Characterizing Pairwise Inter-contact Patterns in Delay Tolerant Networks

Vania Conan,* Jérémie Leguay,* Timur Friedman†
* Thales Communications
† Université Pierre et Marie Curie, Laboratoire LiP6–CNRS

Abstract— **A good understanding of contact patterns in delay tolerant networks (DTNs) is essential for the design of effective routing schemes. Prior work has typically focused on inter-contact time patterns in the aggregate. In this paper, we argue that pairwise inter-contact patterns are a more refined and efficient tool for characterizing DTNs. We provide a detailed statistical analysis of pairwise inter-contact times in three reference DTN data sets. We characterize heterogenities in inter-contact times, and find that the empirical distributions tend to be well fit by log-normal curves, with exponential curves also fitting a significant portion of the distributions.**

I. Introduction

Understanding mobility of nodes in DTNs is of utmost importance. A large number of design issues such as routing, content dissemination or resource management much depend upon what one expects in terms of node mobility.

Initial DTN work focused on exploiting scheduled meeting times [6]. Focus then turned to the sort of randomness in meeting times [4], [8], [9], and characterised in mobility models such as Random Way-Point, and Random Walk. These models yield homogeneous patterns, where all nodes share a single inter-contact time distribution. More recent work has analysed experimental data sets [5], [1], [3] that record actual inter-contact patterns that occurred between people in a number of different environments. Chaintreau et al. [1], from observations on those data sets, proposed to model the sequence of contacts as a discrete renewal process, and study power-law distributed inter-contacts. Karagiannis et al. [7] analyse the mobility traces and explain the observed exponential tail behavior of inter-contact times with a simple random walk on a two dimensional torus followed by all nodes in the network.

In this paper, we advocate that researchers should look at pairwise inter-contact patterns. We provide a detailed statistical analysis of pairwise inter-contact patterns in three reference DTN data sets. Previous work has studied inter-contact times in the aggregate, across all pairs of nodes. It has combined, and thus obscured, the individual effects of pairwise inter-contacts. We characterize heterogenities in inter-contact times, and find that distributions of inter-contact times tend to be well modeled by log-normal curves. Exponential curves also tend to fit a fair portion of distributions.

II. Pairwise interactions

A. Experimental data sets

We describe here the contexts in which the data sets that we used have been collected.

Dartmouth data

This connectivity data set has been inferred from traces collected in the Wi-Fi access network of Dartmouth College [5]. The traces track users' sessions in the wireless network, noting the time at which nodes associate and dissociate from access points. Although the Dartmouth data is not from a DTN network, we use it because it is perhaps the richest data set publicly available that tracks users in a campus setting, and because of its quality.

A few judicious assumptions are required to adapt the Dartmouth data for DTN studies. First, we only consider the subset of users who were present in the network every day between January 26th 2004 and March 11th 2004, an academic period during which we expect nodes' activity to be fairly stationary. This data set contains 834 users, or nodes. Then, we assume that two nodes are in contact if they are attached at the same time to the same access point (AP). Finally, we filter the data to remove the well known *ping-pong* effect. Wireless nodes, even non-mobile ones, can oscillate at a high frequency between two APs. To counter this, we filter all the inter-contact times below 1,800 seconds.

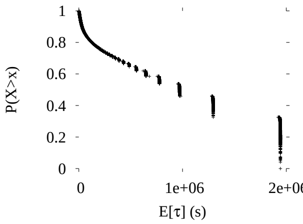
iMote data

Chaintreau et al. [1] used iMotes to acquire proximity contacts that occurred between participants in the student workshop at the *Infocom 2005* research conference. Students were asked to carry one of these sensors in their pocket at all times. Due to Bluetooth's short range, authors logged instances when people were close to each other (typically within 10 meters). They collected data from 41 iMotes over 3 days. The devices performed Bluetooth inquiry scans every 2 minutes. For each pair of nodes (i, j), we considered that i and j were in contact if either one saw the other.

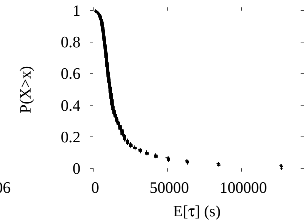
MIT data

The Reality Mining experiment [3] conducted at MIT captured proximity information from 97 subjects over the course of an academic year. Each participant had an application running on their mobile phone to record proximity with others through periodic Bluetooth scans (every 5 minutes) in a similar fashion to that of the iMote experiment. We used the first 95 days of data.

