Trace-driven analysis of data forwarding in opportunistic networks

Merkourios Karaliopoulos Panagiotis Pantazopoulos Eva Jaho Ioannis Stavrakakis
Department of Informatics and Telecommunications
National & Kapodistrian University of Athens
Ilissia, 157 84 Athens, Greece
Email: {mkaralio, ppantaz, ejaho, ioannis}@di.uoa.gr

***Abstract*—We summarize undertaken and ongoing work on the direct and exclusive use of mobile phone traces for assessing the performance of different opportunistic forwarding schemes. Our methods draw on graph-expansion techniques and circumvent the need for more custom simulation software packages. They address a wide range of opportunistic dissemination schemes including controlled flooding and socioaware protocol variants. We outline the general approach and exemplify it with an assessment of centrality metrics as drivers of data dissemination decisions. Finally, we report results on the benefits of identifying community structure out of the similarity of interests across the opportunistic network and discuss their implications for trace-based evaluation.**

I. Introduction

Both the motivation and concerns for the use of real data traces in evaluating protocols and algorithms relate exactly to the words 'real' and 'trace'. On the one hand, they promise realistic performance evaluation and credible results when compared to synthetic input data. On the other hand, they always raise concerns about their representativeness and the generality of the evaluation results. Nevertheless, the use of such traces has become the de facto approach to the evaluation of data dissemination in user-oriented network paradigms such as the opportunistic networking.

We report herein work we have been carrying out on trace-based performance evaluation of different opportunistic forwarding schemes. Our methods draw on graph-expansion techniques and circumvent the need for more custom simulation software packages. They address a wide range of opportunistic dissemination schemes coming under the controlled flooding family of protocols and can be extended to socioaware protocol variants. We focus, in particular, on two main directions that socioaware protocol design in opportunistic networks has taken: the introduction of social metrics into the individual node-oriented relaying utility functions (*e.g.,* [1]), and the explicit a priori assumption that such networks avail community structure that can be detected and exploited in disseminating data (*e.g.,* [2]).

II. Computation of shortest paths over traces

The computation of shortest path(s) for a given message proceeds in three sequential processing steps that differentiate depending on the forwarding scheme and whether shortest paths correspond to minimum-delay or minimum-hopcount space-time paths. In every case, input to this process are time-ordered traces of node contacts, *i.e.,* sequences of contact records with the general format shown in Fig. 1(a). Each contact record $c = (n_1, n_2, t_s, t_e)$ includes four fields: the two nodes that meet, the time their encounter starts, t_s, and the time the encounter ends, t_e; the difference of the last two fields corresponds to the contact duration. The fourth field becomes redundant under the infinite link capacity assumption, *i.e.,* messages need minimal (zero) time to traverse a link between two nodes once this becomes available to them.

1) From the original contact trace to the forwarding contacts: In this step, the original full contact trace is filtered with criteria that account for the different opportunistic schemes so that only those contacts that can result in forwarding of data, hereafter called *forwarding contacts*, are retained. If t_s is the time a message becomes available at source node s for destination node d, then the filtering step first excludes all contact records up to the first one $c_0 = (s, n, t_s)$ involving node s after time t_s. It then initializes an ordered list, hereafter called forwarding list, with the nodes s and n. The forwarding list stores at each time *potential* forwarders of the message, nodes that have acquired the message and may, depending on who they encounter, forward it further.

Contact records after c_0 are scanned sequentially. These contacts may belong to one of three typologies, depending on the encountered nodes: (a) neither node lies in the forwarding list; (b) both nodes are already listed in the forwarding list; and (c) one of the two nodes is in the forwarding list (1-entry contacts). Contacts of the first type do not contribute to the forwarding process and are ignored. On the other hand, contacts of the second type do not represent real additional forwarding opportunities since the assumption in all schemes is that nodes with a message copy will forward it to a node that does not have it and is eligible to acquire it upon the first encounter with it. The most interesting type of contacts is the third one, whose manipulation directly depends on the opportunistic forwarding scheme.

