Prediction Quality of Contact History in Opportunistic Networks

Muhammad Arshad Islam*   Marcel Waldvogel*
* Distributed Systems Laboratory
University of Konstanz – Germany
firstname.lastname@uni-konstanz.de

Abstract—**Routing in opportunistic networks heavily relies on past behavior of the mobile devices it is formed of to predict their future and thus making routing decisions. While almost every protocol relies on this history, its prediction quality has never been studied in a realistic setting. Using extensive simulations on real traces, we are able to describe for the first time how well predictions can be. Unlike oracle-based prediction comparisons, we do not try to predict a contact, but compare the expected user result, namely message delivery probabilities. The analysis also provides guidance on the importance of multi-path routing and the path diversity required, as well as on the impact on forward error correction on the delivery probability. Our results show that the repetitive nature of path is directly proportional to the mobility extent of the devices and, consequently, history obtained from dense opportunistic networks is reliable.**

I. INTRODUCTION

Since the initial introduction of Delay Tolerant Networks on the research horizon for interplanetary communication[2], several offshoots have spawned, e.g. *Vehicular Networks, Mobile Social Networks* and *Opportunistic Networks*. Similarly, several practical applications, such as an emergency response in case of a catastrophe, military operations and non-interactive Internet access in rural areas[6] have vastly increased the usability of such networks.

Every routing protocol in literature is one way or another dependent on history of the devices to extract routing information. Moreover, each routing protocol deploys its own way of collecting the history that is distinct with respect to several aspect including (a) what kind of history information is collected, (b) how frequent is it collected, and (c) what measures are taken to maintain the minimum device storage consumption. Moreover, due to hardware limitations, the size of routing information must be limited, which introduces inaccuracies in the measures. Consequently, obtaining accurate and precise traffic measures for participating devices is a great challenge. One may expect that more accurate paths and traffic measures will lead to better message delivery.

Given these arguments, it is easy to conclude that researchers face the challenge of acquiring accurate and precise information to make correct routing decision. Delays and device mobility makes the access to information like network topology and traffic volume very difficult. Jain, Fall and Patral[5] have proposed several oracles with future insight. Although, such methods are unrealistic, they can help to understand the nature and behavior of underlying networks. Opportunistic networks can be seen as good examples of distributed systems[?], which can be simulated and analyzed
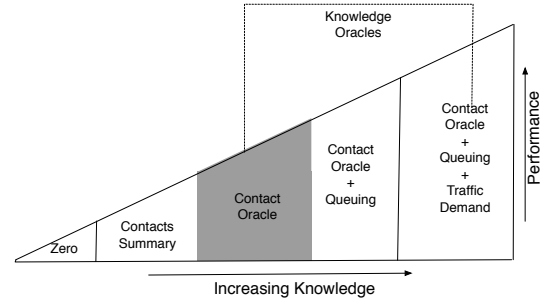

Fig. 1: Conceptual performance vs. knowledge trade-off for different oracles.[5]

with the help of oracles that have the capability of delivering different kinds of network measures without delay, throughout the network . Mechanisms that provide information to predict the device and traffic behavior, and which are difficult or impossible to gather in realistic scenarios, are known as *oracles*[5]. Jain, Fall and Patra [5] have presented classification of several oracles based on the extent of information they can deliver as shown in Fig. 1. Motivated by the work in[5], [?], we have performed experiments to verify the repetitive nature of routing paths on three different opportunistic networks. First we use contact oracle to find the paths and then the re-occurrence of these paths is verified by checking their delivery ratio in the a different time period by assuming that history information is spread instantaneously throughout the network. Our results show that as the history grows older, it looses its precision as well as accuracy.

II. SIMULATION SETUP

We have considered three different kinds of data sets, all of which have been obtained from CRAWDAD. The motivation behind choosing these three traces have been a broad spectrum between dense and sparse networks. Two of the data sets have been synthesized from reality mining project [4] from MIT spans on 16 months, i.e. February 2004 to August 2005 whereas, the third data consist of the SNMP logs for one month from a IBM campus[3].

