Show me your friends and I'll tell you what you like

Alessandro Mei*, Giacomo Morabito†, Paolo Santi‡ and Julinda Stefa*
*Department of Computer Science, Sapienza University of Rome, Italy
Email: {mei,stefa}@di.uniroma1.it.
†Dipartimento di Ingegneria Informatica e delle Telecomunicazioni,
University of Catania, Italy
Email: giacomo.morabito@diit.unict.it.
‡ Istituto di Informatica e Telematica del CNR Pisa, Italy
Email: paolo.santi@iit.cnr.it.

***Abstract*—Pocket Switched Networks (PSN), where humans carry short-range communicating devices such as PDAs, lap-tops, or smart phones, have attracted the attention of many researchers in recent years. These networks, whose main feature is the social-guided movements of users/devices, are considered to be a key technology to provide innovative services to the users without the need of any fixed infrastructure. Even though many forwarding protocols (statefull social-aware–e.g. PROPHET, BUBBLE–and stateless ones–e.g. Epidemic, BinarySW) have been proposed, there's still a need for efficient killer application, possibly both stateless and social-aware, that would transform the PSNs in a useful living reality.**

Here we discuss on the advantages/disadvantages/scalability of already existing stateful social-aware and simple stateless forwarding protocols for PSNs and introduce a stateless interest-based forwarding mechanism that combines the advantages of both forwarding approaches.

I. Introduction

The Pocket Switched Network (PSN) [13], where powerful, short-range communicating devices such as PDAs, smart-phones, etc., are carried around by humans, has attracted more and more researchers recently. In such networks the social-based node mobility coupled with a *store-and-forward* mechanism is the fundamental mean of communication among users. Thus, characterization of social ties between nodes has been used to optimize performance of both unicast and multicast communication [5], [11], [16], [8], as far as publish-subscribe mechanisms [1], [4], [14].

While social-aware routing protocols have been shown to have superior performance to social-oblivious routing ones such as, e.g., BinarySW [19], 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*. Given that existing routing approaches for PSNs have both pros and cons, it would be interesting to design a routing approach aimed at combining the advantages of both approaches, while reducing the cons as much as possible.

In order to exploit both approaches benefits, while keeping out their drawbacks, we base our work on the observation, qualitatively well-known in sociology [17], that individuals with similar interests tend to meet more often. Indeed, a first significant contribution of this paper is a *quantitative* validation of this observation, based on the only real-world mobility trace enriched with user profiles information we are aware of [11], [12]. Upon such observation we build the first *stateless* and *social-aware* routing protocol for PSNs called UN-SANE (UNicast Social Aware NEtworking).

Moreover, we believe that PSNs can create innovative services realized within the PSN itself, without the need of resorting to pre-existing communication facilities. *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 should 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, though). Thus, along with UN-SANE, in this paper we also present the first *social-aware* and *stateless* interest-cast protocol for Pocket Switched Networks called SANE (Social Aware NEtworking). We believe such interest-cast primitive, convenient in many real-world scenarios, to be one of the killer apps that will transform the Pocket Switched Network in a useful reality.

The rest of the paper is organized as follows: Section II reports on on current work in the field. In Sections III and IV we respectively validate our intuition on the real-world data trace, and present our unicast and interest-cast approaches to forwarding in Pocket Switched Networks. Section V presents experimental results of the comparison of our SANE approaches to existing well-known forwarding protocols such as Epidemic [20], BinarySW [19], and BUBBLE [11]. We finally conclude with section VI.

II. Related work

The idea of exploiting information regarding social ties between network nodes in PSNs is not new. For instance, in [5] the authors propose using the notions of "ego-centric betweenness" and "social similarity" to improve end-to-end routing performance. In [11], the authors propose the use of a social "centrality" metric to achieve the same purpose. In [16], 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 [8].

The above approaches have shown how the social structure underlying a PSN can be successfully exploited to improve communication performance with respect to 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. Considering limited memory size when comparing performance of stateful approaches (such as, e.g., [5], [11], [16]) to that of stateless, social-oblivious approaches such as Epidemic [20] and Binary SW [19] is very important, since using memory to store state information clearly reduces the amount of memory that can be used to store the messages circulating in the network, with a negative effect on routing performance. Hence, comparing routing performance without taking the effect of limited memory size into account gives an unfair advantage to stateful approaches over stateless one. Given $i)$ and $ii)$, whether social-aware approaches are actually effective in improving routing performance is still not clear, as well as their scalability properties.