For example, let us consider the manipulation of 1-entry contacts for the two-hop scheme, where nodes other than the message source availing a message copy cannot forward it but only to the destination node. Therefore, two types of 1-entry contacts are retained: those $(s, *, t), t \geq t_s$ involving the source node s as the single node already logged down in the forwarding list, as well as the first 1-entry contact involving the destination node. All other 1-entry contacts are filtered out of the trace (contacts C_1-C_6 in Fig. 1(b)). In fact, when we

The work referred to in this abstract has been supported in part by the European Commission IST-FET project RECOGNITION (FP7-IST- 257756) and the Marie Curie grant RETUNE (FP7-PEOPLE- 2009-IEF-255409).

are after minimum-delay space-time paths, the filtering step terminates upon the first appearance of the destination node d in an 1-entry contact.

2) Building the forwarding contact graph: Outcome of the first processing step is the reduced set of contact records corresponding to forwarding contacts. The next step is to derive the graph construct $G_c = (V_c, E_c)$ that can capture these contacts and their timing relationship. The construct draws on the "temporal" graph representation in [3].

For forwarding schemes under the controlled flooding category, the graph construct is built out of the first contact c_0 and 1-entry forwarding contacts occurring thereafter. Each one of them adds to the graph: a) a pair of vertices, one for each node involved in the contact; b) one *space-spanning* directed edge connecting the two encountered nodes; and c) one *time-spanning* directed edge towards the node that is already included in the forwarding list, originating from the vertex that represents its most recent forwarding contact. Hence, every time a node $v \in V$ appears in an 1-entry forwarding contact, it generates a new vertex $v_c \in V_c$ for construct G_c. As a result, each network node $v \in V$ is eventually identified with a single global index in $[1, |V|]$ for the node set V and the forwarding list, and multiple non-successive indices for the construct V_c.

The graph edge set E_c is weighted. When we are interested in minimum-delay space-time paths, time-spanning edge weights equal the time differences between successive occurrences of the node in forwarding contacts and express the time over which a message may be stored and carried by a given node. Space-spanning edges express the time it takes to forward a message upon a contact and, under the infinite link capacity assumption, are assigned zero weights. On the contrary, when we are interested in minimum-hopcount space-time paths, time-spanning edges are assigned zero weights and space-spanning ones unit weights. The G_c constructs resulting from the contact trace 1(a) for the two-hop forwarding scheme is given in Fig. 1(c).

3) Computing shortest paths: The last processing step consists in the computation of shortest space-time paths over the expanded graph G_c. The size of the graph depends on whether we want to compute minimum-delay or minimum-hopcount space-time paths. When we are after the minimum-delay path(s), the parsing of contacts ends upon the first appearance of the message destination node d in an 1-entry contact. This may be anywhere from the first till the $(|V|-1)^{th}$ contact that is retained in the set of forwarding contacts. On the contrary, to include all possible minimum-hopcount paths, the parsing should continue until either all network nodes enter the forwarding list or the source nodes contact the destination node directly, whatever happens first. It can be shown that the graph constructs G_c are directed acyclic graphs (DAGs) and running Dijkstra will yield the s-d minimum-hopcount space-time path in $O\big((|V_c|+|E_c|)log_2|V_c|\big)$ = $O(|V|^2 log_2|V|)$ time [4].

III. Centrality-based data dissemination

Our trace-based performance evaluation approach can be applied to socioaware opportunistic forwarding schemes. SimBetTS [1] and BubbleRap [2] protocols compute metrics borrowed from Social Network Analysis (SNA) over contact graphs, which effectively aggregate the sequence of node encounters over certain time windows T. Both protocols have identified betweenness centrality (BC) as the dominant user-centric metric, even when it is combined with more metric when making forwarding decisions.

Centrality computation caveats: There are three main concerns regarding the realization of a user-centric approach relying solely on BC. First of all, node centrality values are destination-agnostic; namely, the node relaying utilities are averages computed across all node pairs in the network. Secondly, SNA metrics are computed over graphs. The derivation of these graphs out of the sequence (history) of contacts has been shown to be highly sensitive to the time window T during which all past contacts are aggregated into a contact graph [5]. Thirdly, the original centrality metrics need to be approximated by egocentrically computed centrality variants [6], which, in principle, offer only limited views of the node's utility in the network. We have employed real human mobility traces of pairwise node encounters to experimentally study these three factors and their impact on the BC-based data dissemination [7].