In the case of the IBM access point trace, SNMP is used to poll access points (AP) every 5 minutes, from July 20, 2002 through August 17, 2002. A total of 1366 devices have been polled over 172 different access points during approximately 4 weeks. To turn these samples into continuous data, we assume that the snapshot data remains constant for the next 5 minutes. We further assume that two nodes that are connected to one

access point during the overlapping time period are connected to each other.

The second trace of the MIT cell tower is utilized according to the similar principal that was used for the IBM traces. The only difference is that instead of access points, cell towers are used to gather the contact times of the nodes, thus the resulting network can be characterized as a very dense network due to the high range of the cell tower. As the duration span of the the MIT reality mining is longer than the IBM trace, we have filtered the MIT data to match the time span of the IBM traces. The span time of the IBM trace is approximately one month whereas for MIT is more than one year, we have chosen one month from cell tower on the basis of the activity, so that the results can be compared.

The most sparse network is obtained from bluetooth logs (MITBT) where each node scans every five minutes for active bluetooth neighbors and stored the duration of contact times. Like the MIT trace, we selected one month from bluetooth traces, i.e. November 2004 showed 1858 bluetooth nodes suggesting a huge number of undesignated nodes as compared to the designated[1] 81 nodes that were designated to gather the data. Here it is noteworthy that a few undesignated devices had more connectivity and interaction with the network than the designated nodes.

A. Analysis Methodology

Typically, network analysis requires finding a maximal-flow solution to identify bottlenecks when there are capacity constraints on the arcs. The maximum flow problem is structured on a network, however, the arc capacities or upper bounds, are the only relevant parameters. Given a graph where one vertex is considered a source and another is the sink, some object then flows along the edges of the graph from the source to the sink consuming the corresponding capacities of the paths. Readers interested in background and theoretical proofs of problems related to max-flow may consult[1].

We have chosen 10 src,dst pairs, where max-flow mechanisms establish the throughput of the variable number shortest path. As opportunistic networks are practically time varying graphs, different max-flow starting point in time may result in different measurements. For this reason, we have created multiple contact oracle based max-flows for one trace week period at different starting times. Moreover, each max-flow comprises of a variable number of shortest paths. It is important to notice that these shortest paths are not necessarily disjoint and one path may overlap with the others paths belonging to either the same or the next starting time. Message delivery to the destination using the stored path may fail due to two reasons, (a) The stored path is not repetitive in nature and thus failed to connect the source with the destination. (b) The stored path does connect the source to the destination but is not able to carry the prescribed message volume.

III. RESULTS DISCUSSION

In all the figures presented in the results, Y-axis represents the amount of data delivered and X-axis represent the delay in days encountered during the propagation to destination.

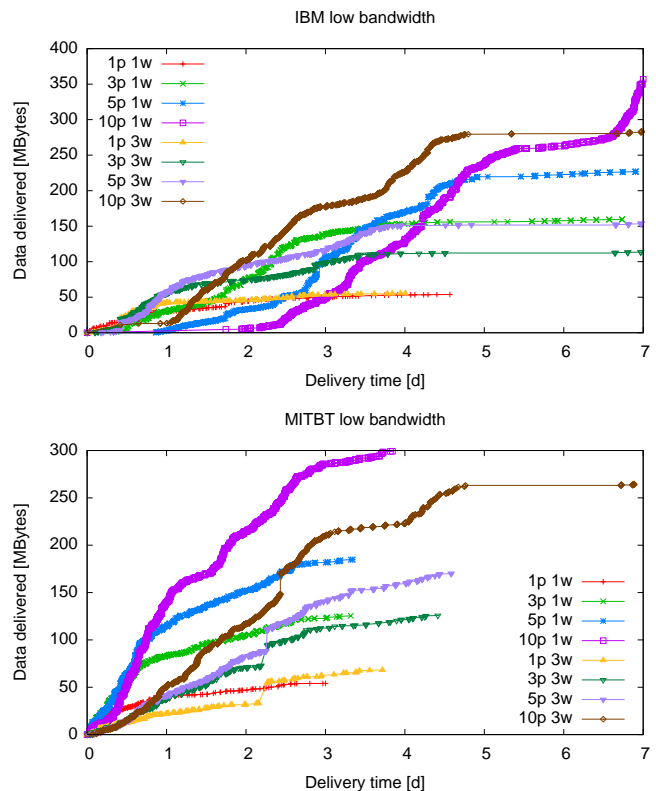[1]Nodes running the scanning software are referred to as designated

