta(A, B)$, is defined as follows:

$$\Theta(A, B) = \cos(\angle AB) = \frac{A \cdot B}{\parallel A \parallel \parallel B \parallel} ,$$

where $\parallel X \parallel$ represent the length of vector X.

Note that, given the definition of interest space, $0 \leq \Theta(A, B) \leq 1$ in our model, with higher values of $\Theta(A, B)$ corresponding to a higher "homophily" degree.

Our stateless protocols are based on a simple and natural observation from everyday life: Our movements are guided by our interests. To validate this intuition, we use traces collected during an experiment done with real Bluetooth communicating devices distributed to participants of the Infocom 2006 conference [6], [7]. This data trace contains not only contact logs, also does it report information on participants' nationality, residence, languages spoken, affiliation, scientific interests, etc.. From this information we generate interests profile as defined above. In the process, we discard participants that have not declared any interest, in order to remove erroneous profiles. This way, the number of participants reduces to 61 nodes.

To support our intuition, we first calculate the cosine similarity between the interest profiles for every pair of participants. Then, we compute the Pearson correlation index among this value and the total meeting duration/meeting frequency among every couple. We then compute the correlation coefficients among profile similarities and meeting duration/meeting frequency, only for pairs of individuals who spend, on the average, more than a certain amount of time together. This way the effect of casual short meetings is attenuated. The results are presented in Table I. As can be seen, when we focus on longer meetings, the correlation of meeting frequency and similarity of interest profiles is considerably high, reaching 0.67. These results support the conclusion that our intuition is sound and that it can be used as the basic mechanism of social-aware, stateless forwarding protocols.

AVG meet time	C_d	C_f	Nodes
> 0 (min)	.28	.08	61
> 5 (min)	.55	.57	53
> 10 (min)	.67	.67	26

TABLE I
CORRELATION BETWEEN INTERESTS PROFILES AND PARTICIPANTS' ENCOUNTERS. C_d AND C_f INDICATE THE PEARSON CORRELATION COEFFICIENT BETWEEN PARTICIPANTS' COUPLES PROFILES AND RESPECTIVELY TOTAL MEETING DURATIONS AND MEETING RATES.

IV. Social Aware NEtworking (SANE)

In this section, we describe *Social Aware NEtworking* (SANE), a protocol suite that enables the efficient delivery of information to relevant destinations in PSNs. SANE supports a novel communication service, that we call *interest-cast* (see Section IV-B), besides the traditional unicast.

We assume that each node can be a forwarder and therefore maintains a buffer of messages that must be relayed to the respective destinations. Each message M has a header that contains a target interest profile that we call *message relevance profile*, an integer value N_{replicas} representing the number of replicas of the message that the node is allowed to forward to other relays, and a *time-to-live* value TTL. In PSNs nodes can exchange information as a communication opportunity arises. Accordingly, SANE procedures are triggered each time a node, say A, enters within the radio coverage of another node, say B. Initially, nodes exchange their interest profile (IP), then each node start scanning its buffer for messages to relay.

A. Unicast

In the unicast case the message relevance profile is set equal to the interest profile of the destination. According to our interest-based approach, a message M should preferably be forwarded to individuals whose interest profile closely resembles the one of the destination. More specifically, as in [14], we assume that in order to keep the communication overhead under control, the same message can be relayed at most for N^*_{replicas} times. Message relaying obey the following rules: Message M should be relayed to a node B if an only if both the two following conditions hold:

- the current value of N_{replicas} is higher than 1.
- the cosine similarity metric between the relevance of message M, denoted as $R(M)$, and the IP of B, denoted $IP(B)$, is higher than a given threshold ρ that we call *relaying threshold*.

In case, the value of N_{replicas} in the message header is halved and a copy of the message is sent to B. Obviously the message is transmitted to node B regardless of the value of N_{replicas} if B the destination of the message. In this case, node A removes the message from the buffer after this is relayed to B.

