Characterizing Pairwise Inter-contact Patterns in Delay Tolerant Networks

(Invited Paper)

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

ABSTRACT

A good understanding of contact patterns in delay tolerant networks (DTNs) is elemental to the design of effective routing or content distribution 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. First, we provide a detailed statistical analysis of pairwise contact and inter-contact times in three reference DTN data sets. We characterize heterogeneities in contact times and inter-contact times, and find that the empirical distributions of inter-contact times tend to be well fitted by log-normal curves, with exponential curves also fitting a significant portion of the distributions. Second, we investigate analytically the relationship between pairwise and aggregate inter-contact times. In particular, we consider both the exponential and log-normal cases and show analytically how the aggregation of pairwise inter-contacts may lead to aggregate inter-contacts with power laws of various degrees.

1. INTRODUCTION

In delay tolerant networks (DTNs) [7] nodes are typically mobile and have wireless networking capabilities. They are able to communicate with each other only when they are within transmission range. The network suffers from frequent connectivity disruptions, making the topology only intermittently and partially connected. This means that there is no guarantee that an end-to-end path exists between a given pair of nodes at a given time. Examples from the recent literature include the DieselNet project [26], which features communication devices deployed in a regional bus system, and Pocket Switched Networks (PSNs) [3], which are formed by devices that people carry every day, such as cell phones, PDAs, and music players. In contexts such as these, end-to-end paths can exist temporarily, or may sometimes never exist, with only partial paths emerging.

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. We provide in this work a detailed statistical analysis of pairwise contact and inter-contact times in three reference DTN data sets which allow us to draw conclusions about the characteristics that DTN models should integrate. We characterize heterogeneities in contact times and inter-contact times, and find that the empirical distributions of inter-contact times tend to be well fitted by log-normal curves, with exponential curves also fitting a significant portion of the distributions. Second, we investigate the relationship between pairwise and aggregate inter-contact times. In particular, we consider both the exponential and log-normal cases and show analytically how the aggregation of pairwise inter-contacts may lead to aggregate inter-contacts with power laws of various degrees.

Initial DTN work focused on scheduled meeting times [12]. Focus then turned to the sort of randomness in meeting times encountered in mobile ad-hoc networks [10, 23], and characterised in mobility models such as Random Way-Point [13], and Random Walk [6]. These models yield homogeneous patterns, where all nodes share a single inter-contact time distribution. Spyropoulos et al. model mobility of nodes as independent random walks on a torus, and use it to analyse the performance of different routing schemes [25]. Their model considers all pairs of nodes to follow the same law, with the same parameters. This results in a homogeneous DTN (a network where the mobility of nodes results in all pairwise inter-contacts to follow exactly the same law).

More recent work has analysed experimental data sets [11, 3, 5] that record actual inter-contact patterns that occurred between people in a number of different environments. Chaintreau et al. [3], 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. [15] 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 make two contributions along these lines: First, 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 heterogeneities in contact times and 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.

Second, we provide analytical studies of the relationships between aggregate and pairwise distributions. One often adopts the distributions estimated on aggregate data for modelling pairwise inter-contacts. We show that this may be a too strong simplification which leads to paradoxical situations. For example, when one

Autonomics, October 28–30, 2007, Rome, Italy.

wants to route messages, results are too pessimistic. We describe how distributions with finite means and variances can be composed to yield distributions whose tails follow a power law. This corresponds to the power-tailed distributions that Chaintreau et al. [3] observed. In particular, we show that exponential and log-normal pairwise inter-contact time distributions can be combined to create an aggregate power law distribution.

The rest of this paper is structured as follows. Sec. 2 provides a statistical analysis of pairwise contacts in the three real life data sets we used in this work. Sec. 3 discusses the power law paradox. Sec. 4 discusses the results and describes related work concerning mobility in DTNs. Sec. 5 concludes the paper, discussing directions for future work.

2. PAIRWISE INTERACTIONS

This section introduces and analyses the three different data sets that we use in the rest of this paper. We characterize interactions that may occur in DTN scenarios and highlight the different kinds of heterogeneities that arise.

2.1 Experimental data sets

We describe here the contexts in which the data sets have been collected and the acquisition methodologies that were used. All of these data sets are publicly available in the CRAWDAD archive [1].

Dartmouth data

This connectivity data set has been inferred from traces collected in the Wi-Fi access network of Dartmouth College [11]. The traces that we use were pre-processed by Song et al. [24] for their prior work on mobility prediction. They 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. Jones et al. [14], Leguay et al. [18], and Chaintreau et al. [3] have recently used these traces in a similar way.

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). We miss other contacts between users that are not logged by Wi-Fi devices, because the users are not carrying the devices or have turned them off. These contacts might have been logged in a true DTN network, by lighter-weight wearable devices that remain on at all times. When more extensive DTN data in campus settings becomes available, researchers will need to revisit the studies made using the Dartmouth data, to see if the lack of such contacts has an impact on their conclusions. 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 (30 minutes). Note that defining better filtering methods, although challenging, would be of interest for the community. As this is not the purpose of this work, we choose here the threshold that Yoon et al. [27] used for the same purpose. We use this inferred data set for the remainder of this paper.

