UNIVERSITÀ DEGLI STUDI DI MILANO
Dipartimento di Informatica e Comunicazione

RAPPORTO INTERNO Nº RT 30-09

Contact Analysis in Workplaces

Sabrina Gaito, Elena Pagani,
Gian Paolo Rossi

Contact Analysis in Workplaces

Sabrina Gaito, Elena Pagani, Gian Paolo Rossi
Computer Science Dept., Università degli Studi di Milano
e-mail: gaito@dsi.unimi.it, pagani@dico.unimi.it, rossi@dico.unimi.it

Technical Report 30-09

Abstract

So far, the search for Opportunistic Network (ON) applications has focused on urban/rural scenarios where the combined use of mobility and the *store-carry-and-forward* paradigm helpfully recovers from network partitions and copes with node sparsity. This paper explores the chance of using ONs in workplaces, where the node distribution is denser, thus contributing to reduce the message delivery latency, and where we still find similar needs for informal and unplanned network platforms to support human social relationships and interactions. Both a survey and trace recording experiments have been used to support the analysis of this mobility setting. The ability of recording very short contact times (i.e. lasting few seconds) allowed to interestingly show the slightly different role the social relationships play in dense scenarios and how the large amount of contacts (both short and long), occurring in densily populated spaces, actually contribute to reduce the message-delivery latency and to increase the delivery probability.

1 Introduction

Opportunistic Networks (ONs) have recently received growing attention because of their ability to create unplanned and improvised urban/rural wireless connectivity among mobile nodes. Researchers in this area envision an urban scenario where people carry radio devices that can be dynamically networked by exploiting human contact opportunities and, as a result, the term of pocket switched networks (PSN) has been used [8]. The growing interest in ONs is motivated by three main factors: 1. the pervasive multitude of portable devices has a huge amount of unused system and networking resources that may be profitably exploited to support a wide range of human interactions; 2. mobility can increase the capacity of wireless networks, as has been proved in [6]; 3. human social relationships provide quite a stable network of contact opportunities that can be profitably utilized to design forwarding algorithms [9].

So far, the search for ON applications envisions scenarios where the combined use of mobility and the *store-carry-and-forward* paradigm helpfully recovers from network partitions and copes with node sparsity. This paper focuses on a changed and enlarged application perspective. In fact, when we scale down from the metropolitan area to workplaces, buildings and small campuses, we obtain denser node distribution and observe similar needs for maintaining human social (or working) relationships that are likely to be profitably supported by an ON infrastructure, outside the institutional IT platform. The growing success of applications such as Twitter [13] – which keeps friends and co-workers frequently connected – and the constant growth of demand for mobile messaging (MM) applications[1] are showing this trend in human interaction and mandates verifying if ONs are suitable to support them. Apparently, this challenging idea is in contrast with the delay tolerant nature of ONs. There is the expectation, however, that the larger amount of contacts in a densily populated space can significantly contribute to reducing message-delivery latency. This motivates a research effort to understand the role of contacts and whether or not human mobility in workplaces shows behavior similar to mobility in more sparsely distributed settings.

When considering the application of PSNs in dense urban scenarios, people (and referees, as well) tend to ask why we should introduce a novel, delay-prone network when we have plenty of delay-free infrastructures. We too have wondered, without obtaining a definitive answer. The willing reader can find a more detailed analysis of this issue in [4], although we believe that we need not always have the applications before the technology. By contrast, we expect that, once deployed, the informal, easy, spontaneous access to this further form of connectivity will encourage the growth of a new family of applications centred on the innovative notion of 'on-line social network'.

To improve our understanding of human social attitudes in a workplace scenario, we performed a survey involving nearly 300 computer science faculty members and students. The primary observations obtained are the following:
(*i*) among faculty members the need for ubiquitous reception of notices about upcoming institutional meetings or events and for extemporaneous contacts with co-workers emerges. Students also need notifying and would like to be able to exchange messages with friends. Less than 30% of these communications are likely to have some attached file, so most of them could be fruitfully forwarded during brief contacts. In fact, the people surveyed seem to prefer downloading files over regular Internet when required (i.e. they want to be notified via thin device of attachments to download later);
(*ii*) 80% of people surveyed usually see each other at least once a day;
(*iii*) how much delivery latency people can tolerate varies according to service required, ranging from less than one to a few hours.

