

Power law and exponential decay of
inter contact times between mobile devices¹

Thomas Karagiannis, Jean-Yves Le Boudec, Milan Vojnović

March 2007

Technical Report
MSR-TR-2007-24

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
<http://www.research.microsoft.com>

¹Alphabetical author order. Author affiliations: T. Karagiannis and M. Vojnović are with Microsoft Research Cambridge, UK. J.-Y. Le Boudec is with EPFL, Switzerland. Contact email: milanv@microsoft.com.

Abstract— We examine the fundamental properties that determine the basic performance metrics for opportunistic communications. We first consider the distribution of inter-contact times between mobile devices. Using a diverse set of measured mobility traces, we find as an invariant property that there is a characteristic time, order of half a day, beyond which the distribution decays exponentially. Up to this value, the distribution in many cases follows a power law, as shown in recent work. This power law finding was previously used to support the hypothesis that inter-contact time has a power law tail, and that common mobility models are not adequate. However, we observe that the time scale of interest for opportunistic forwarding may be of the same order as the characteristic time, and thus the exponential tail is important. We further show that already simple models such as random walk and random waypoint can exhibit the same dichotomy in the distribution of inter-contact times as in empirical traces. Finally, we perform an extensive analysis of several properties of human mobility patterns across several dimensions, and we present empirical evidence that the *return time* of a mobile device to its favorite location site may already explain the observed dichotomy. Our findings suggest that existing results on the performance of forwarding schemes based on power-law tails might be overly pessimistic.

Keywords—opportunistic communications, mobile ad-hoc networking, wireless networking, serendipitous communications, information dissemination, inter-contact time distribution, power law

1. INTRODUCTION

Over the past year, empirical studies have provided evidence suggesting that power laws characterize diverse aspects of human mobility patterns, such as inter-contact times, contact, and pause durations. These studies are of high practical importance for (a) informed decisions in protocol design, and (b) realistic mobility models for protocol performance evaluation.

Specifically, Chaintreau et al [2] were perhaps the first to report credible empirical evidence suggesting that the CCDF (complementary cumulative distribution function) of inter-contact time between human-carried mobile devices follows a power law over a wide range of values that span the timescales of a few minutes to half a day. This empirical finding has motivated Chaintreau et al to pose the hypothesis that inter-contact time has a CCDF with power law *tail*. Under this assumption, they derived some interesting results on feasibility and performance of opportunistic forwarding algorithms. In particular, their hypothesis implies that for any forwarding scheme the mean packet delay is *infinite*, if the power-law exponent of the inter-contact time is smaller than or equal to 1 (the case suggested to hold in prac-

tice by the empirical results so far). These results are in sharp contrast with previously known findings on similar packet forwarding algorithms (e.g. Grossglauser and Tse [7]) which were obtained under a hypothesis of exponentially decaying CCDF of inter-contact time. Furthermore, the authors argued that the power-law tail is not supported by common mobility models (e.g. random waypoint [9]), thus suggesting a need for new models.

In this paper, we find that the CCDF of inter-contact time between mobile devices features a *dichotomy* described as follows. On the one hand, in many cases the CCDF of inter-contact time follows closely a power-law decay up to a characteristic time, which confirms earlier studies. On the other hand, beyond this characteristic time, we find that the decay is exponential. This exponential decay appears to be a new finding, which we validate across a diverse set of mobility traces. The dichotomy has important implications on the performance of opportunistic forwarding algorithms and implies that recent statements on performance of such algorithms may be over-pessimistic.

We further provide analytical results showing that simple mobility models such as simple random walk on a circuit (one-dimensional version of the Manhattan Street Network model dating from the 80's [12] and used recently [6, 1]) and random waypoint [9] on a chain can exhibit the same qualitative properties observed in empirical traces. Whilst our results do not suggest that the considered mobility models are sufficient for realistic simulations, they stress that existing models should not be discarded on the basis of not supporting the empirically observed dichotomy of inter-contact time.

To understand the origins of the observed dichotomy, we then examine several properties of device contacts across various dimensions. We provide empirical evidence suggesting that the return time of a mobile device to its favorite location site features the same dichotomy as the one observed for the inter-contact time between device pairs. This is an interesting hypothesis as it refers to the return time of a device to a site, which is a more elementary characterization of human mobility than that of the inter-contact times. Also, it is a particularly important metric for cases where some devices have fixed locations (e.g. throwboxes [19]). It is also noteworthy that we established the same qualitative equivalence between return time to a site and inter-contact time for our simple random walk model. Further, our empirical analysis suggests that mobile devices are typically in contact in a few sites which are specific to the given pair of devices. Combining the latter with the observed dichotomy of the return time to a site may already explain the inter-contact time dichotomy in real settings.

We also investigate different viewpoints on device in-

Table 1: Traces studied.