The line of research closer to the ideas presented in this paper is the design of social-based publish-subscribe mechanisms for PSNs. For instance, the authors of [14] consider a network in which a service provider (e.g., a cell phone operator) selectively sends dynamic content updates to users, updates that can be shared with other users when a communication opportunity arises. The authors show that performance can be considerably improved if the relative frequency of updates sent to users from the service provider takes into account the strength of social ties between network users. In [1], the authors approach the problem of sharing data within an opportunistic network from a social-based perspective, with the goal of optimizing content availability through careful, social-aware data placement.

The work that is most closely related to ours is [4], in which the authors present the design of a mechanism whose underlying routing framework, called SocialCast, exploits predictions based on metrics of social interactions to drive the forwarding process. While the underlying idea of interest-based routing is similar in spirit to our approach, implicit in SocialCast is the assumption that an individual implicitly or explicitly subscribes to one or more "interests". On the contrary, our approach builds on the notion of interest profile, in which a PSN member compactly encodes not only the *degree* (not necessarily binary) of interest in different topics, but also his/her habits (e.g., where he/she lives, works, etc.), etc. Thus, our approach allows a more complete characterization of a PSN member's habits and social relationships. Finally, SocialCast still remains a publish-subscribe scheme and requires storing of a considerable amount of state information at the nodes, which should be contrasted with the stateless approach taken herein.

III. Interest-similarity among friends in real life

In our everyday life our movements are guided in a large part by our interests. As a matter of fact, we go to cinemas because we are interested in movies, we go to the library because we are interested in reading, and so on. As a consequence, we often meet people that have similar interests, hobbies, and habits to ours. To validate this intuition in a quantitative fashion, we use traces built during an experiment done with real Bluetooth communicating devices distributed to part of the participants of the Infocom 2006 conference [11], [12]. The devices were configured to perform a Bluetooth baseband layer "inquiry" discovering the MAC addresses of other blue-tooth nodes in range of communication. The results of the inquiry were written to flash RAM, recording contact periods between devices, in the form of {MAC, start time, end time}. Along with this information, the data trace includes also reports on participants' nationality, residence, languages spoken, affiliation, scientific interests, etc.; from which we can easily generate an interests profile vector of 0/1 coordinates: We count all the possible nationalities, countries and cities of residence, languages spoken, affiliations, possible scientific interest topics, declared by the participants. Then, we build, for each participant, a profile vector that has as many coordinates as the sum of all these possibilities put together. A 1 in the i-th coordinate of a given participant's profile vector corresponds to the fact that that participant is either interested in the scientific topic, or speaks that particular language, or comes from that particular country (depending on what interest dimension i represents). In the process, we discard participants that have not declared any of the above interests, in order to remove erroneous profiles. The number of the participants involved after this cut reduces to 61. Although there are other data-traces available on line describing contact among participants in different experimental settings ([7], [11], [12], [13]), they do not include any information on participants' profiles. To

Experimental data set	Infocom 06
Device	iMote (Bluetooth)
Duration (days)	3
Granularity (sec)	120
Participants number	78
Participants with profile	61
Internal contacts number	191,336
Average Contacts/pair/day	6.7

TABLE I
DETAILED INFORMATION ON THE INFOCOM 06 TRACE.

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

TABLE II
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.

the best of our knowledge, Infocom 06 is the only available data-trace that includes also these type of information, thus in this paper we focus on this data-trace. More details about the data-trace can be found in Table I.

To support our intuition, we first measure the profile-similarity among node-couples. For this we use the well known cosine similarity metric [6], which measures similarity between two points A and B in a certain vectorial space as the cosine of the angle $\angle AB$ between the vectors corresponding to A and B. Formally:

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.