Fig. 1 presents, for all the data sets, the evolution over time of the total number of contacts that occurred between nodes (left column) and the number of contacts for every pair of nodes having at least one contact, ranked in decreasing order (right column). Fig. 1(a) and Fig. 1(d) are the plots for the Dartmouth data set. As Fig. 1(a) shows, the interactions between nodes are quite stable over time. We observed 13,901.7 contacts per day on average, with a standard deviation of 796.9 contacts. We conjecture that this stability comes from the fact that we choose only nodes that are present every day. Fig. 1(d) shows that a few node pairs had a high number of contacts, and that this number then decreases very rapidly. Just 10.7% (i.e., 37,424) of node pairs had contacts between each other, and these are the ones that are plotted. Among these, the mean number of contacts was 15.4, with a standard deviation of 32.9 contacts.

iMote data

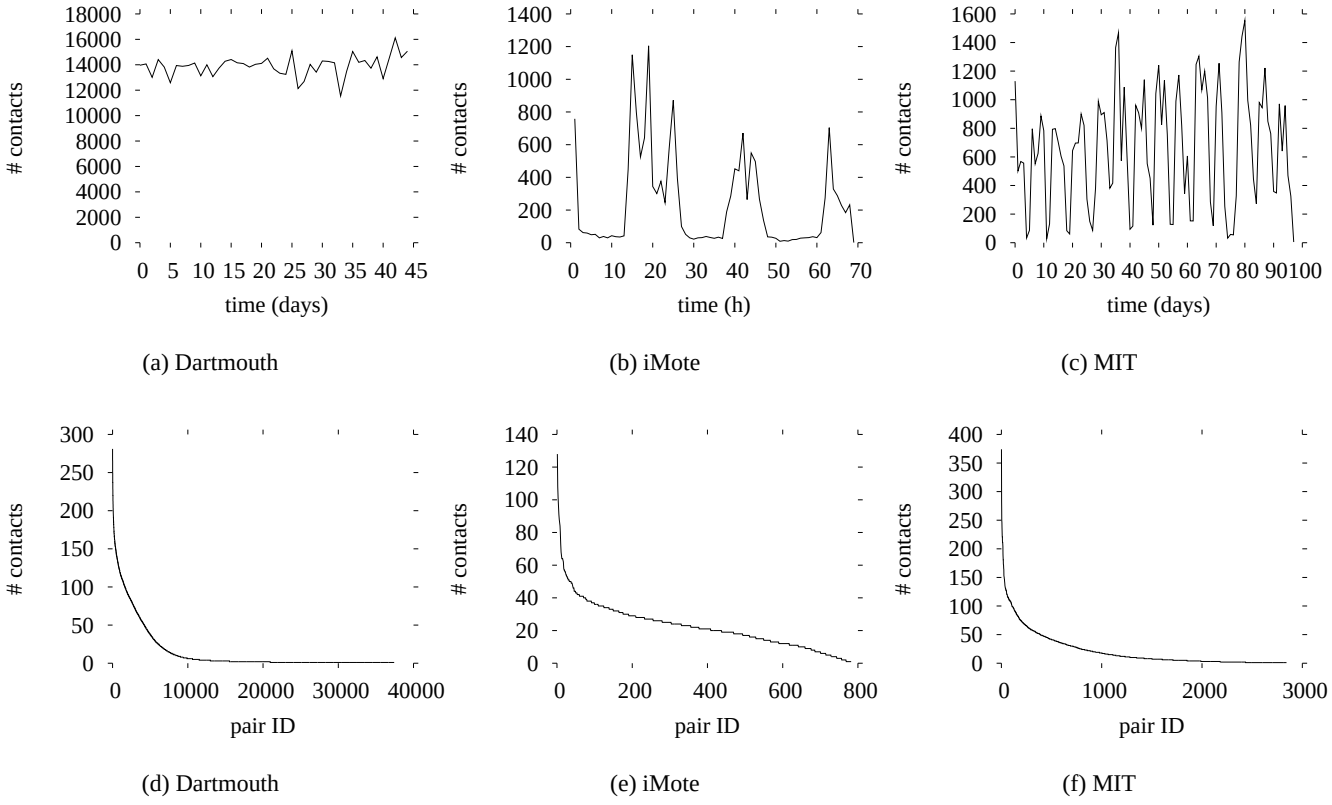
Chaintreau et al. [3] used iMotes (Bluetooth contact loggers from Intel) to acquire proximity contacts that occurred between participants in the student workshop at the *Infocom 2005* research conference. Students were asked to carry one of these sensors in their pocket at all times. Due to Bluetooth's 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. Note that, as with the Dartmouth data, many contacts might be missed. Those that occur between the 2 minute scans are not registered, and two nodes that are scanning simultaneously will not see each other.

In this data set, the evolution of the number of contacts between participants shows diurnal variations, as seen in Fig. 1(b). We observed 231.7 contacts per hour on average, with a standard deviation of 281.3 contacts. Fig. 1(e) only plots node pairs that had contacts, but these represent fully 95.4% of the pairs. For these pairs, the mean number of contacts was 22.8, with a standard deviation of 14.8 contacts. The iMote data shows more contacts for the typical node pair than does the Dartmouth data.

MIT data

The Reality Mining experiment [5] conducted at MIT captured proximity, location, and activity information from 97 subjects (mainly students) 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. Locality information comes from knowing which GSM network cell the phone is attached to. We only make use of the Bluetooth proximity data to determine whether two nodes were in contact. We selected 95 days of data corresponding to the first semester of the academic year 2004-2005 where activity was high in the traces in terms of the number of phones that collected data and the number of contacts that were recorded.

Fig. 1(c) displays weekly variations in the number of contacts between participants. The mean number of contacts per day is 660.0 contacts per day, with a standard deviation of 405.072 contacts. The number of interactions is lower than in the iMote data set, where the mean number of contacts per day was 1,378.39, and that was among only 41 nodes. For the 60.4% of node pairs that had at least one contact, and are plotted in Fig. 1(f), there is a mean of 22.3 contacts, with a standard deviation of 32.8 contacts. This plot more closely resembles the plot for the Dartmouth data set than the iMotes data set.



(a) Dartmouth (b) iMote (c) MIT

(d) Dartmouth (e) iMote (f) MIT

Figure 1: Evolution of the total number of contacts over time (top). Number of contacts for each pair of nodes (bottom); pairs are sorted in decreasing order of their number of contacts.

These data sets represent three different DTN scenarios which are of interest for the understanding of interactions between people that might carry communication devices. We will refer to these data sets as *Dartmouth*, *iMote* and *MIT*.