In the scenario outlined, this paper is a first attempt to answer questions such as the following: 1. Do social relationships in workplaces lead to the same structure of the contact topology as in the scenarios that have been considered so far? 2. has this new setting some different impact on the forwarding algorithms

[1] A recent Report of Forrester Research claims that the use of MM has grown of the 9% in 2008 and that its demand is expected to grow up to 80 million users in 2013, when it is supposed to replace 13% of SMS traffic.

in ONs? 3. can the application needs, as they emerged from the survey, be satisfied by the changed mobility setting and social attitudes?

To answer these questions, we developed a test bed with on purpose designed devices, named Pocket Mobility Trace Recorders, or PMTRs. The main and original behaviour of PMTRs is the ability to observe fine-grained contacts, so that even contacts of few seconds can be recorded. Moreover, unlike other experiences reported in the literature, the trial involved people profiled in order to enable the verification of the results emerging from trace analysis. Both the above aspects distinguish our experiments from other experiments conducted in restricted areas, e.g. [2].

The trace analysis enables to characterize the considered setting in terms of contact times distribution, latency times, social attitudes and their impact on the forwarding algorithms. The traces we obtained in a dense setting are compared with the traces obtained in similar experiments, where, however, it was not possible to observe short contact times [11]. We can identify three main contributions of this paper: firstly, we show that the node density actually helps to reduce the delivery latency and, secondly, that short contact times (few seconds), produced by non-intentional mobility, give a great contribution to reduce the delivery latency and to increase the delivery probability. Finally, we show how the human social relationships can be profitably exploited for message forwarding in dense spaces.

2 Pocket Trace Recorder

The design of a specific device for trace recording is mainly motivated by the need of observing and recording very short contact periods, few seconds, that arise from random mobility in a dense area. As a consequence, PMTRs [5] have been designed to operate with beaconing times ranging from 1 sec. to some configurable value which depends on the mobility environment we wish to observe. Secondly, the devices have to enable unmanned experiments lasting 3-4 weeks without batteries substitution. Layer 2 beacons are the unique frames a PMTR broadcasts to its neighbors. The first time a beacon is received from a given encounter, the current time is recorded for the contact start, together with the encounter ID; a contact ends when beacons from a certain encounter have been missing for more than t seconds, with $t = 60$ in our experiments. The local memory size should be dimensioned to store the contacts of the experiments. Our test beds have generated on average 2000 contacts per device with beaconing time set to 1 sec. No specific bandwidth and processing requirements have been envisaged for PMTRs.

The PMTR architecture is described in Fig.1. It uses the Cypress CY8C29566 micro-controller and the radio module AUREL, model RTX-RTLP. The radio range has been limited to 10 meters in order to reduce the power consumption and to maintain multi-hop paths between end-systems. This combination allows a very low power consumption that let the experiments last for the required time with common batteries NiMh, AA 1.2 V. Each PMTR has a 1 MB flash memory

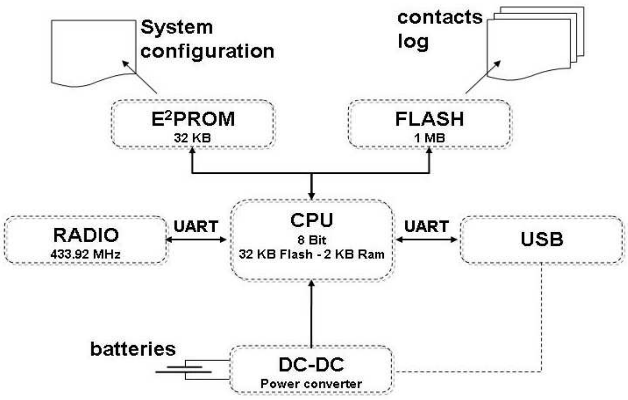


Figure 1: PMTR architecture.

where more than 50K contacts can be stored. The PMTR implements a CSMA non-persistent MAC protocol. The local clock value is set at the configuration time. Each PMTR uses a USB interface to communicate with the Pocket Viewer PC, running the Desktop application software, which has been used to configure the devices, collect the recorded data at the end of the experiment and support data analysis and device monitoring.