Note that, as the threshold ρ decreases, the forwarding strategy becomes more aggressive. This results in the decrease of the delivery delay, and in the increase of both the delivery success probability and the communication overhead (cost) incurred for the delivery of the message M, that we denote

as $c(M)$. Observe that the cost $c(M)$ is proportional to the number of copies of the message M spread in the network. A few extreme cases can be considered:

- $N^*_{\text{replicas}} = \infty$: in this case there is no bound on the number of copies of the message circulating in the network. We call the resulting version of our protocol suite *epidemic* unicast SANE, and we denote it with UN-SANE EP. The SANE version corresponding to the case $N^*_{\text{replicas}} < \infty$ is instead called *spray & wait* unicast SANE and denoted UN-SANE SW.
- $\rho = 0$: in this case, the relay threshold is not used, and the proposed forwarding strategy becomes the same as BinarySW [14]. Furthermore, if N^*_{replicas} is set equal to ∞ then our protocol behaves like epidemic forwarding [15], which is the policy achieving the lowest delivery delay (but also the highest cost).
- $\rho = 1$: in this case, only direct message delivery from source to destination is possible: Message delivery cost is minimized, but message delivery delay is very high.

In Section V, we will show the impact of the threshold ρ on the performance of the forwarding strategy through numerical examples.

B. Interest-cast

PSNs can create innovative services. *Interest-cast* is an example of such services in which a user wants to communicate a certain information (for instance, a movie at a local theater about opera composer Puccini) to the maximum possible number of interested users, within a certain time (e.g., the time of the last movie show). Interested users might have an interest in opera, or cinema, or both, and may be located in the "neighborhood" of the theater, so to be able to reach the theater if interested. This type of communication paradigm matches very well with the localized nature of PSN communications: the information is spread relatively fast in the neighborhood of the sender, while it takes longer to propagate to remote areas (which are typically less interested in the information).

Assume individual C wants to send a message M to all the individuals within the network that are potentially interested. First, C sets $R(M)$, the message relevance profile of M, which is done by setting a "relevance" value in $[0, 1]$ for each of the m interest dimensions. Whether a message M is relevant for a certain individual B is determined using a certain *relevance metric*. As we already explained, in this paper we use the well-know cosine similarity metric [4] to determine whether message M is relevant for individual B.

Note that, since both individuals' interests and message relevance profiles take values in the same m-dimensional interest space, we have that, for any individual B and message M, the angle between $IP(B)$ and $R(M)$ is in $[0, \pi/2]$, implying that $\Theta(B, M)$ is indeed in $[0, 1]$. In this paper, we use the following simple rule to determine whether message M is relevant to individual B: The message is relevant if and only if $\Theta(IP(B), R(M)) \geq \alpha$, where α is a suitably chosen *relevance threshold*.

V. EXPERIMENTAL SETUP AND RESULTS

Here we present experimental results on the performance of SANE (the interest-cast version) and UN-SANE (the unicast version) compared to that of well known opportunistic forwarding protocols. For the evaluation we use both real-world traces (Infocom 06) and synthetic ones obtained with the SWIM mobility model [12]. We use also synthetic mobility traces to evaluate protocol performance because of the limited real-world traces enriched with user profiles, which does not allow evaluating performance under different conditions for what concern, e.g., the degree of correlation between individual meeting rates and similarity of their profiles. Such different mobility scenarios can instead be easily realized in SWIM by properly tuning the model parameters.

A. SANE with limited buffer in Infocom 06

To validate the protocols on the Infocom 06 trace we average the results of the following experiment, repeated 100 times: We generate a message with a uniform traffic pattern (source-destination chosen uniformly at random), and set message's relevance profile to the destination's interest profile. Then, we let the message to be forwarded in the network according to the different forwarding schemes. We have evaluated all the protocols with limited buffer size of the nodes, for different limits. Here, for the sake of space, we present only the results where the limit is 40 packets per node.

As already discussed, the correlation between node interest profiles and their meeting frequencies may be low (see first row of Table I) without filtering out short meetings; on the other hand, filtering out short meetings to increase correlation would considerably reduce the size of the data set, making simulation results scarcely significant. In view of this, we have decided to keep the user population as large as possible (61 users, with a 0.08 meeting frequency correlation); thus, reasonably low values for the relay and relevance thresholds ρ and α should be chosen ($\rho = .25$ and $\alpha = .45$ in our case) for both unicast and inter-cast.