2.2 Heterogeneity in expectations

This section looks at the durations of contacts between pairs of nodes (*contact times*) and the time that elapses between any such contacts (*inter-contact times*). Both parameters influence performances of DTN schemes. Inter-contact patterns characterize the overall network topology and identifies message forwarding opportunities. It has a direct impact on opportunistic routing decisions. Contact time, along with other individual node characteristics such as memory, disk or power, adds constraints that further impact in particular the capacity of the DTN. This study focuses on heterogeneity, looking at the distributions for all node pairs. In the data sets just described in Sec. 2.1, we have already observed heterogeneity in the number of contacts per node pair. However, a deeper look is required to understand the impact of contact patterns on routing.

Fig. 2 shows, on the top row, the complementary cumulative distribution functions (CCDF), for all node pairs, of mean inter-contact times. We denote with $E(\tau)$ the expectation of inter-contact times, with τ being the process of inter-contact times for a given pair. Similarly, Fig. 2 shows, on the bottom row, the CCDF of $E(\Omega)$, 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.

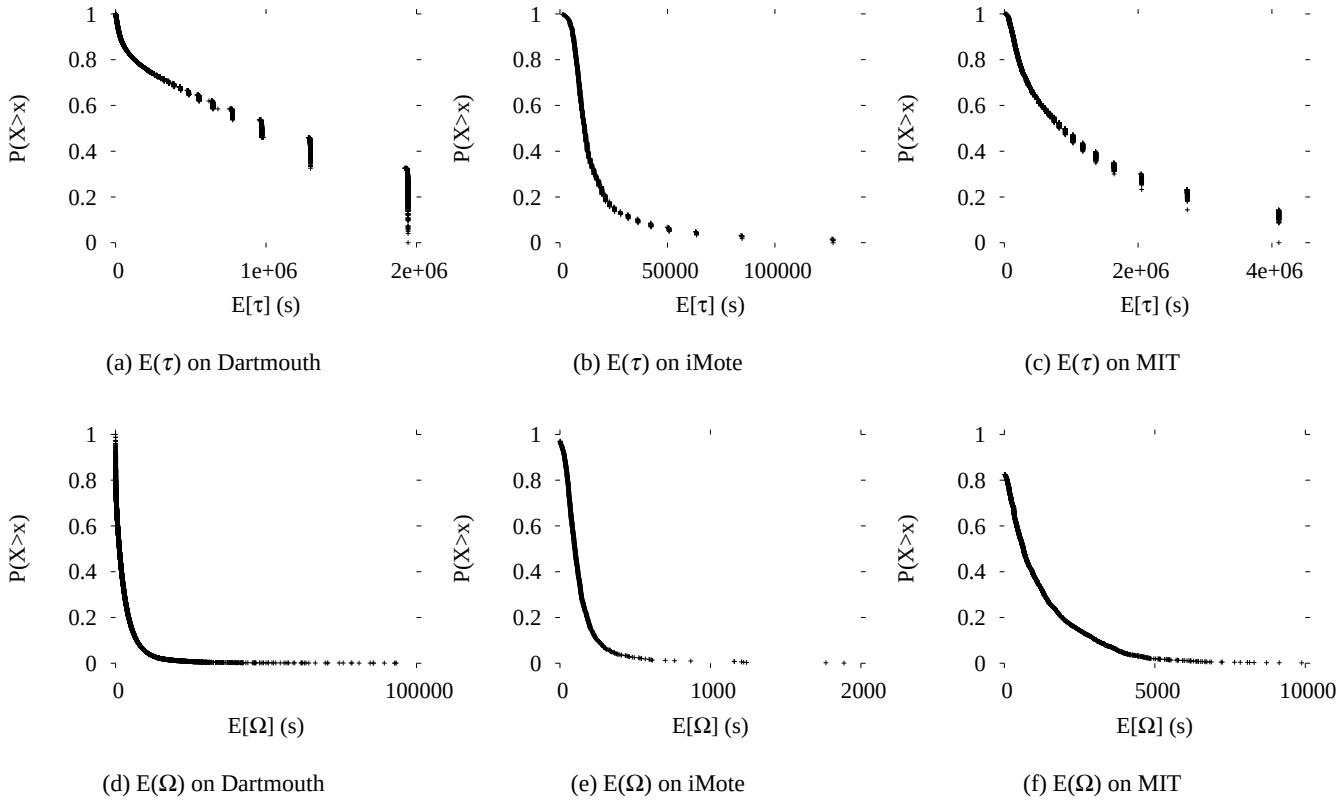
The mean contact times are also heterogeneous, as shown by plots in the right column of Fig. 2. The mean expected contact times are: 0.8 hours for Dartmouth, with a standard deviation of 3.0 hours; 0.03 hours for iMote, with a standard deviation of 0.04 hours; and 0.3 hours for MIT, with a standard deviation of 0.4 hours.

These results demonstrate that pairwise contact and inter-contact times processes should not be considered homogeneous in DTN models as they plot a high level of heterogeneity in expectations when looking at real data. We also observe, for all three data sets, that mean contact times are much shorter than inter-contact times. This leads us to conjecture that understanding inter-contact times processes is more crucial that understanding contact times processes if one needs to choose which to focus on.

2.3 Nature of inter-contact times distributions

Concentrating on aggregate inter-contacts provides very compact descriptors of the overall network behavior. It is summarized in a single distribution (e.g. Pareto) and the estimates of its parameters (in the Pareto case, two parameters). Although being very synthetic, using such information for DTN modeling might lead to erroneous conclusions. As a consequence, in order to better understand node interactions in DTN, we will be looking in this section at characterizing pairwise inter-contact distributions.