Then, we compute the Pearson correlation among this value and the total meeting duration/meeting frequency among every couple. These values result to be .28 and .08, respectively. The second correlation coefficient is small: This is more than reasonable, being this trace the result of the mobility pattern in a big conference, where there is a high "mixing" of people and thus a high number of short-casual meetings, for example, almost all the attendees meet during the coffee break. Yet, the first correlation coefficient (the one related to the duration of the contacts between people) shows that even in the presence of a high number of casual meetings, people with similar profile tend to meet for longer times. To confirm this observation, 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 the casual short meetings is attenuated and, at the same time, the number of participants that satisfy this condition decreases significantly. The results are presented in Table II. As can be seen, the higher the average meeting time used to filter out casual meetings, the higher the correlation among interest profiles and meeting duration/frequency. 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.

IV. SOCIAL AWARE AND STATELESS NETWORKING

In order to at least partially address the issues with current social-aware routing approaches, in this paper we advocate a different perspective on how information related to the user social behavior is used to optimize PSN routing performance. In particular, we propose to characterize each individual belonging to the network with an *interest profile* belonging to the network's *interest space*, and to base the forwarding strategy of the routing protocol upon a similarity metric between individual interest profiles. When individual A carrying a message M destined to individual D meets another individual B, he/she compare D and B interest profiles, and, based on the outcome of this comparison, he/she decides whether to forward M to B. It is important to observe that this forwarding approach is *stateless*, since A discards B's profile after the forwarding decision has been taken. Furthermore, forwarding decision is based on a notion of similarity between individual interests, somewhat taking the social ties between individuals into account. Thus, ours is, to the best of our knowledge, the first social-aware, stateless routing approach for PSN introduced so far.

Each message M's header 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 utilized to remove obsolete messages. Furthermore, the header of unicast messages contains the destination user identifier, whereas, the header of interest-cast messages contains a threshold value α that is used to select the relevant destinations as explained in Section IV-B. The treatment of each message depends on its type, (i.e., unicast or interest-cast), and will be described respectively in Sections IV-A and IV-B. After all messages in the buffer have been analyzed, the node updates the buffer. This is achieved by

- *removing messages that are obsolete*: To this aim a deadline instant, t_{dead}, is assigned to each message in the buffer.
- *handling the messages relayed by the other node*: More specifically, if the node is a destination then the message will be forwarded to the application; if the node is a relay then it will insert the message in the buffer. As described above, a deadline instant t_{dead} is assigned to the message which is calculated as the value of the current time plus the TTL value reported in the message header.

A. Unicast

In the unicast case we aim at the best tradeoff between communication overhead and the probability of delivery success (i.e., the probability that the packet reaches the destination before it elapses), as well as the delivery delay. 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 [19], 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*, that is

$$\Theta(R(M), IP(B)) \geq \rho \tag{1}$$

The values of N_{replicas} and TTL contained in the message header are updated as follows: The value of N_{replicas} is halved, whereas the value of TTL is set equal to the difference between the deadline instant and the current time. Then, a copy of the message is sent to B. Note that, since N_{replicas} is equal to half the initial number of replicas at the sender node A, this is equivalent to handling node B half of the copies of M currently in node A's buffer, as done in BinarySW [19]. 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 will remove the message from the buffer after this is relayed to B.

The source is responsible of initializing the values of N_{replicas}, which must be a power of 2 and represents the maximum values of replicas of the message in the network, and the value of TTL, which represents the maximum delay acceptable for the delivery of the message. The message relevance profile is set equal to the interest profile of the destination.

Note that, as the threshold ρ decreases, the forwarding strategy becomes more aggressive. This results in the decrease of the delivery delay, and an 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. Note that 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* SANE, and we denote it with SANE EP. The SANE version corresponding to the case $N^*_{\text{replicas}} < \infty$ is instead called *spray & wait* SANE and denoted SANE SW.
- $\rho = 0$: in this case, the relay threshold is not used, and the proposed forwarding strategy becomes the same as BinarySW [19]. Furthermore, if N^*_{replicas} is set equal to ∞ then our protocol behaves like epidemic forwarding [20], 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.

B. Interest-cast

Assume individual C wants to send a message M to all or the largest possible number of potentially interested individuals within the network. First, C must set the message relevance profile of M, which can be done assigning for each of the m interest dimensions a "relevance" value in the $[0, 1]$ interval. Such m-dimensional vector associated with a message is used (coupled with the individuals' interest profiles) to drive information propagation within the PSN. Note that the notion of message relevance profile allows to represent message M–similarly to individuals–as a point in the interest space. In the following, the relevance profile of message M is denoted $R(M)$. The set of *relevant destinations* for M, denoted $RD(M)$, is the set of individuals within the PSN for which message M is deemed relevant. As a consequence, $RD(M)$ is the set of nodes to which message M should be delivered, subject to an upper bound on the delivery time that we have called TTL^*. 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 [6] 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*.