3 Experimental Results

The experiment has been run for 19 days in November 2008; 49 PMTRs were involved, distributed to faculty members, PhD students, and technical staff. People work in offices and laboratories located in a three-floors building, large roughly 200×100 m., and they take lunches or coffee breaks in a nearby cafeteria. Some of the lessons take place in a different building 3.5 Km far, where faculty members and students may temporarily displace. These locations are equipped with fixed PMTRs. At the end of the experiment, 5 PMTRs (with IDs 5, 8, 15, 30 and 35) showed misbehaviors in data recording. Excluding weekends, we were able to trace contacts among 44 PMTRs for 15 working days with no gaps in the traces. A collection of 11895 contacts remained for analysis.

Characterization of the Environment We consider the following statistics to usefully characterize the environment we consider: the aggregated inter-contact times, defined as the time elapsed between the end of a contact and the beginning of the successive contact for every fixed pair of nodes, the duration of the contacts called intra-contact times, and the number of neighbors seen by every node at the beginning of a contact. A coarse analysis of the traces shows the well known power law distribution of inter-contact times [1, 8, 12, 2, 10] whose complementary cumulative distribution function (ccdf) is shown in fig.2.

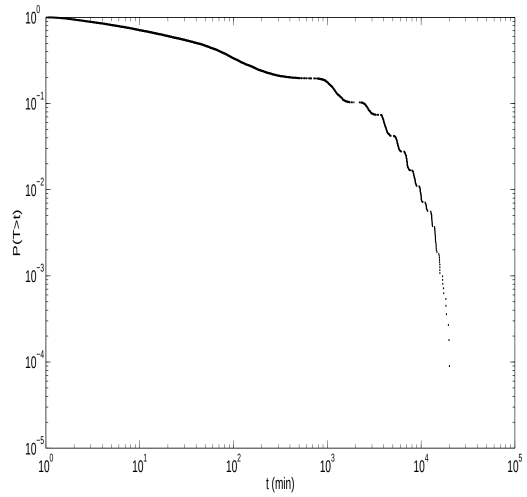
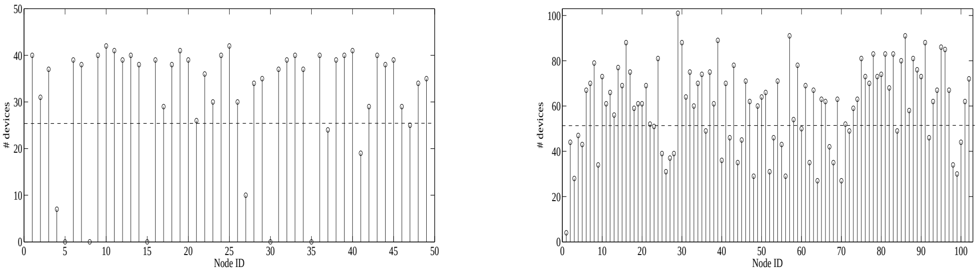


Figure 2: Inter-contact time ccdf along the whole experiment.

Table 1: Comparison between PMTR and MIT traces.

	PMTR	**MIT**
power law exponent	0.21	0.15
% contacts within a day	80%	47%
mean inter-contact time	11.81 hours	4.12 days
median inter-contact time	41.49 minutes	16 hours
mean intra-contact time	8.69 min.	57 min
median intra-contact time	0.8 min.	32 min.
mean # neighbors	2.18	2.79
median # neighbors	2	2

In Table 1, we compare these statistics estimated on our traces with those yield by the MIT Reality Mining experiment [11]. The MIT traces are considered for comparison as they characterize a scenario different from that of our experiment. Differently from MIT, PMTRs are distributed over a reduced area. As a consequence, bearers are more likely to be in range. This is confirmed by the first six parameters. As far as applications are concerned, fine-grained beaconing allows to detect a high number of short contacts, as evidenced by the median intra-contact time lower than 1 min. These contacts are not present in other traces publicly available, as the usual beaconing time adopted by other experiments [3] is no less than 120 sec. However, short contacts can effectively contribute to message forwarding, according to the observation that the most part of sent data has a small size. The difference between the two scenarios also