In a network of n nodes, there are $n(n-1)/2$ inter-contacts dis-



(a) E(τ) on Dartmouth (b) E(τ) on iMote (c) E(τ) on MIT

(d) E(Ω) on Dartmouth (e) E(Ω) on iMote (f) E(Ω) on MIT

Figure 2: CCDF of mean inter-contact times E(τ) (top) and mean contact times E(Ω) (bottom).

tributions. Fitting a different distribution for each pair, along with its parameters, may lead to intractable models. In order to keep a model of manageable size, we will be looking at families of distributions that can best capture pairwise behaviors. In the sought model, a single family (e.g. exponential, Pareto or log-normal) governs all pairwise inter-contacts, variations between individual pairs being summarized by different parameter values.

To this end, and as a first step, 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 use the Cramer-Smirnov-Von-Mises [4] statistical hypothesis test. Recall that such a statistical test can only *reject* or *fail to reject* a given hypothesis. So, when the hypothetical distribution is rejected by the test, we are certain that the distribution computed over the data does not match. On the other hand, when the test fails to reject the hypothesis, we only know that this is true to a confidence level $1-\alpha$. We used a relatively high level of confidence ($\alpha = 0.01$) and also visually cross-checked the goodness of fits.

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 hypothesised cumulative distribution functions (CDFs), $F_{ij}(x) = P(T_{ij} < x)$, given by the three following formulas:

- Exponential distribution: $F_{ij}(x) = 1 - e^{-\lambda_{ij}x}$

  λ_{ij} is the constant decay rate of the exponential distribution characteristic of a light tail behavior.

- Pareto distribution: $F_{ij}(x) = 1 - \left(\frac{x_{\mathrm{m}_{ij}}}{x}\right)^{k_{ij}}$

  k_{ij} is the shape parameter of the Pareto distribution and represents the degree of the power-tail (the slope of the linear decay of its CCDF on a log-log plot).

- Log-normal distribution: $F_{ij}(x) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left[\frac{\ln(x)-\mu_{ij}}{\sigma_{ij}\sqrt{2}}\right]$

  This corresponds to a variable whose logarithm follows a Normal (i.e. Gaussian) distribution with parameters μ_{ij} and σ_{ij}. The distribution is light-tailed and decays exponentially.

Note that, for a given node pair, several distributions may fit the inter-contact distribution. We see an example in Fig. 3 of the inter-contact times for an iMote node pair. These inter-contact times are found to follow a log-normal distribution.

Table 1 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. The main reason is that the log-normal distribution offers a more versatile model to capture the variability in inter-contact patterns across the different pairs of nodes. Almost no pair of nodes has been found to fit only an exponential or a Pareto distribution. 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.

Fig. 4 plots the distribution of the σ parameters of the log-normal in the three data sets. This parameter governs the shape of the

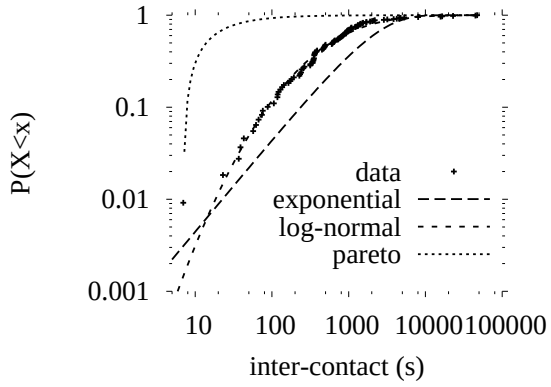


Figure 3: CDF in log-log scale of inter-contacts and their corresponding fits for a given node pair in iMote data.

	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 1: Fitting results.

log-normal distribution. For small values of σ the log-normal distribution is bell shaped around its mean e^{μ}. Indeed its skewness $(2+e^{\sigma^2})\sqrt{e^{\sigma^2}-1}$, becomes, for $\sigma << 1$, equivalent to 3σ, which, being close to 0, indicates a fairly symmetric distribution. For large values of σ the distribution becomes very skewed. Its unique mode is given by $e^{\mu-\sigma^2}$, so that, for $\sigma >> \mu$, the mode gets close to zero and appears as a vertical asymptote around the origin.

In other words, the log-normal family of distributions is capable of modelling all types of behaviors of the CDF around the origin, from smooth horizontal asymptotes for small values of σ, to nearly vertical asymptotes for larger values. In our data sets many samples exhibit vertical asymptotes around the origin, which translates into σ values: they are higher in Dartmouth, with an average of 3.5, compared to 2.2 for iMote and 2.1 for MIT. Both the exponential and Pareto have a linear behavior around the origin so the ability of the log-normal distribution to cope with vertical asymptotes near the origin is a clear advantage.

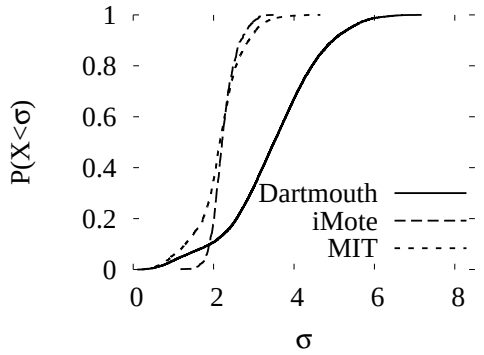


Figure 4: CDF of σ_{ij} for log-normal node pairs in data.

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 over a certain range, while always maintaining a finite expectation and variance (contrary to power laws with degrees lower than 2).

As we have examined only three data sets, albeit often-used ones, we cannot draw firm general conclusions about what will be revealed elsewhere. But one might reasonably expect that other mobility traces captured in similar environments will show similar characteristics.

3. THE POWER LAW PARADOX

In this section we wish to investigate and understand better the interplay between pairwise inter-contacts and aggregate inter-contacts.

Let's first first introduce why it seems important to make this difference: Chaintreau et al. [3] report that aggregated inter-contact times follow power laws in a number of DTN traces (including ones based on Dartmouth and iMote data). At the same time, we have just ruled out a power-law distribution to model pairwise inter-contacts in favor of the log-normal which exhibits a light-tail (with exponential decay). And this holds for traces based on the same Dartmouth and iMote data. This constitutes what we called the *power law paradox*.