1) Unicast: We compare the unicast version of SANE (UN-SANE) to well known stateless forwarding protocols such as BinarySW [14] and Epidemic [15], and to a state-of-the-art of social-aware forwarding protocol, BUBBLE [6]. In implementing BUBBLE, we took care of putting the protocol in the best possible conditions, i.e., complete knowledge of the social graph and of the local/global ranking metrics. We consider both the SW and the uncontrolled version of UN-SANE in our experiments, denoted UN-SANE SW and UN-SANE EP, respectively. Being the network considered of only 61 nodes, parameter N^*_{replicas} (number of message copies) of BinarySW and UN-SANE SW is set to 4. The experiments are repeated for various values of the TTL's, and in each case we measure the *average delay* (average delivery time for successfully delivered messages), the *cost* (average number of message copies in the network per delivered message, computed only for successfully delivered messages), and success percentage.

As you can see in Figure 1, both versions of UN-SANE provide significantly higher success percentage than that of

all the competing protocols, at a lower cost and at a similar or smaller delay. Quite interestingly, buffer limit yields the surprising result that UN-SANE EP has higher success rate than Epidemic itself.

2) Interest-cast: Here, we show results related to the two interest-cast versions of our protocol: SANE SW, and SANE EP. Since there is no immediate way of extending BUBBLE into an interest-cast protocol, we compare SANE protocols only to Epidemic and BinarySW, whose interest-cast versions are straightforward (simply delivers a copy of the message to all relevant destinations). The way we generate messages and the input tuning parameters of BinarySW and SANE SW are the same as in unicast. The results are shown in Figure 2. In this case, *coverage* refers to the percentage of relevant destinations holding a copy of the message when the TTL expires. As seen from the figures, SANE protocols perform very well, providing comparable coverage of relevant destinations to that of Epidemic (for TTLs values large than 30 min), but with a much reduced cost (as much as 10-fold cost reduction with respect to Epidemic, in case of SANE SW). The benefits of social-aware forwarding are evident comparing the relative performance of BinarySW and SANE SW: with a comparable cost, SANE SW provides higher coverage and lower delay than BinarySW.

B. SANE with Synthetic Traces

The synthetic traces we use for evaluation have been obtained from the SWIM mobility model [12], [13]. In SWIM, nodes are assigned a home point in the network area, assumed to be a square. Each time a node has to choose its next destination, it tradeoffs distance from its home point and popularity of the possible destinations. Thus, nodes with relatively close home points (*neighbors*) tend to go to the same locations and get in contact more often. In order to run SANE on SWIM's traces, we do the following setup: First, we generate a the network, and a given number of network nodes. For each node, a 4-dimensional interest profile vector is randomly generated, with entries chosen independently and uniformly at random in $[0, 1]$. Each profile vector is then normalized to 1—this way, we make sure that no node has very low interests or no interests at all.

In SWIM, neighbors tend to have a higher meeting rate. The amount of correlation between vicinity of home points and meeting rate in SWIM is controlled by a parameter that here we denote as η (see [12], [13] for details): The higher this parameter, the higher the correlation. Therefore, we can generate SWIM mobility traces controlling the resulting correlation between node profile similarity and their meeting frequency by tuning SWIM's η parameter. Due to space limitation, in the following we will only show results for a SWIM simulation with 200 nodes scattered in a square area of $500m \times 500m$ and with η set in such a way that the correlation between interest profile similarity and pairwise meeting rates is about .7.

Unfortunately, due to lack of space, here we do not present SWIM-based comparison results of SANE with the afore-

mentioned well-known forwarding based protocols. Still we want to stress that due to the high correlation between node-profiles and pairwise meeting rates the advantage of the SANE protocols over the competitors becomes even more evident than in Infocom 06 simulations. Figures 3 and 4 we show the success rate, average cost and average delay per received copy when the relevance threshold is $\alpha = .95$, versus the value of the relaying threshold ρ. As expected, the communication cost increases as the value of ρ decreases.

VI. CONCLUSIONS

In this paper, we have first validated the intuition that individuals with similar interests tend to meet more often than individuals with diverse interests, and then used this intuition to design the first social-aware, stateless forwarding mechanism for opportunistic networks, called SANE. A nice feature of the SANE forwarding approach is that it can be used not only for traditional unicast communication, but also for realizing innovative networking services for PSNs, such as interest-casting. When collectively considered, the experimental results clearly show the superiority of SANE protocols over both social oblivious, stateless and social-aware, stateful approaches. Quite astonishingly, SANE provides better performance than competitors even when the degree of correlation between interest profile similarity and pairwise meeting rates is modest, as in the Infocom 06 scenario. If this correlation is higher, as it might be expected in practical situations, we expect that the advantages of SANE protocols over competitors become substantial.