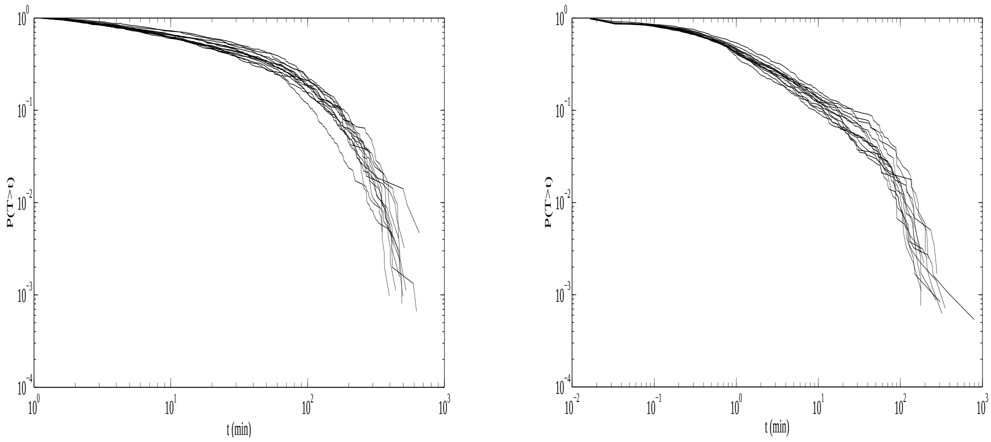


(a) (b)

Figure 3: Number of persons encountered throughout the experiment in (a) PMTR and (b) MIT traces.

reflects in a different exponent of the power law distribution of the inter-contact times. More in detail, 80% of inter-contact times in PMTRs are within a day since people, sharing the same workplace, repeatedly meet. But 50% of contacts lasts less than one minute because people meet by chance in corridor. In fig.3, the number of devices encountered by each device along the experiment is shown, for both MIT and PMTR. Only a 10% of PMTRs encounters less than 50% of the other devices, thus confirming a lower sparsity of the environment with respect to MIT, where the experiment lasted 9 months and 30% of the devices fall below the 50% line. This measure apparently contrasts with the mean number of neighbors when a contact occurs, reported in Table 1. In fact, the merge of the two results could indicate that the PMTR environment is less sparse: it is thus easier for a node to frequently detect other devices passing by, but these contacts have short duration showing a prevalence of opportunistic rather than planned encounters. Opportunistic encounters are likely to happen in presence of a lower number of neighbors with respect to planned meetings. On the other hand, the PMTR capability of detecting short encounters might more accurately capture the real connectivity graph.

Detailed Trace Analysis As a consequence of the latency requirements emerged from the survey, trace analysis focused on one-day behavior. In fig.10, the inter-contact and intra-contact time ccdfs are plotted for each day of experimentation. All days show comparable behavior; in the following, the traces for a fixed day are analyzed. According to relatively short delays requested by people surveyed, we implemented a simple greedy algorithm that computes over the trace the *optimal in latency* diffusion tree for every source. The tree provides statistics for both unicast and broadcast communications. In the considered day, 32 nodes were present, on which unicast measures were averaged. Two classes of statistics have been collected; in Table 2, the obtained indexes are summarized for both classes. In the former, we consider the whole day and evaluate the diffusion performance along the tree. An epidemic broadcast algorithm spends around 6.63 hours to reach all nodes. In fig.5, the progress of node



(*a*) (*b*)

Figure 4: (*a*) Inter-contact and (*b*) intra-contact time ccdfs for every day of experimentation.