Fig. 2: Oracle behavior among all traces

a) Oracle behavior: All the three trace figures as shown in Fig. 2 have four plots using the contact oracle. The number prefixed with p in each plot represents the number of the shortest paths that have been considered for the corresponding plot, while the number prefixed with w represent the week during which the corresponding measures have been computed. We have simulated on 1,3,5 and 10 first paths(1p, 3p, 5p, 10p) during the first, second and third week(1w, 2w, 3w) of the trace timespan. The frequent repetitive nature of device contacts is evident in the MIT trace where all the paths arrive within 1.8–2.1 days at the destination, and the curves for all the path counts are similar to each other during the first week. *Oracle-1p* that utilizes the first shortest path, delivers approx. 5MB after 1.8 days whereas *Oracle-10p* delivers approx. 38MB in 2.1 days using 10 first shortest paths.

We have different starting points as well as dissimilar curves in the IBM trace during the first week(1w), while all the curves for the third week(3w) have the same shape for the first day. This behavior can be coupled with the first week as there is a significant delay between the first and second path for the first week(1w). This behavior is absent during the third week(3w). Presence of a dissimilar curves of *Oracel-10* in the case of the MITBT trace show a lower frequency of repetitive paths when a bigger number of the first shortest paths is taken into consideration. This feature is more evident in the case of the IBM trace where all the curves have different starting points.

b) History quality: Each of the figures has 4 plots representing propagation during the 4 weeks of the trace periods. In each figure, there is one plot for propagation with the help of the contact oracle, while the remainder of the plots represent the propagation using the stored path of the oracle.
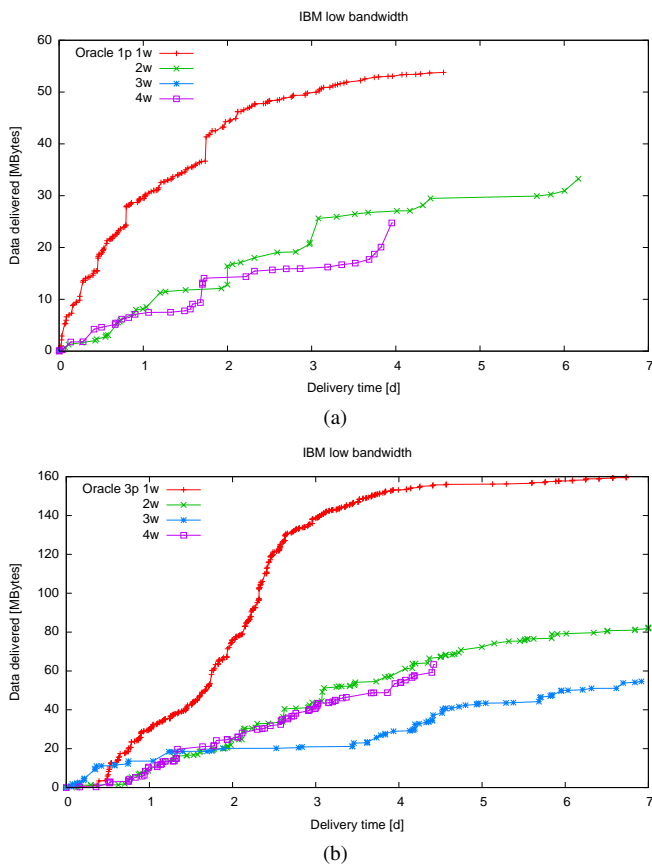
Fig. 3: History quality of IBM trace

Fig. 4: History quality of MIT trace

The number prefixed with letter p represents the number of first paths obtained from the oracle, while the number prefixed with letter w shows the week of the corresponding plot in the figure.

When we compare oracle delivery times of the first week belonging to the IBM trace (Fig. 3 (a)) with successive weeks, we see prominent degradation during the capability of the paths reaching the destination. This suggests that most of the paths failed to occur in the later 3 weeks of the trace with the same throughput as shown in first week. In an extreme case of the first shortest path (Fig. 3)(a), the shortest paths among all the chosen source and destination pairs are totally absent in the period of the third trace week. As far as the figures of 1p and 3p (Fig 3(a),(b)) are concerned, the second trace week has a relatively higher throughput from other weeks, while in the remainder of the cases there is no such difference visible. This suggests that advantages of recent history prediction may diminish when larger number of shortest paths are considered.