We want to stress the difference between the notion of interest-casting defined herein and more traditional communication paradigms and services such as multi-casting and publish-subscribe. In interest-casting, the only action taken by a "content provider" (an individual generating a message) is determining the message relevance profile. After that, the message is injected in the network, and information propagation is driven by the notions of relevance and interest profile. As we shall see, these notions are used not only to dynamically determine the set of relevant destinations, but also to govern the forwarding process. Thus, in interest-casting the content-provider is not aware of the set of destinations the content should be delivered to, which is in sharp contrast with the traditional notion of multi-casting in which multicast groups are explicitly defined and typically known to the content provider. Furthermore, in interest-cast destinations must not explicitly subscribe to a specific "topic", as an individual is able to dynamically "capture" all (or most) relevant messages circulating in the PSN. This is also in sharp contrast with publish-subscribe mechanisms, which typically requires explicit subscription to one or more "topics" to be able to receive relevant information.

The forwarding discipline of interest-cast is similar in philosophy to the unicast case. In fact, if the two conditions given in Section IV-A for the unicast case hold then the message is relayed to B in the same way. If the above two conditions are not met but B is a relevant destination, then the message is transmitted with N_{replicas} set to one and TTL evaluated as explained in Section IV-A. Note that the above transmission does not have impact on the communication overhead.

V. Experimental Setup and Results

In this section, we present some experimental results in order to show the performance of SANE, in both its unicast and interest-cast version, as compared to that of well known opportunistic routing protocols. Performance will be evaluated implementing the protocols in a trace-driven simulator feed by the Infocom 06 trace.

To this end, we average the results of the following experiment, repeated 100 times: We generate a message with a source and destination chosen uniformly at random (uniform traffic pattern), and we set message's relevance profile to be equal to the destination's interest profile. Then, we let the message to be forwarded in the network according to the different forwarding schemes. As already discussed in Section III, the correlation between node interest profiles and their meeting frequencies is low (see first row of Table II) 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); consequently, reasonably low values for the relay and relevance thresholds ρ and α should be chosen. In this experiment, we have set them to $\rho = .25$ and, for the interest case, $\alpha = .45$, respectively.

A. Unicast

In a first set of experiments, we compare the unicast version of SANE (UN-SANE) to that of well known stateless forwarding protocols such as BinarySW [19] and Epidemic [20], as well as with a state-of-the-art of social-aware forwarding protocol, namely BUBBLE [11]. 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. The results are presented in Figure 1.

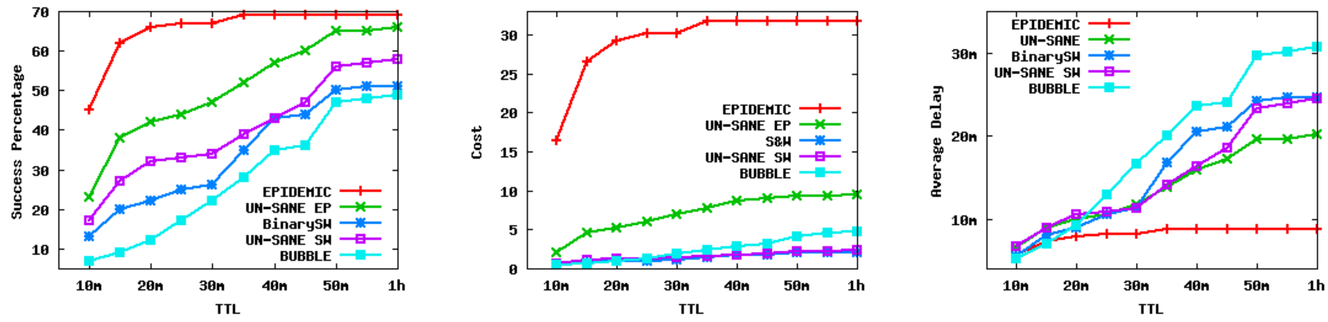
As can be seen, both versions of UN-SANE provide significantly higher success percentage than that of competing protocols (excluding, of course, Epidemic); also, the delay provided by the two versions of UN-SANE is better than that of both BinarySW and BUBBLE. In a sense, the two versions of UN-SANE provide different routing performance/cost trade-offs, with the SW version providing reduced success percentage with respect to the EP version (around 60% instead of about 68%), but with a much lower cost (factor 4 reduction in cost with respect to UN-SANE). Note also that the cost of UN-SANE SW is about the same as that of BinarySW, and lower than that of BUBBLE.