Name	Technology	Duration	Devices	Contacts	Mean Inter-contact Time	Year
UCSD	WiFi	77 days	275	116,383	24 hours	2002
Vehicular	GPS	6 months	196	9,588	20.8 hours	2004
MITcell	GSM	16 months	89	1,891,024	3.5 hours	2004
MITbt	Bluetooth	16 months	89	114,046	87 hours	2004
Cambridge	Bluetooth	11.5 days	36	21,203	14 hours	2005
Infocom	Bluetooth	3 days	41	28,216	3.3 hours	2005

ter contacts such as that of a specific device pair and the random observer viewpoint. The examination of these viewpoints are of interest in order to evaluate how representative the CCDF of inter-contact time is, which following earlier studies, is defined as the CCDF of inter-contact time samples aggregated over all device pairs over a measurement period. We argue that the aggregate CCDF of inter-contact time when used to derive residual time until a contact, refers to a viewpoint that corresponds to a random observer and for a device pair picked uniformly at random. Finally, we also report on various breakdowns of the aggregate behavior across pairs to investigate time nonstationarity due to the synchronization of time-of-day human activities. Overall our results provide valuable insights towards the design of opportunistic packet forwarding schemes. Our contributions can be summarized in the following points:

- Using 6 distinct traces, we verify the power-law decay of inter-contact time CCDF between mobile devices. The analyzed datasets range from campus-wide logs of device associations to existing infrastructure, to direct contact bluetooth traces and GPS logs of vehicle movements over several months.

- We demonstrate that beyond a characteristic time of the order of half a day which is present across all datasets the CCDF exhibits exponential decay. (Section 3.) The observed dichotomy and exponential decay appear to be new results.

- We provide analytical results that show that already simple mobility models can exhibit precisely the same qualitative dichotomy. (Section 4.) These results contradict statements that current mobility models cannot support power law CCDF of inter-contact time.

- We examine several dimensions of device contacts in order to better understand the observed dichotomy, and we study the time nonstationarity due to human synchrony with the time of day. In particular, we provide evidence that there exist real-world cases in which return time of a mobile device to its frequently visited site follows the same dichotomy as that of the inter-contact time between devices. (Section 5.)

Implications to a practitioner. Our findings suggest that: (a) The issue of infinite expected packet for-

warding delay of opportunistic forwarding schemes under the power tail assumption [2] does not appear relevant with an exponentially decaying tail of the inter-contact CCDF. The exponential decay beyond the characteristic time is of relevance as available data traces suggest that the mean inter-contact time is in many cases of the same order as the characteristic time. (b) Widely-used mobility models should not be abandoned on the claim that they cannot support power-law decay of the inter-contact CCDF. Finally, (c) our analysis implies that time nonstationarity and potential environmental or other idiosyncracies that may influence human behavior (e.g., conference sites vs. working environments) should be taken into account during the design of future systems as they can significantly affect its primary performance metrics.

2. DATASETS & DEFINITIONS

2.1 Datasets

To study the properties of contacts between human-carried mobile devices, we analyze several traces with diverse characteristics in terms of their duration, wireless technology used and environment of collection. Table 1 presents a summary of the different datasets that we used, with aggregate statistics regarding the total duration of the trace, the number of mobile devices, the number of contacts and the mean inter-contact time. We can group the datasets in three distinct types:

- *Infrastructure-based* traces that reflect connectivity between existing infrastructure, e.g., Access Points (APs) or cells, and wireless mobile devices (UCSD [13] & MITcell [5, 4] datasets in Table 1). These datasets describe association times of a specific mobile device with an AP or cell.

- *Direct contact* traces that record contacts directly between mobile devices (e.g., imotes) and were collected by distributing devices to a number of people, usually students or conference attendees (Cambridge [2, 14], Infocom [2, 15] & MITbt [5, 4] datasets in Table 1). These datasets describe start and end contact times for each pair of mobile devices.

- *GPS-based* contacts through a private trace collected by tracking the movements of individual people