coverage is plotted for some sources, evidencing that for most of the sources 90% of the destinations is reached within three hours; this is confirmed by the length of the paths followed by the epidemic. Yet, some nodes experiment very long latencies, which can be explained by the consideration that some of the bearers may arrive in the Department at midday, after taking/giving lesson in the other building. Due to this consideration, a second group of measurements focused on the time a source needs to reach all the nodes present in the considered environment when a message diffusion is started. In this case, we considered as sources a sample of 12 PMTRs (37% of the population in the considered day) chosen among the people most present in the Department, and we collected the same statistics as before limited to nodes present at the first appearance of the source in the trace file. These measures estimate the behavior achievable by the extemporaneous communications taken as typical application in the considered environment, where a user may want to advertise (for instance: "Remind the seminar in Room 102 at 3 pm") to all the people currently present in the workplace. The results are reported in Table 2, "one shot" columns. In this setting, the network actually meets the latency requirements for both unicast and broadcast. When considering only contacts lasting more than 5 minutes (as in many traces available from [3]), performances degrade (last column of Table 2). Excluding short contacts – detected thanks to the fine-grained monitoring available with the PMTRs– leads to loose around 80% of the encounters in this environment. The experimented unicast latency is around 30% higher; for broadcast, latency almost doubles. This could lead to the consideration that short contacts do not contribute to forwarding proportionally to their number. In fact, two aspects must be considered. First, due to the dense environment, if a contact is missed another (longer) contact occurs between the same pair of nodes soon

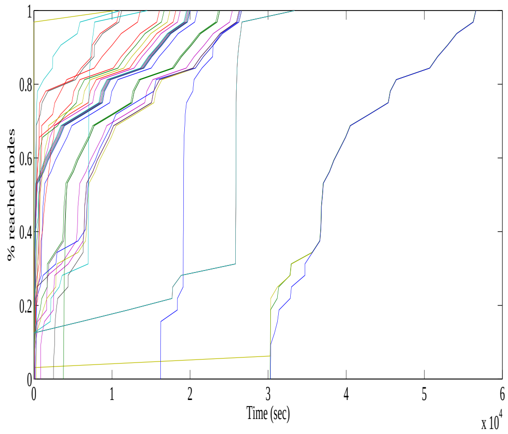


Figure 5: Coverage vs. time for broadcast transmission in one day.

after, as indicated by the low median inter-contact time (Table 1). Secondly, disregarding short contacts leads to eliminate encounters between nodes that meet only occasionally; the sources experiment an increasing difficulty in reaching other nodes – as more accurately discussed in the next subsection – and the number of destinations delivering the data is thus more than halved. As a consequence, the latencies reported in Table 2 are misrepresented. For broadcast diffusion, where measures are averaged over all the sources, the latency refers to the last reached node; but less than half of the nodes are reached. Similar considerations apply to unicast diffusions, where measures are averaged on all sources and all *reached* destinations; nodes not reached could be interpreted as observing infinite latency. Hence, the contribution of short contacts to latency is greater than that apparent from Table 2.

Forwarding Strategies In order to devise a policy both effective (in terms of delivery probability) and efficient (in terms of bandwidth consumption) to characterize relay nodes, we analyzed the paths forming the tree, and the nature of the contacts they exploit. In fig.6, the relationships between pairs of nodes are depicted as they emerge from the taxonomy proposed in [14], where the following categories are distinguished:

- community: high number of contacts with long duration;
- familiar stranger: high number of contacts with short duration;
- stranger: low number of contacts with short duration;
- friends: low number of contacts with long duration.

In particular, the mean has been used as a criteria to discriminate among classes, for both the number of contacts and the intra-contact times [7]. The taxonomy has been validated by a single-blind test, asking bearers to classify every other node in one of the above four categories: a correspondence between 75% and 98% has been confirmed. However, in the considered environment,

Table 2: Routing over PMTR traces.

		whole day	one shot	
			all	> 5'
broad-cast	min latency	2.87 h	1.48 h	2.83 h
	mean latency	6.63 h	2.55 h	4.74 h
	median latency	6.68 h	3.85 h	3.95 h
	max latency	15.73 h	5.04 h	24.41 h
	mean # hops	4-9	3-7	2-10
	coverage	98%	100%	44%
uni-cast	min latency	1"	1"	1"
	mean latency	2.91 h	40.42'	57.73'
	median latency	2.21 h	33.62'	46.12'
	mean # hops	2.98	2.81	2.97