Computing the cumulative distribution of aggregated inter-contact times for the Dartmouth data set confirms this observation. The log-log plot in Fig. 5 shows that it follows a power law of the form $f(x) = cx^{\delta}$, with exponent $\delta = -0.16$ and scale parameter $c = 3.45$.

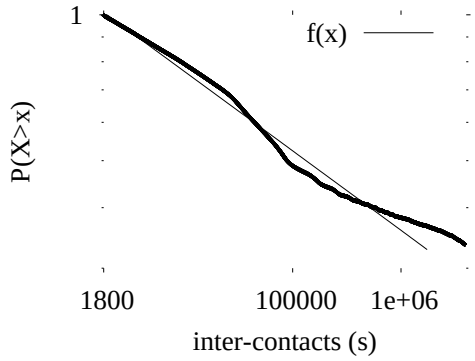


Figure 5: Log-log plot of the CCDF of aggregate inter-contacts in Dartmouth.

This section studies analytically this power law paradox by looking at the relationship between pairwise inter-contacts and inter-contacts aggregated over the entire set of pairs of nodes. The main purpose is to identify conditions in which aggregating different pairwise inter-contacts may lead to power-tailed distributions. We show how specific combinations of heterogeneous pairwise inter-contact times, both for the exponential and log-normal cases, can explain the aggregated power laws that have often been observed in experimental data sets.

First, we introduce a generic model of a heterogeneous DTN. The model supports any inter-contact time distribution. To analyze the power law paradox, we study two special cases which correspond to the exponential and log-normal distribution families. This

leads to explicit analytical formulas that provide insight into the phenomenon at play.

3.1 Heterogeneous DTN Model

Let us first consider a generic model for a heterogeneous DTN composed of n nodes. All nodes are given a unique ID in $1, 2, ..., n$. For any two nodes (i, j) we denote as $(t_{ij}^{(n)})_{n\in Z} = ... < t_{ij}^{-1} < t_{ij}^{0} < t_{ij}^{1} < ...$ the sequence of time instants at which a contact between i and j occurs.

The inter-contact pattern of the network is defined in the following way:

- For each pair of nodes (i, j) the pairwise inter-contact sequence $(t_{ij}^{(n)})_{n\in Z}$ is a renewal process; in other words, inter-contact times between nodes i and j are independent identically distributed *(iid)* random variables, let's say T_{ij}. Note that for each pair (i, j) the distribution law of T_{ij} and its parameters may differ.
- A given joint distribution of the $n(n-1)/2$ pairwise inter-contact sequences serves in particular to characterize the possible correlations between two inter-contact sequences.

In this section we are interested in looking at variable Θ, the inter-contact times aggregated for all pairs of nodes in the network. By definition the aggregated inter-contact time is greater than t, $t > 0$, if the inter-contact time is greater than t for at least one pair of nodes. Formally, this means that $\Theta > t$ iff $\exists(i, j), i < j, T_{ij} > t$, which means that $\Theta = \sup_{i<j} T_{ij}$.

The major hypothesis made by the model is that inter-contact time distributions are stationary. The model focuses on the temporal dynamics of node connectivity in a DTN. It does not model node mobility directly, but captures inter-contact patterns. In this way it provides a common framework to analyse different DTNs. The traditional Random Way Point and Random Walk mobility models for ad-hoc networks fall in this category (see Carreras et al. [2]). More generally, the model would allow one to capture the different forms of heterogeneity that we have identified in the data sets of Sec. 2.

3.2 Aggregated Exponentials

In this section we address the exponential case. The exponential hypothesis is not the best fit for the data we analysed, but appears often as the second best choice (as one can see from Table 1). It constitutes also the extreme case – exponential decay being the very prototype of light tailed distributions – and the opposite of the power tailed behavior of the Pareto. We are then able to formally derive the aggregate distribution of inter-contact times in the case of exponential pairwise inter-contacts.

The heterogeneous exponential case corresponds to the model with the following complementary hypotheses: the pairwise contact sequences are homogeneous Poisson processes (HPPs), i.e., the inter-contact times follow exponential laws with parameters λ_{ij}. Furthermore the HPPs are independent. The purpose is here to focus on the heterogeneity of pairwise mean inter-contact times, which are given by $1/\lambda_{ij}$ and to study the effect of aggregating the inter-contact patterns. Choosing the exponential case may be seen as an extreme; the tail distribution of the exponential is the very opposite of the heavy tailed pattern that power laws capture best. Moreover, it may seem that the pairwise exponential assumption is too strong to yield a power law in the aggregate.

The key to resolving the paradox in the exponential case is given by a classical result by Bernstein which states that any completely monotone PDF can be obtained as the mixture of exponential PDF's (see for example [8]). Let's examine how this translates in our case.

Let Θ be the aggregate inter-contact time for all pairs of nodes. Let's write $K = n(n-1)/2$ and renumber the pairwise inter-contacts T_k from 1 to K. We then have $\Theta = \sup_{1\le k\le K} T_k$. Let's imagine that all inter-contact processes are exponentially distributed with various parameters λ_k. Different distributions of the λ_k parameters can model different global properties of DTNs. For example, in some cases, a node will meet most of the others several times a day, and the remaining ones on a weekly basis. Let p_k denote the proportion of pairs with parameter λ_k.

Conditioning on the event that in the aggregate $\sup_{1\le k\le K} T_k$ the pair is pair number k, we have:

$$P(\Theta > t) = \sum_{k=1}^{k=K} P(T_k > t) p_k \tag{1}$$

In the case of a DTN with a very large number of pairs (such as in the Dartmouth case) we consider the analogue formula with a continuously varying probability distribution $p(\lambda)$ of the λ parameters:

$$P(\Theta > t) = \int_{\lambda=0}^{\infty} e^{-\lambda t} p(\lambda) d\lambda \tag{2}$$

What Eq. 2 says is that, for the exponential case, the aggregate inter-contact time distribution is fully characterized by the distributions of the λ parameters. More precisely, the tail cumulative distribution of the aggregated inter-contact times is given by the Laplace transform of the probability density function (PDF) p of the λ parameters.