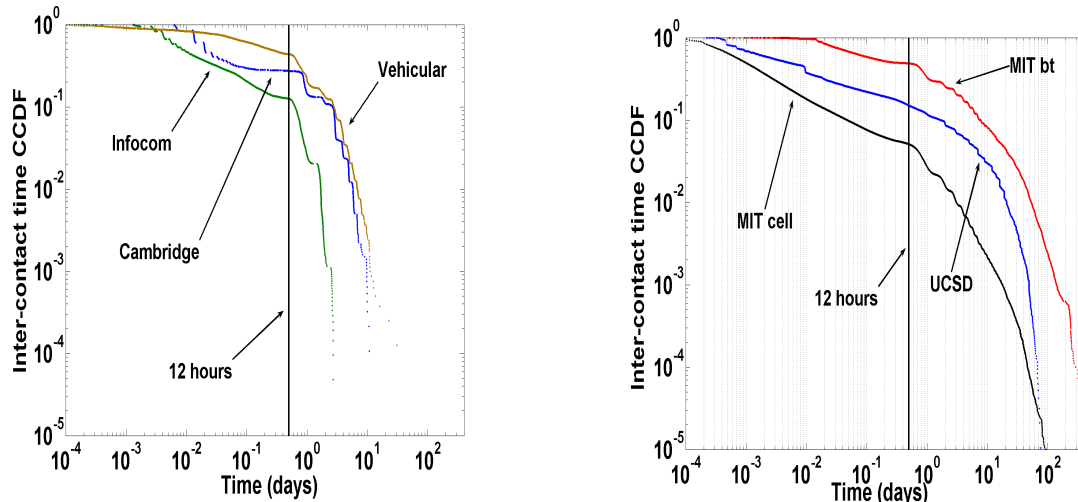


Figure 1: Inter-contact time CCDF for the six datasets.

of a large corporation through GPS units. The GPS units were placed in volunteer’s cars for approximately four months and overall the trace covers the metropolitan area of a large US city (Vehicular dataset [10] in Table 1). The dataset logs the latitude and longitude coordinates of each mobile device every approximately 10 seconds.

Note that the first two types of traces have been used in previous studies (e.g., [2]), while the latter which features exact mobility patterns based on the GPS trackers is unique to our study. Due to space limitations, we refer the interested reader to the references provided for a complete description of the experimental setting, the devices used and the limitations of each trace.

Apart from the datasets specifying direct connectivity, contacts need to be inferred in the rest of the traces. For the infrastructure-based traces, we assume that two devices are in contact if they reside within the range of the same AP or cell in accordance with previous studies. For the vehicular trace, we assume that two mobile devices are in contact if their distance is less than or equal to a parameter r . For our experiments we chose $r = 500$ meters [8], while we experimented with values from 100m to 1km and found qualitatively similar results.

Throughout the paper we will use all datasets interchangeably. Unless otherwise specified, our observations apply to all traces listed in Table 1.

2.2 Definitions

We use the following definitions. An *inter-contact time* between two devices is defined as the length of the time interval over which the two devices are not in contact and are in contact at the end points of this

interval. For a device pair, we call *residual inter-contact time*, the time until the next contact of this device pair from a given observation time. A *return time* of a device to a set of a space is defined as the minimum time until the device enters the set, from a time instance at which the device exited the set.

We call *CCDF of inter-contact time between two devices*, the CCDF obtained for the inter-contact time sampled per contact of these two devices. We further call the *aggregate CCDF* of inter-contact time between *all* devices, the CCDF of per contact samples of inter-contact time over all distinct pairs of devices. For simplicity, we often abuse this notation by omitting explicitly to mention the “aggregate” but the meaning should be clear. Finally, we consider the *CCDF of the residual inter-contact time* at a specific observation time, defined for a value $t \geq 0$ as the fraction of device pairs for which the residual inter-contact time at the observation time is larger than t .

3. INTER-CONTACT TIME DICHOTOMY

In this section, we examine the empirical distributions of inter-contact times between mobile devices inferred from the mobility traces introduced in the previous section. We have carefully examined all datasets in Table 1 and confirmed the hypothesis that in many cases the aggregate CCDF of the inter-contact times follows a power-law up to a characteristic time. We find this time to be in the order of half a day. Note that this hypothesis was already tested in previous work [2].

However, we demonstrate here that beyond this characteristic time, the CCDF exhibits an exponential decay. To the best of our knowledge, the hypothesis that the CCDF of the inter-contact times beyond the char-

characteristic time exhibits an exponential decay has been neither posed nor tested before. We then argue that the exponential decay is an important property since it bears significant impact on the mean inter-contact time or more generally on the CCDF of the inter-contact time observed from an arbitrary point in time. Finally, we discuss the practical implications of the observed dichotomy to opportunistic forwarding.

3.1 Power law and exponential decay

In this section, we provide empirical evidence of a dichotomy in the CCDF of inter-contact time. Up to a characteristic time in the order of half a day, the decay of the CCDF is well approximated as a power law, while beyond this characteristic time, the decay is exponential.

Power law. We first revisit the power law hypothesis in the examined datasets. To this end, we have inferred the inter-contact time for each of the traces and estimated the aggregate CCDF of inter-contact time between all devices. Fig. 1 shows the respective aggregate CCDFs of inter-contact times in log-log scale. The CCDF values follow a straight line over a range of values spanning the order of a few minutes to half a day, thus suggesting a power law. These results are in line with observations of previous studies for datasets MIT, UCSD, Cambridge and Infocom. A new piece of information is however that the same property holds for the vehicular trace which is significantly different in nature from the rest of the datasets.