social links also involve geographical considerations, due to the relative density of the nodes. For instance, two bearers having near offices appear in the same community also if they do not work together. Familiar strangers are bearers with offices in the same area of the workplace – although not near – such that when one of the two moves out of office, s/he has a high probability of contact with the other one. Indeed, the light grey cluster at low node IDs corresponds to persons located in the same corridor, but not in the same room. The medium grey/white cluster at IDs in range [31,34] corresponds to a group of students. In fig.7, the contour plot (flattened 3D graph) of usage of each relay (in columns) on behalf of each source (in rows) is shown, computed on the tree. Two considerations emerge. First, each source often acts as relay; this could be a consequence of nodes with scarce mobility. Second, social relationships are useful for message forwarding. The number of encounters in the trace of the considered day, and the number of relay relations in the friendship map for the same day, are shown in Table 3 for the different classes. In the tree, around 50% of the potential relays has been used for each class, but some of them has been used up to 30-40 times. We computed on fig.7 the probability $\mathcal{P}$ that an encounter of a certain class $\mathcal{C}$ is used for message relaying, as the ratio $(\sum_{i \in \mathcal{C}} \# \ uses \ for \ relay_i)/(\# \ encounters \in \mathcal{C})$. The result may seem counterintuitive, as it shows that the most used relays are either strangers or friends for the relaying node. However, both strangers and friends usually stay elsewhere with respect to a certain node. Hence, they are the most useful to bring the message to different areas of the environment and/or to destinations not familiar with a certain forwarder, thus increasing delivery probability. As a further validation, we considered – over all the relays – 12 of them that are the most used (between 31 and 43 times each; peaks in fig.7): 9 of them (75%) are strangers, 2 of them (17%) are friends and 1 (8%) is a familiar stranger; relays belonging to the same community are not useful. The same analysis over both other days and the whole experiment yields comparable results: in fig.8, the percentage of the different classes of encounters and of the different classes of

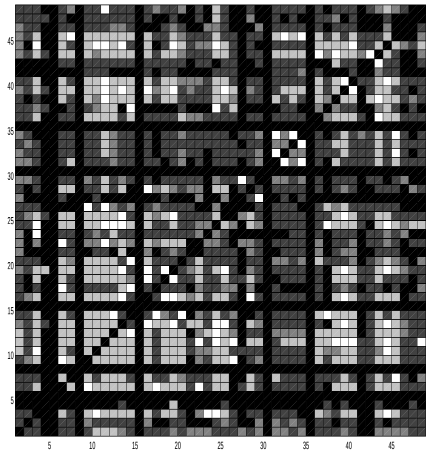


Figure 6: Friendship levels according to the taxonomy in [16]: black → no encounter; dark grey → stranger; light grey → familiar stranger; white → community; medium grey → friend.

Table 3: Statistics about forwarding.

class	#encounters	#relays	$\mathcal{P}$
community	311	68	0.3119
familiar stranger	1012	196	0.3883
stranger	392	324	0.9949
friends	133	90	0.9850

relay relationships are reported, computed over the whole experiment. In fig.9, the percentage of relays used for optimal routing and $\mathcal{P}$ are reported, computed over the whole experiment. In fig.10, the value of $\mathcal{P}$ normalized over the number of bearers present in a day is reported, for each day of experiment where at least 60% of the bearers appeared.

However, strangers are characterized by short intra-contact times; hence, disregarding short contacts could lead to eliminate encounters between strangers. In fact, classification is not that sharp, as shown by the mean and median intra-contact times (in sec.) reported in the last two rows of Table 4. For instance, two nodes belonging to the same community have a lot of contacts, some of which very long, but also many short. When short contacts are excluded, they correspond to encounters scattered in all four classes, although in prevalence concerning strangers. Two effects then follow: the prevalent elimination of encounters with strangers reduces the possibility of passing a message to nodes that bring it to different parts of the system, and as a consequence yield the de-

Table 4: Relay characteristics.

	community	fam. stran.	stranger	friends
all	10%	39%	39%	12%
$> 1'$	15%	41%	27%	17 %
$> 2'$	20%	38%	26%	16%
$> 5'$	21%	32%	18%	29%
mean intra	2314	188	120	2983
median intra	75	40	23	3040

creased coverage observed before. Moreover, the proportion among the different classes of relays changes with the length of the contacts used for forwarding : in Table 4, we report the percentages obtained by using all the contacts, and with different cut-off thresholds. By excluding short contacts, the preferred relays progressively move from strangers to friends and communities.