We would thus like to know if there exist cases of PDFs p that can generate power law tail behaviors in the aggregate. Let's consider the case of the Pareto law with shape parameter $\alpha > 0$ and scale parameter $b > 0$, which reads, for $t \ge 0$:

$$P(\Theta > t) = (\frac{b}{t+b})^{\alpha} \tag{3}$$

Since the Laplace transform is invertible, Eq. 2 tells us that taking the inverse Laplace transform of $P(\Theta > t)$ gives the distribution p of the λ parameters. We then have, Γ being the Gamma function, for $\lambda \ge 0$:

$$p(\lambda) = \frac{\lambda^{\alpha-1} b^{\alpha} e^{-b\lambda}}{\Gamma(\alpha)} \tag{4}$$

This provides an answer to our initial question: even if all pairwise inter-contacts follow an exponential distribution, it is still possible to regain the power law distribution in the aggregate. One could have thought *a priori* that it would require the distribution of the λ parameters to be power-tailed. In that case a power law would still have come into play, not directly in the pairwise inter-contacts, but at a global scale of the DTN (its distribution of parameters). In fact Eq. 4 shows that this is not necessary, since the tail of the Gamma distribution drops off exponentially.

What this analytical result shows is that when considering a DTN with independent pairwise exponential inter-contacts, one can regain the power law behavior for the aggregated inter-contacts when the distribution of the parameters is a Gamma.

Let us apply this result to the Dartmouth data set; recall that inter-contact patterns are not all exponential, so to confront the result to the data, we proceed in the following way: we estimate parameters α and b from the cumulative distribution of the λ parameters for pairs that were shown to follow an exponential behavior (the ones that "pass" the Cramer hypothesis test). We find $b = 113{,}766.9$ and

$\alpha = 2.26$. Fig. 6(a) shows the estimated cumulative gamma distribution $g(x)$ with the experimental lambda cumulative distribution for all pairs that have shown to be exponential. Then, we plot in Fig. 6(b) the corresponding power-law $h(t)$ with cumulative distribution of aggregated inter-contact times.

The fit of the Gamma distribution in Fig. 6(a) captures the shape of the distribution of parameters with a small underestimate for larger values of λ. The predicted Pareto distribution for the aggregated inter-contacts in Fig. 6(b) shows good fit for most of the values. The tail that the Pareto distribution generates appears however heavier than the actual inter-contacts which might be explain by the fact that aggregated inter-contacts do not fit perfectly a power-law (see Fig. 5).

3.3 Aggregated Log-normals

In this section we consider the same hypotheses for the DTN model as in the exponential case above, but we replace the individual exponential laws with parameters λ_{ij} by log-normals with parameters μ_{ij} and σ_{ij}. Similarly to the section above we would like study how different distributions of the parameters yield aggregate power law inter-contacts.

Limited analytical results exist on mixture of log-normals as they do not lead as easily as exponentials to simple closed form formulas. Montroll and Shlesinger [20] show that a geometric mixture of log-normals yields a power-law, and Reed [22] [19] show that stopping a log-normal multiplicative growth at a random exponential time yields a power-law distribution. None of these results deal with aggregate estimates nor do they apply directly. In this section we are going to show that it is possible, for an appropriate distribution of parameters, to regain here again a power-law from log-normal pairwise inter-contacts.

As for the exponential study, let's consider the continuous setting. Let's first consider the impact of the shape parameters, keeping the scale constant. Calling p_σ the PDF of the σ parameters, one can write a similar equation to Eq. 2 for the PDF $p_\Theta(t)$ of the aggregate Θ:

$$p_\Theta(t) = \int_{\sigma=0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}} p_\sigma(\sigma) d\sigma \tag{5}$$

Looking at the sigmoid shape of the σ distributions in the three data sets in Fig. 4, it seems reasonable to consider modeling them with a Weibull distribution, whose PDF p_W is given by:

$$p_W(t) = \frac{k}{\lambda}(\frac{t}{\lambda})^{k-1} e^{-(\frac{t}{\lambda})^k} \tag{6}$$

In the Weibull distribution λ is a scale parameter and k a shape parameter. We will now consider the specific case of $k = 2$. Eq. 5 becomes after simplification:

$$p_\Theta(t) = \frac{2}{\sqrt{2\pi}\lambda^2 t} \int_{\sigma=0}^{\infty} e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}} e^{-\frac{\sigma^2}{\lambda^2}} d\sigma \tag{7}$$

This is of the form $\int_{x=0}^{\infty} e^{-\frac{b}{x^2}} e^{-ax^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} e^{-2\sqrt{ab}}$ with $a = \frac{1}{\lambda^2}$ and $b = \frac{(\ln t-\mu)^2}{2}$, which yields:

$$p_\Theta(t) = \frac{1}{\sqrt{2}\lambda t} e^{-\frac{\sqrt{2}}{\lambda}|\ln t-\mu|} \tag{8}$$

p_Θ follows two different behaviors depending on whether $t \geq e^\mu$ or $t \leq e^\mu$.