When we look at the second set of 4 figures corresponding to the MIT trace (Fig. 4(a)(b)), we observe that all the figures show a strong repetitive nature of paths during the proceeding 3 weeks with relatively comparable throughput shown by the oracle in the first trace week period. One interesting characteristic in the MIT trace is the variance of throughput with a changing number of utilized shortest paths, among the several weeks. For example, the throughput obtained during second week with 3 and 5 shortest path as shown in Fig. 4(a)(b) is the lowest among all, whereas it improves when we utilize 1 shortest paths as shown in Fig. 4(a).
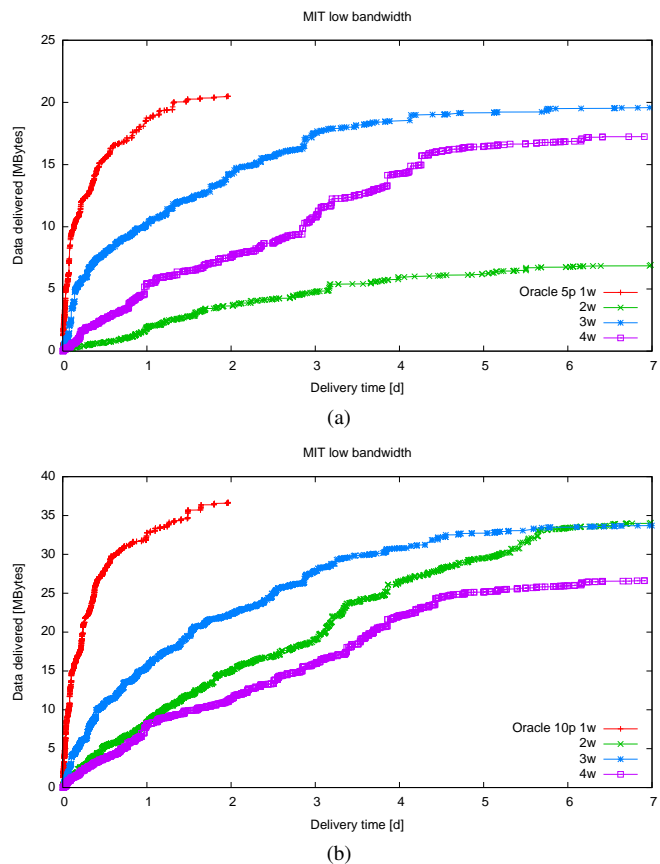
IV. CONCLUSION AND FUTURE WORK

Our results show that the density of a network does not necessarily relates to the reliability of the gathered history information. In future, we will design more innovative experiments to establish a relationship between the way history information is gathered with the nature of the network. Further on, we will design mechanism that will help a routing protocol to adapt to a suitable way of utilizing history to make optimal routing decisions.

REFERENCES

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice Hall, 1993.
[2] I. F. Akyildiz, Özgür B. Akan, C. Chen, J. Fang, and W. Su, "Interplanetary internet: state-of-the-art and research challenges," *Computer Netwroks*, vol. 43, no. 2, pp. 75–112, 2003.
[3] M. Balazinska and P. Castro, "CRAWDAD data set ibm/watson (v. 2003-02-19)," Downloaded from http://crawdad.cs.dartmouth.edu/ibm/watson, Feb. 2003.
[4] N. Eagle and A. S. Pentland, "CRAWDAD data set mit/reality (v. 2005-07-01)," Downloaded from http://crawdad.cs.dartmouth.edu/mit/reality, Jul. 2005.
[5] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," in *Proceedings of SIGCOMM 2004*. ACM Press, 2004, pp. 145–158. [Online]. Available: http://portal.acm.org/citation.cfm?id=1015484
[6] A. S. Pentland, R. Fletcher, and A. Hasson, "DakNet: Rethinking connectivity in developing nations," *IEEE Computer*, vol. 37, no. 1, pp. 78–83, Jan. 2004.