As a next step, we plan to design forwarding strategies such that nodes infer their friendship relations with other nodes by monitoring the number and duration of contacts. According to the social links, a subset of relays could be used in order to maintain good performance in terms of delivery latency while limiting the bandwidth consumption.

4 Conclusions

The study reported in this work provides insight into the use of ONs in the workplace for unicast and broadcast communication. The results presented in this paper do not consider group communication, although many applications might take advantage of the availability of multicast service primitives. As future work, we will extend our analysis to evaluate the routing performance for multicast, with source and destinations in one of the four friendship levels of the taxonomy considered. It is worth noting that the people surveyed claimed they seldom use multicast addressing. Rather they adopt less expensive tree-like messages dissemination (A sends a message to B, asking him/her to forward it to C and D, who in turn will forward the message to other friends, and so on until the whole community is covered). This communication pattern is a perfect fit for ON dissemination schemes.

5 Acknowledgments

This work has been partially funded by the Italian Ministry of University and Research in the framework of the “Context-Aware RouTing Over Opportunistic Networks (CARTOON)” PRIN Project (http://cartoon.nptlab.dico.unimi.it).

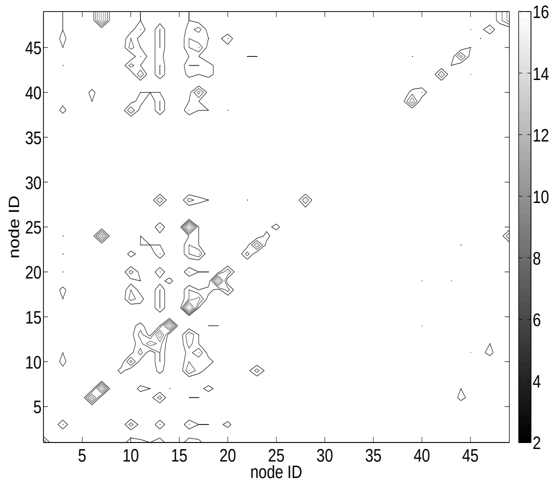


Figure 7: Relay usage.

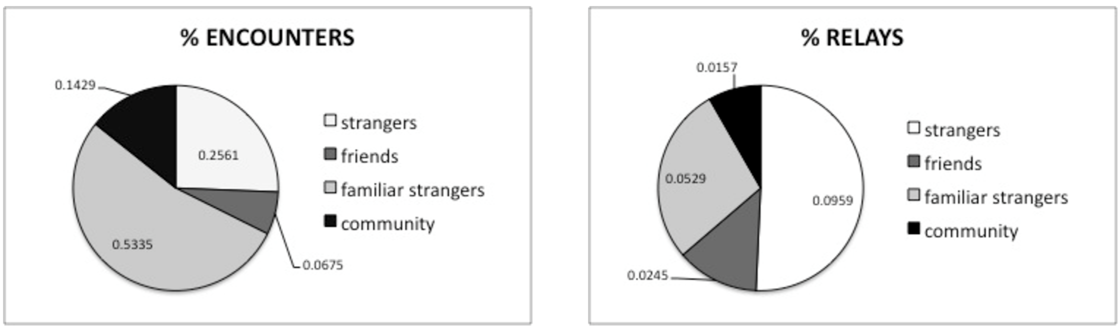


Figure 8: Percentage (a) of the different classes of encounters and (b) of the different classes of relay relationships, computed over the whole experiment

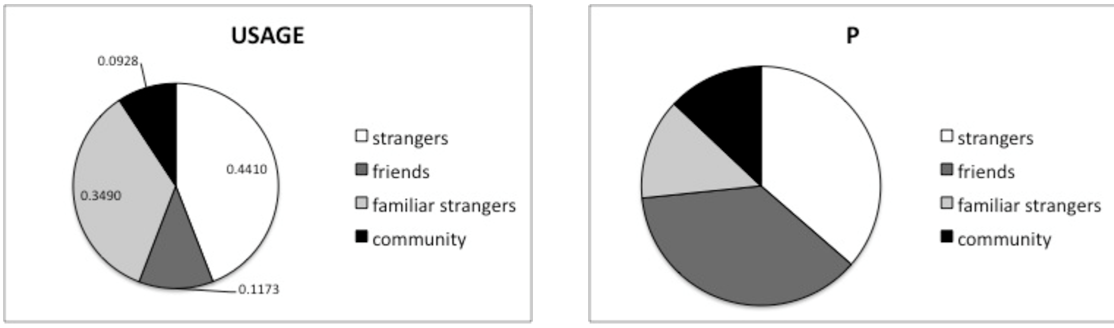


Figure 9: (a) Percentage of relays used for optimal routing and (b) $\mathcal{P}$, computed over the whole experiment