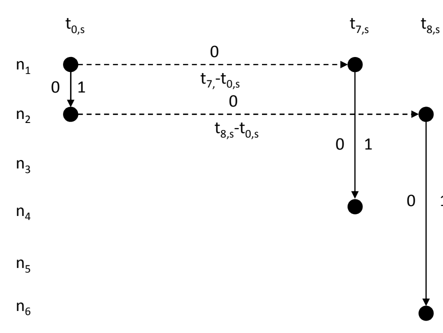
Mobile traces: We have used five well-known experimental traces, part of the iMote-based traceset available in [8]. The traces cover a rich diversity of environments with an experimental period from few days to almost one month. All traces, gathered over the last five years, include Bluetooth sightings of users carrying iMotes. Each Bluetooth sighting is assumed

contact id	involved nodes		contact start time	contact end time	additional fields
...	...	...	...	...	...
C_0	n1	n2	$t_{0,s}$	$t_{0,e}$	...
C_1	n3	n4	$t_{1,s}$	$t_{1,e}$	...
C_2	n4	n5	$t_{2,s}$	$t_{2,e}$	...
C_3	n2	n5	$t_{3,s}$	$t_{3,e}$	...
C_4	n5	n3	$t_{4,s}$	$t_{4,e}$	...
C_5	n3	n2	$t_{5,s}$	$t_{5,e}$	...
C_6	n3	n6	$t_{6,s}$	$t_{6,e}$	...
C_7	n1	n4	$t_{7,s}$	$t_{7,e}$	...
C_8	n2	n6	$t_{8,s}$	$t_{8,e}$	...
...	...	...	...	...	...

(a) Original trace.

contact id	involved nodes		contact start time	contact end time	additional fields
...	...	...	...	...	...
C_0	**n1**	**n2**	**$t_{0,s}$**	**$t_{0,e}$**	**...**
C_1	n3	n4	$t_{1,s}$	$t_{1,e}$	...
C_2	n4	n5	$t_{2,s}$	$t_{2,e}$	...
C_3	n2	n5	$t_{3,s}$	$t_{3,e}$	...
C_4	n5	n3	$t_{4,s}$	$t_{4,e}$	...
C_5	n3	n2	$t_{5,s}$	$t_{5,e}$	...
C_6	n3	n6	$t_{6,s}$	$t_{6,e}$	...
C_7	**n1**	**n4**	**$t_{7,s}$**	**$t_{7,e}$**	**...**
C_8	**n2**	**n6**	**$t_{8,s}$**	**$t_{8,e}$**	**...**
...	...	...	...	...	...

(b) Forwarding contacts: two-hop forwarding



(c) Two-hop forwarding graph.

Fig. 1. Original contact trace, forwarding contacts (entries in bold), and resulting graph construct for message $m = (1, 6, t_{0,s})$: two-hop forwarding scheme.

TABLE I
CHARACTERISTICS OF EMPLOYED DATASETS

Configuration	Intel	Cambridge	Infocom05	Content	Infocom06
Device type	iMote	iMote	iMote	iMote	iMote
Network type	B/T	B/T	B/T	B/T	B/T
Duration (days)	6	6	4	24	4
Scan time (sec)	5-10	5-10	5-10	5-10	5-10
Granularity (sec)	120	120	120	120-600	120
Mobile Devices	8	12	41	36	78
Stationary Dev.	1	0	0	18	20
External Dev.	119	211	233	11368	4421
Average internal contacts/pair/day	9.09	12.09	8.60	0.66	9.03
# of Contacts	2766	6732	28216	41330	227657

to be a contact where nodes can exchange information. In Table I scan time is the time needed by iMotes to perform a complete scan for Bluetooth devices and takes approximately 5 to 10 seconds; time granularity represents the idle time between two consecutive scans and affects significantly the measurement accuracy. We analyze only the contacts between iMotes (*i.e.*, internal contacts), which represent the data transfer opportunities among participants.

Emulation of optimal routing over the traces: The theoretically optimal paths (of minimum delay and hopcount) to the destination, have been computed directly out of the dataset sequence of encounters according to the expanded graph technique, introduced in Section II. The outcome values are naturally considered as performance benchmarks.

Emulation of BC-based routing over the traces: To account for the relative social standing of each node we need to aggregate the encounters' history to an unweighted or a more "informed" weighted static graph with link weights equal to the frequency of contacts, over which the centrality values are computed. The trace is again replayed (sequentially read) but now we compute five different centrality-variants for each contact record; a message is forwarded provided that the encountered node exhibits higher value of the corresponding variant than the one of the current holder. These values may be either the sociocentric (including the destination-aware BC variant called Conditional Betweenness Centrality (CBC) [9]) or their egocentric counterparts [6] computed over the full set of contacts within the past T time window.