B. 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 the previous section. 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 as compared to BinarySW.

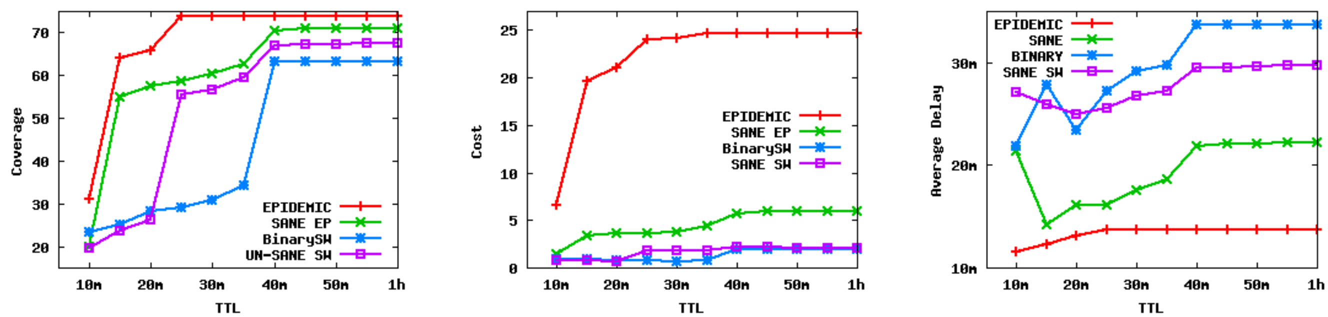
VI. Conclusions

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. The results of simulations based on real-world mobility traces have shown a clear superiority of our SANE approach over existing competitors.



(a) Success Percentage (b) Average Cost (c) Average Delay

Fig. 1. Performance of unicast protocols on Infocom 06 traces.



(a) Coverage (b) Average Cost (c) Average Delay

Fig. 2. Performance of multicast protocols on Infocom 06 traces.

References

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

[2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, "Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms", *Proc. IEEE Infocom*, 2006.

[3] M. Claypool, D. Brown, P. Le, M. Waseda, "Inferring User Interest", *IEEE Internet Computing*, Vol. 5, No. 6, pp. 32–39. November/December 2001.

[4] P. Costa, C. Mascolo, M. Musolesi, G.P. Picco, "Socially-Aware Routing for Publish-Subscribe in Delay-Tolerant Mobile Ad Hoc Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 26, n. 5, pp. 748–760, May 2008.

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

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

[7] N. Eagle, A. Pentland, "Reality Mining: Sensing Complex Social Systems", *Personal and Ubiquitous Computing*, Vol. 10, n. 4, pp. 255–268, 2006.

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

[9] M. Grossglauser and D. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks", *IEEE/ACM Transactions on Networking*, Vol 10, No 4, pp. 477–486, August 2002.

[10] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network", *Proc. MobiCom*, 2004.

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

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

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

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

[15] L. Kissner and D. Song, "Privacy Preserving Set Operations", CRYPTO, 2005.

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

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

[18] A. Noulas, M. Musolesi, M. Pontil, C. Mascolo, "Inferring Interests from Mobility and Social Interactions, *Proc. Workshop on Analyzing Networks and Learning with Graphs*, 2009.

[19] T. Spyropoulos, K. Psounis, C.S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multi-copy Case", *IEEE Trans. on Networking*, Vol. 16, n. 1, pp. 77–90, 2008.

[20] A. Vahdat, D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks", *Tech. Rep. CS-200006*, Duke Univ., 2000.

[21] W. Zhao, M. Ammar, E. Zegura, "Multicasting in Delay Tolerant Networks: Semantic Models and Routing Algorithms", *Proc. ACM WDTN*, 2005.