Let's now allow the μ parameter to vary as a random variable independent of λ, with a PDF p_μ. Eq. 7 thus becomes, after simplification using Eq. 8:

$$p_\Theta(t) = \int_{\mu=-\infty}^{\infty} \frac{1}{\sqrt{2}\lambda t} e^{-\frac{\sqrt{2}}{\lambda}|\ln t-\mu|} p_\mu(\mu) d\mu \tag{9}$$

Since $t > 0$, we have:

$$p_\Theta(t) = \frac{1}{\sqrt{2}\lambda t}\Big(\int_{-\infty}^{\ln t} e^{-\frac{\sqrt{2}}{\lambda}(\ln t-\mu)} p_\mu(\mu) d\mu + \int_{\ln t}^{\infty} e^{-\frac{\sqrt{2}}{\lambda}(\mu-\ln t)} p_\mu(\mu) d\mu\Big) \tag{10}$$

and

$$p_\Theta(t) = \frac{1}{\sqrt{2}\lambda}\Big(\frac{1}{t^{1+\sqrt{2}/\lambda}} \int_{-\infty}^{\ln t} e^{\frac{\sqrt{2}}{\lambda}\mu} p_\mu(\mu) d\mu + \frac{1}{t^{1-\sqrt{2}/\lambda}} \int_{\ln t}^{\infty} e^{-\frac{\sqrt{2}}{\lambda}\mu} p_\mu(\mu) d\mu\Big) \tag{11}$$

Supposing that $\int_{-\infty}^{\infty} e^{\frac{\sqrt{2}}{\lambda}(\mu)} p_\mu(\mu) d\mu$ converges, let's show that the first term dominates as $t \to \infty$. Writing

$$r(t) = \int_{\ln t}^{\infty} e^{-\frac{\sqrt{2}}{\lambda}\mu} p_\mu(\mu) d\mu = \int_{\ln t}^{\infty} e^{\frac{-2\sqrt{2}}{\lambda}\mu} e^{\frac{\sqrt{2}}{\lambda}\mu} p_\mu(\mu) d\mu \tag{12}$$

And using the fact that function $\mu \mapsto e^{\frac{-2\sqrt{2}}{\lambda}\mu}$ is decreasing, one has:

$$0 \leq r(t) \leq t^{-\frac{2\sqrt{2}}{\lambda}} \int_{\ln t}^{\infty} e^{\frac{\sqrt{2}}{\lambda}(\mu)} p_\mu(\mu) d\mu \tag{13}$$

From Eq. 11, we have

$$\frac{p_\Theta(t)}{\frac{\sqrt{2}\lambda}{t^{1+\sqrt{2}/\lambda}}} = \int_{-\infty}^{\ln t} e^{\frac{\sqrt{2}}{\lambda}\mu} p_\mu(\mu) d\mu + t^{\frac{2\sqrt{2}}{\lambda}} r(t) \tag{14}$$

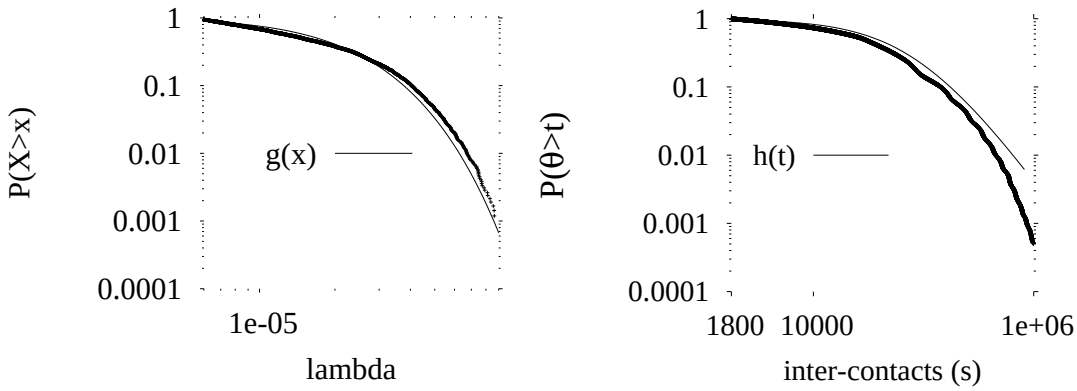
Using the hypothesis that $K = \int_{-\infty}^{\infty} e^{\frac{\sqrt{2}}{\lambda}(\mu)} p_\mu(\mu) d\mu < \infty$, one can see that the first term in Eq. 14 converges to K as $t \to \infty$. Using also Eq. 13, one sees that the second term goes to 0 as $t \to \infty$. With this hypothesis we have:

$$p_\Theta(t) \sim C_2 t^{-1-C_1} \tag{15}$$

with $C_2 = K\sqrt{2}/\lambda$ and $C_1 = \sqrt{2}/\lambda$ which shows that p_Θ then follows a power law. The condition that we use is that the integral $\int_{-\infty}^{\infty} e^{\frac{\sqrt{2}}{\lambda}(\mu)} p_\mu(\mu) d\mu$ converges. For instance, if μ follows a distribution with an exponentially decreasing tail with parameter ν, the condition is ensured as soon as $\nu < \sqrt{2}/\lambda$, where λ is the scale parameter for the distribution of the shape parameters of the log-normal family.

The important lesson that we can draw is that it is not necessary to introduce pairwise power laws to generate aggregate power laws. This is good news for routing in DTNs. Chaintreau et al. [3] have reported that pairwise power-laws (in particular for degrees lower than 2) have an adverse impact on the opportunistic Spray and Wait routing strategy, casting some doubt on the mere possibility of efficient finite-time delivery of messages across a network. With pairwise inter-contacts with both first moments (mean and variance) finite, the picture looks much brighter.

Fitting a Weibull distribution to the shape parameter distributions in the data sets gives values of $k = 3.50356$ for Dartmouth, $k = 4.88741$ for MIT and $k = 8.903$ for iMote. In these data sets the shape parameters are 2 to 5 times greater than 2. The power law result that we demonstrated thus does not apply directly in these cases. This may result from the fact that the data sets under study follow a power-law behavior only on a finite range (although covering several orders of magnitude), but that their asymptotic behavior



(a) λ_{ij} (b) Inter contacts

Figure 6: Log-log plot of the CCDF of aggregate inter-contacts with exponential pairs in Dartmouth.

is not a power-law and would be consistent with similar observations reported by Karagiannis et al. [15].