REFERENCES

[1] C. Boldrini, M. Conti, A. Passarella, "ContentPlace: Social-Aware Data Dissemination in Opportunistic Networks", ACM MSWiM 2008.

[2] P. Costa, C. Mascolo, M. Musolesi, G.P. Picco, "Socially-Aware Routing for Publish-Subscribe in Delay-Tolerant Mobile Ad Hoc Networks", *IEEE JSAC*, Vol. 26(5), 2008.

[3] E. Daly, M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs", ACM MobiHoc 2007.

[4] M.M. Deza, E. Deza, *Encyclopedia of Distances*, Springer, Berlin, 2009.

[5] W. Gao, Q. Li, B. Zhao, G. Cao, "Multicasting in Delay Tolerant Networks: A Social Network Perspective", ACM MobiHoc 2009.

[6] P. Hui, J. Crowcroft, E. Yoneki, "BUBBLE Rap: Social-based Forwarding in Delay Tolerant Networks", ACM MobiHoc 2008.

[7] P. Hui, E. Yoneki, S-Y. Chan, J. Crowcroft, "Distributed Community Detection in Delay Tolerant Networks", ACM MobiArch 2007.

[8] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, "Pocket-Switched Networks and Human Mobility in Conference Environments", ACM WDTN 2005.

[9] S. Ioannidis, A. Chaintreau, L. Massoulie, "Optimal and Scalable Distribution of Content Updates over a Mobile Social Networks", IEEE Infocom 2009.

[10] F. Li, J. Wu, "LocalCom: A Community-Based Epidemic Forwarding Scheme in Disruption-tolerant Networks", IEEE Secon 2009.

[11] M. McPherson, "Birds of a feather: Homophily in Social Networks", *Annual Review of Sociology*, vol. 27, n. 1, pp. 415–444, 2001.

[12] A. Mei, J. Stefa, "SWIM: A Simple Model to Generate Small Mobile Worlds", IEEE Infocom 2009.

[13] S. Kosta, A. Mei, J. Stefa, "Small World in Motion (SWIM): Modeling Communities in Ad-Hoc Mobile Networking", IEEE Secon 2010.

[14] T. Spyropoulos, K. Psounis, C. S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multi-copy Case", IEEE Trans. on Networking, vol. 16(1), 2008.

[15] A. Vahdat, D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks", *TR CS-200006*, Duke Univ., 2000.
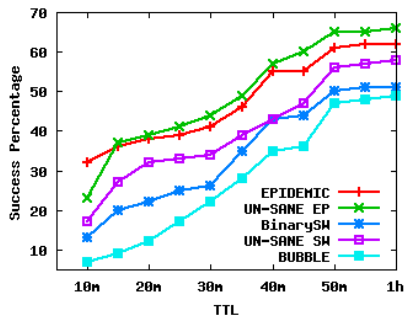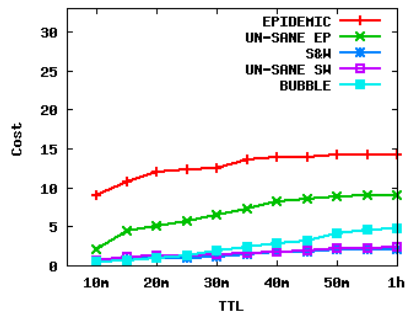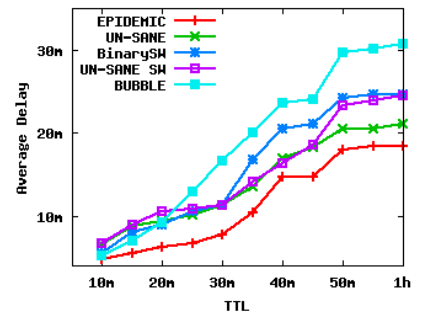
Fig. 1. Performance of unicast protocols on Infocom 06 traces. Buffer limit set to 40 messages.
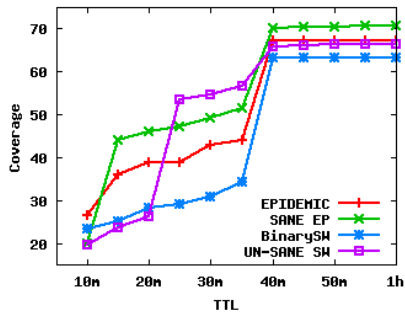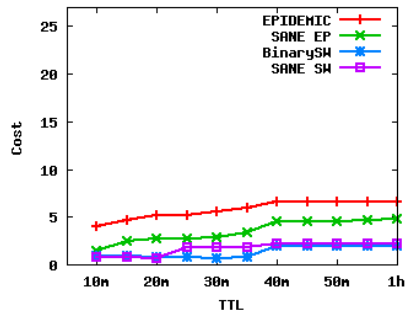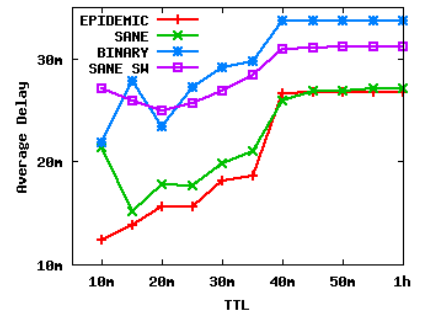
(a) Success Percentage