Summary of results: We have generated messages with randomly chosen source and destination and emulated their paths over the traces. The message delivery delay and number of forwarding hops have been computed and compared with those of the optimal (opt) scheme described above. Our findings are summarized in the following discussion. The centrality-based approaches perform considerably worse than the optimal method both in terms of message delay and hops to the destination. Forwaring performance across the different traces depends on the way nodes' mobility patterns mix with each other. Less intuitively, replacing BC with its destination-aware counterpart (CBC) does not give benefits in terms of message delay but results in significant energy savings by reducing the message hops. When computing BC over weighted graphs our study reports that the performance does not consistently improve for all traces. Nevertheless, the routing protocol is more resilient to variations of the time window used for contact aggregation. Finally, we have found that using the egocentric BC variant penalizes the forwarding performance only marginally, when computed over unweighted and even less when computed over weighted graphs. This is further supported by the observed strong positive correlation between socio- and egocentric BC in almost all traces. A more detailed discussion as well as explanatory plots appear in [7].

IV. ENRICHING TRACES WITH USERS' INTERESTS

In [10], we have enhanced primitive push mechanisms for data dissemination in opportunistic settings with social information concerning the preferences and interests of network nodes. It is shown that *interest-based forwarding* can improve considerably the information dissemination process. Inspired by this result, in [11] we have proposed a framework called *ISCoDe*, which identifies communities of nodes with similar interests. We have applied *ISCoDe* to the Delicious (www.delicious.com) platform, showing how end-user interests can be inferred out of a real online social networking (OSN) application.

Adding further realism to this thread would call for data traces that, besides encounters, provide information about the preferences of users. Enriching mobile phone datasets with such information is one option that could be considered in future datasets to become available. This information encoded in the form of user preference distributions (profiles) over a set of certain thematic areas (such as music, sports, art), can be inferred out of tags annotating data that users save in their mobile phones. A more demanding alternative would be to indirectly infer such information from online social networks, trying to correlate traces of encounters with OSN user profiles (*e.g.*, [12]).

REFERENCES

[1] E. M. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant manets," *IEEE Trans. Mob. Comput.*, vol. 8, no. 5, pp. 606–621, 2009.

[2] P. Hui *et al.*, "Bubble rap: Social-based forwarding in delay tolerant networks," *IEEE Trans. Mob. Comput.*, (To Appear) 2011.

[3] V. Kostakos, "Temporal graphs," *Physica A-statistical Mechanics and Its Applications*, vol. 388, pp. 1007–1023, 2009.

[4] M. Karaliopoulos, "Trace-based performance analysis of opportunistic forwarding under imperfect cooperation conditions," University of Athens, Tech. Rep., July 2011.

[5] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing," in *Proc. IEEE Infocom 2010*, San Diego, CA, USA, Mar 2010.

[6] P. Marsden, "Egocentric and sociocentric measures of network centrality," *Social Networks*, vol. 24, no. 4, pp. 407–422, October 2002.

[7] P. Nikolopoulos, T. Papadimitriou, P. Pantazopoulos, M. Karaliopoulos, and I. Stavrakakis, "How much off-center are centrality metrics for routing in opportunistic networks," in *ACM MobiCom 2011 CHANTS Workshop)*, Las Vegas, NV, USA, Sep 2011.

[8] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD data set cambridge/haggle (v. 2009-05-29)," Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle, May 2009.

[9] P. Pantazopoulos, I. Stavrakakis, A. Passarella, and M. Conti, "Efficient social-aware content placement for opportunistic networks," in *IFIP/IEEE WONS*, Kranjska Gora, Slovenia, February, 3-5 2010.

[10] S. Allen, M. Chorley, G. Colombo, E. Jaho, M. Karaliopoulos, I. Stavrakakis, and R. Whitaker, "Exploiting user interest similarity and social links for microblog forwarding in mobile opportunistic networks," in *Elsevier Pervasive and Mobile Comuting (submitted)*, 2011.

[11] E. Jaho, M. Karaliopoulos, and I. Stavrakakis, "ISCoDe: A framework for interest similarity-based community detection in social networks," in *Proc. 3rd Int'l Workshop on Network Science for Communication Networks*, 2011.

[12] G. Bigwood and T. Henderson, "Bootstrapping opportunistic networks using social roles," in *Proc. 5th IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC)*, 2011.