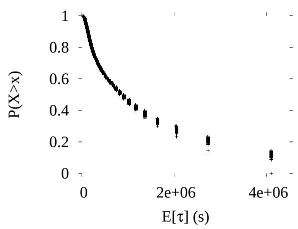
We will refer to these data sets as *Dartmouth*, *iMote* and *MIT*.



(a) E(τ) on Dartmouth



(b) E(τ) on iMote



(c) E(τ) on MIT

Fig. 1. CDF of mean inter-contact times E(τ).

B. *Heterogeneity in expectations*

This section looks at times that enlapse between instants when pairs of nodes are in contact (*inter-contact times*). We focus on heterogeneity, looking at the distributions for all node pairs.

Fig. 1 shows the cumulative distribution, for all node pairs, of mean inter-contact times. We denote with E(τ) the expectation of inter-contact times, with τ being the process of inter-contact times for a given pair. Similarly, Fig. 1 shows the distribution of E(Ω), the expected contact times of node pairs. We can see that the distributions are heterogeneous, with the means spanning over three orders of magnitude. The mean inter-contact time is 280.6 hours for Dartmouth, with a standard deviation of 210.5 hours; 4.9 hours for iMote, with a standard deviation of 5.6 hours; and 387.1 hours for MIT, with a standard deviation of 377.3 hours.

C. *Nature of inter-contact times distributions*

To better understand the inter-contact time processes between node pairs, we test for whether the distribution of inter-contact times between any two nodes can be modelled either by an exponential, a log-normal, or a power law (to be precise, Pareto) distribution. For this purpose, we used the Cramer-Smirnov-Von-Mises [2] statistical hypothesis test with a confidence of 99% and we cross-checked the results visually. For each pair of nodes (i, j) having at least 4 contacts, we compare the cumulative distribution I_N^{ij} of the N inter-contact times observed and the hypothesis functions. Note that, for a given node pair, several distributions may be identified.

Table I presents, for each data set, the proportion of pairs for which the distribution of inter-contact times fits an exponential, a Pareto, and a log-normal distribution. We also show the proportion of pairs that were rejected for all three hypothetical distributions.

One notable observation is that log-normal tends to fit better than exponential or Pareto for all three data sets. Almost no pair of nodes has been found fit only an exponential or a Pareto. For Dartmouth, for example, 0.1% of node pairs are exponential only, and the same proportion are Pareto only, while 36.4% of node pairs only match a log-normal distribution.

	Dartmouth	**iMote**	**MIT**
Number of pairs tested	20,211	755	2,174
Exponential	42.8 %	7.9 %	56.3 %
Pareto	34.2 %	12.3 %	26.5 %
Log-normal	85.8 %	99.4 %	96.9 %
None	12.9 %	0.4 %	2.7 %

TABLE I

FITTING RESULTS.

From these observations, it seems reasonable, in these data sets, to consider pairwise inter-contact time distributions as log-normal rather than power law or exponential. This speaks to the heterogeneity of the distributions. The log-normal family is better capable of modeling the variations of behaviors across the pairs of nodes. The reasons are probably twofold. First, it covers a large span of asymptotic behaviors at the origin (from horizontal to vertical asymptotes). Second, it can capture light tailed behavior as well as some heavy tailed behavior, while always maintaining a finite expectation and/or variance.

III. CONCLUSION

In this paper, we argue for the wisdom of using pairwise inter-contact patterns to characterize DTNs. We have provided a statistical study using widely-used DTN data sets in which we characterize heterogeneity of interactions between nodes. We show that pairwise inter-contact times processes, which have a great impact on routing, are heterogeneous and distributed in log-normal for a large number of node pairs.

REFERENCES

[1] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. INFOCOM*, 2006.

[2] W.T. Eadie. *Statistical Methods in Experimental Physics*. Elsevier Science Ltd, 1971.

[3] N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, 2005.

[4] K. Harras, K. Almeroth, and E. Belding-Royer. Delay tolerant mobile networks (DTMNs): Controlled flooding schemes in sparse mobile networks. In *Proc. Netwoking*, 2005.

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

[6] S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant network. In *Proc. SIGCOMM*, 2004.

[7] T. Karagiannis, J-Y. Le Boudec, and M. Vojnovic. Power law and exponential decay of inter contact times between mobile devices. In *Proc. ACM MOBICOM*, 2007.

[8] N. Sarafijanovic-Djukic and M. Grossglauser. Last encounter routing under random waypoint mobility. In *Networking*, 2004.

[9] T. Spyropoulos, K. Psounis, and C. Raghavendra. Single-copy routing in intermittently connected mobile networks. In *Proc. IEEE SECON*, 2004.