4. DISCUSSION AND RELATED WORK

The data sets may represent partial or biased real life interactions as sampling methods were used for their collection. The iMote and MIT data sets have been collected using periodic Bluetooth scans which may have underestimated the overall number of contacts or the contact times between nodes. In Dartmouth, the two main factors coming into play (see Sec. 2) are: 1) we infer that two people are in contact whenever they are connected to the same AP which might create unrealistic interactions, 2) mobility of laptops is not really representative of human mobility. As a consequence, one has to take carefully these results into account.

Power-law and log-normal distributions are closely related as shown by Mitzenmacher [19]. Choosing one or the other from empirical data only is often difficult. A priori or external information (derived from the underlying phenomena at play, whether coming from physics, biology, geology or other) may be needed to guide statistical inference. Still we have found that the log-normal family offers a synthetic model (with only two sets of parameters) that allows to summarize a large proportion of pairwise inter-contacts. It is the versatility of the proposed family that can make it an attractive model for capturing heterogeneity in inter-contact patterns.

Much ongoing research tries to understand and characterize the mobility patterns in DTNs and mobile ad-hoc networks (MANETs). Due to the limited number of data sets available and the fact that they are generally specific to a scenario, studies often resort to synthetic models. Models such as Random Walk (i.e., Brownian motion [6]) and Random Way-Point [13] have been very popular [10, 23]. More recent work has extended these initial models with proposals to better match patterns observed in real mobility data. Musolesi et al. [21] propose a model in which the movements of nodes are driven by social relationships. Bohacek [16] designed a mobility model of individuals in urban settings based on a recent US Bureau of Labor Statistics time-use study. Legendre et al. [17] question whether microscopic mobility behaviors are valuable to represent mobility with more realness and their influence on important characteristics (e.g., link duration distribution). Francois et al. [9] proposes a framework for formalizing the behavior contact patterns in situations in which each node knows the probability distributions for its contacts with other nodes. Carreras et al. [2] propose a graph-based model able to capture the evolution of the connectivity between nodes over time.

The approach taken in our work is rather to put the stress on pairwise inter-contact patterns as one of the key enablers for the design of routing algorithms in DTNs. This paper is the first to provide a detailed analysis of the pairwise inter-contacts in a number of DTN data sets, and the first to identify the log-normal family of distributions as a promising modeling candidate.

5. CONCLUSION AND FUTURE WORK

In this paper, we argue for the wisdom of using pairwise inter-contact patterns to characterize DTNs. We have first 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. Second, we describe the power-law paradox and investigate analytically the relationship between pairwise and aggregate inter-contact times. In particular, we consider both the exponential and log-normal cases and show analytically how the aggregation of pairwise inter-contacts may lead to aggregate inter-contacts with power laws of various degrees.

Future work along these lines might include studies of correlations between processes, and of short and long term dependencies in DTN data sets. Work on modeling has also to be conducted to provide workable DTN models that integrate heterogeneity in pairwise interaction processes and the use of distributions such as the log-normal.

Acknowledgments

This work was supported by LiP6 and Thales Communications through the laboratory Euronetlab, and the ANRT which provided the CIFRE grant 135/2004. This work was also supported by the French research project RNRT AIRNET. We also thank Augustin Chaintreau for his valuable comments.

6. REFERENCES

[1] Crawdad: A community resource for archiving wireless data at dartmouth. `http://crawdad.cs.dartmouth.edu`.

[2] I. Carreras, D. Miorandi, and I. Chlamtac. A framework for opportunistic forwarding in disconnected networks. In *Proc. of MOBIQUITOUS*, 2006.

[3] 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.

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

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

[6] A. Einstein. *Investigations on the Theory of the Brownian Movement*. Dover Publications, 1906.

[7] K. Fall. A delay-tolerant network architecture for challenged internets. In *Proc. SIGCOMM*, 2003.

[8] W. Feller. *An Introduction to Probability Theory and Its Applications, vol.2*. Wiley & Sons, 1971.

[9] J.-M. Francois and G. Leduc. Predictable disruption tolerant networks and delivery guarantees. Technical Report cs.NI/0612034, Arxiv, 2006.

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

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

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

[13] D. B. Johnson and D. A. Maltz. Dynamic source routing in ad hoc wireless networks. In Imielinski and Korth, editors, *Mobile Computing*, volume 353. Kluwer Academic Publishers, 1996.

[14] E. P. C. Jones, L. Li, and P. A. S. Ward. Practical routing in delay-tolerant networks. In *Proc. WDTN*, 2005.

[15] 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.

[16] J. Kim and S. Bohacek. A survey-based mobility model of people for simulation of urban mesh networks. In *Proc. MeshNets*, 2005.

[17] F. Legendre, V. Borrel, M. Amorim, and S. Fdida. Reconsider microscopic mobility modeling for self-organizing networks. *IEEE Communications Magazine*, December 2006.

[18] J. Leguay, T. Friedman, and V. Conan. Evaluating mobility pattern space routing for DTNs. In *Proc. INFOCOM*, 2006.

[19] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2004.

[20] E. M. Montroll and M. F. Shlesinger. On 1/f noise and other distributions with long tails. In *Proc. of the National Academy of Sciences 79*, 1982.

[21] M. Musolesi and C. Mascolo. A Community based Mobility Model for Ad Hoc Network Research. In *Proc. REALMAN*, 2006.

[22] W. J. Reed. The pareto law of incomes - an explanation and an extension. *Physica A*, (319):469–486, 2003.

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

[24] L. Song, D. Kotz, R. Jain, and X. He. Evaluating location predictors with extensive Wi-Fi mobility data. In *Proc. Infocom*, 2004.

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

[26] UMassDieselNet. A Bus-based Disruption Tolerant Network, `http://prisms.cs.umass.edu/diesel/`.

[27] J. Yoon, B. Noble, M. Liu, and M. Kim. Building realistic mobility models from coarse-grained traces. In *Proc. MobiSys*, 2006.