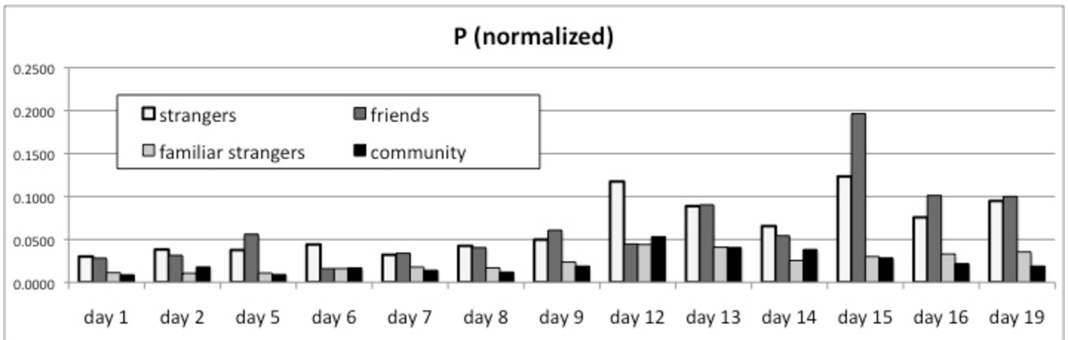


Figure 10: Value of $\mathcal{P}$ normalized over the number of bearers present in a day, for the days where at least 60% of the bearers appeared

References

[1] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott. *Pocket Switched Networks: real world mobility and its consequences for opportunistic forwarding.* Tech. Rep. UCAM-CL-TR-617, Univ. Cambridge, 2005.

[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, pp. 1-13.

[3] CRAWDAD- A Community Resource for Archiving Wireless Data At Dartmouth. http://crawdad.cs.dartmouth.edu/

[4] J. Crowcroft, E. Yoneki, P. Hui, T. Henderson. *Promoting Tolerance for Delay Tolerant Network Research.* ACM Computer Communication Review, 38(5):63-68, October 2008.

[5] S. Gaito, E. Pagani, G.P. Rossi. *Fine Grained Tracking of Human Mobility in Dense Scenarios.* In 6th Annual IEEE SECON Conference - Poster Session, June 2009.

[6] M. Grossglauser, D.N.C. Tse. *Mobility increases the capacity of ad hoc wireless networks.* IEEE/ACM Transactions on Networking, 10(4), pp. 477-486, 2002.

[7] P. Hui. *People are the Network: Experimental Design and Evaluation of Soocial-based Forwarding Algorithms.* Computer Laboratory Technical Report UCAM-CL-TR-713, University of Cambridge, March 2008.

[8] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot. *Pocket switched networks and the consequences of human mobility in conference environments.* Proc. SIGCOMM 2005 Workshop on Delay Tolerant Networking.

[9] P. Hui, J. Crowcroft, E. Yoneki. *Bubble rap: social-based forwarding in delay tolerant networks.* Proc. 9th ACM Int'l Symp. on Mobile Ad Hoc Networking and Computing, May 2008.

[10] T. Karagiannis, J. Le Boudec, M. Vojnovič. *Power law and exponential decay of inter contact times between mobile devices.* Proc. 13th ACM MobiCom, Sep. 2007.

[11] MIT, *Reality Mining*, `http://reality-media.mit.edu/`

[12] J. Su, A. Chin, A. Popivanova, A. Goel, E. de Lara. *User mobility for opportunistic ad hoc networking.* Proc. IEEE WMCSA 2004.

[13] www.twitter.com

[14] E. Yoneki, P. Hui and J. Crowcroft. *Visualizing Community Detection in Opportunistic Networks.* ACM MobiCom Workshop on Challenged Networks (CHANTS), Montreal, Canada, September, 2007.