(b) Average Cost

(c) Average Delay

Fig. 2. Performance of multicast protocols on Infocom 06 traces. Buffer limit set to 40 messages.
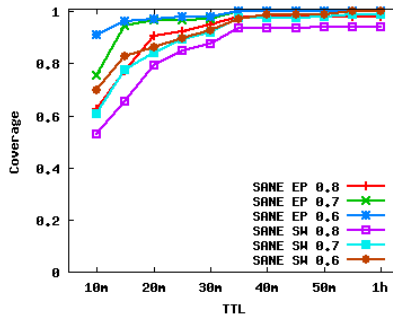
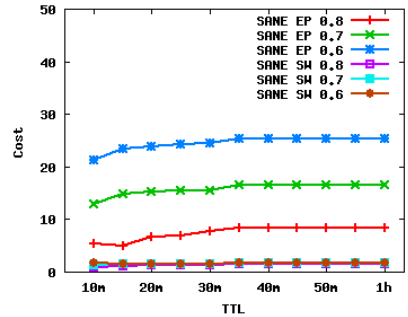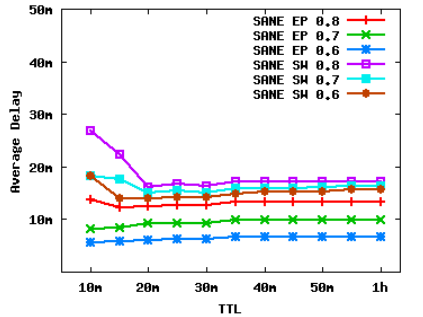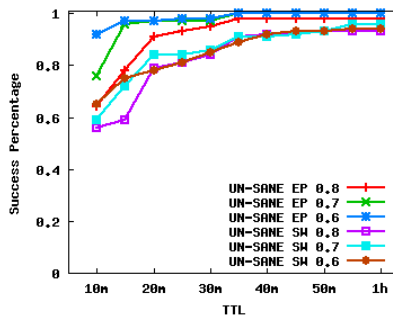(a) Success Percentage

(b) Average Cost
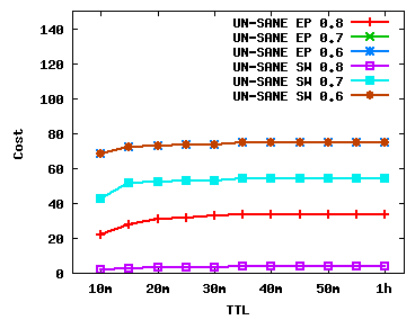
(c) Average Delay
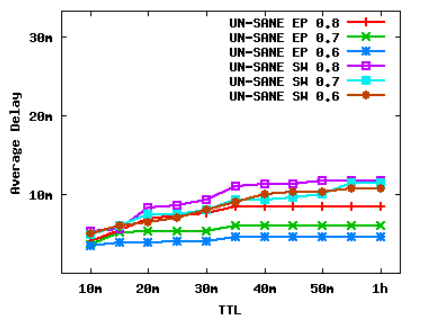

(a) Coverage

(b) Average Cost

(c) Average Delay

Fig. 3. SANE in dependence of the relay threshold ρ.

(a) Succes Percentage

(b) Average Cost

(c) Average Delay

Fig. 4. UN-SANE in dependence of the relay threshold ρ.