Exponential decay. Carefully examining Fig. 1, we observe that at roughly around half a day, the CCDF has a knee beyond which the decay is abruptly faster. We call this knee *the characteristic time*. In order to examine the CCDF of inter-contact time beyond the characteristic time, we replot the same curves of Fig. 1 in Fig. 2 but this time in lin-log scale. We now turn our attention to the distributions beyond the characteristic time. In the lin-log scale (Fig. 2), the CCDF can be closely upper bounded with a straight line, thus indicating an exponential decay. In some traces, e.g., for Infocom and Cambridge, we also observe some variability in the tail that after close examination we found to be in line with daily periodicities (24 hours).

3.2 Implications of dichotomy to opportunistic forwarding

How does the observed dichotomy in the distribution of inter-contact time affect the design of forwarding schemes?

Motivated by the observed power law in the empirical aggregate CCDF of inter-contact time up to half a day, Chaintreau et al [2] made the hypothesis that the CCDF of inter-contact time, denoted as $F^0(t)$, between any two mobile devices is a Pareto distribution, thus

power law over $[t_0, +\infty)$, for some $t_0 > 0$. Concretely, for $\alpha > 0$,

$$F^0(t) = \left(\frac{t_0}{t}\right)^\alpha, \quad t \geq t_0. \quad (1)$$

The authors then argued that the assumption that the CCDF of inter-contact time has a power tail is in sharp contrast with prior work on packet forwarding; previous work would assume exponential tail for the CCDF distribution, such as for example, that of Grossglauser and Tse [7] that considers two-hop packet relying schemes. Under the assumption that inter-contact time between mobile devices are independent and identically distributed, Chaintreau et al [2] derived interesting results on the feasibility of two-hop packet relying schemes. In summary, they show that under the assumptions therein, there exists a two-hop relying scheme that ensures finite mean packet forwarding delay if $\alpha > 1 + 1/m$, where m is the number of packet replicas made from a source to distinct relay nodes, and that if $\alpha \leq 1$, for any packet forwarding scheme the mean packet forwarding delay is infinite. It is precisely the latter case ($\alpha \leq 1$) that was suggested to hold in real-life by the mobility traces analyzed so far.

However, Fig. 2 highlights that the observed dichotomy in the CCDF of inter-contact times between mobile devices, rather suggests to take as a hypothesis that the CCDF of the inter-contact times has exponentially decaying tail. This exponential tail entirely eliminates the issue of infinite packet forwarding delay under the power tail assumption. Furthermore, in the datasets the mean inter-contact time is of the same order as the characteristic time, and thus the exponential tail cannot be ignored by the time separation argument. This is of particular importance for practical schemes that were later proposed, such as throwboxes [19]. There, it is assumed that the mean inter-contact time is finite and can be estimated. This is a valid hypothesis under the dichotomy that we observed in the traces is a general feature, but would not be valid under the hypotheses of the model in [2].

We further contrast the dichotomy in the distribution of inter-contact time with the assumed power-law tail by Chaintreau et al [2]. For analyses similar to those in [2], it is of interest to consider the residual inter-contact time distribution, i.e. the time until the next contact for a node pair from an arbitrary point in time. Intuitively, the residual time reflects how much time a device has to wait before being able to forward a message to another specific device. Suppose for a moment that contacts between a node pair occur at instances of a stationary point process in time with a finite mean inter-contact time. It is well known that the CCDF of residual inter-contact time, $F(t)$, relates to the CCDF of the inter-contact time sampled at contact instances, $F^0(t)$, as

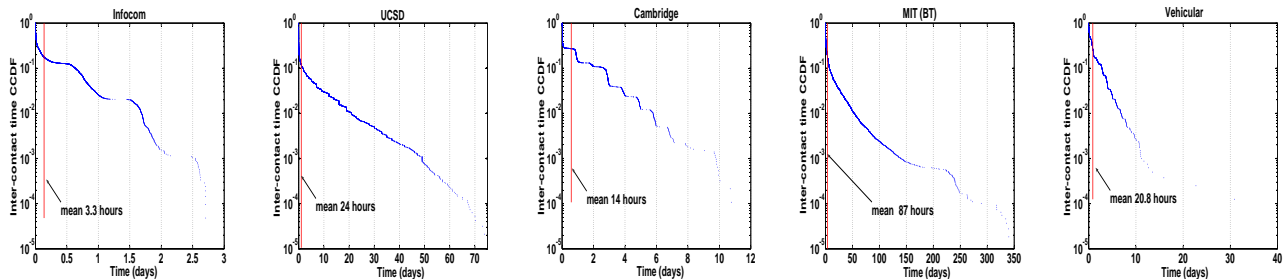


Figure 2: Same as in Fig. 1 but plotted in lin-log scale. The results confirm the exponential decay of the CCDF beyond half a day.

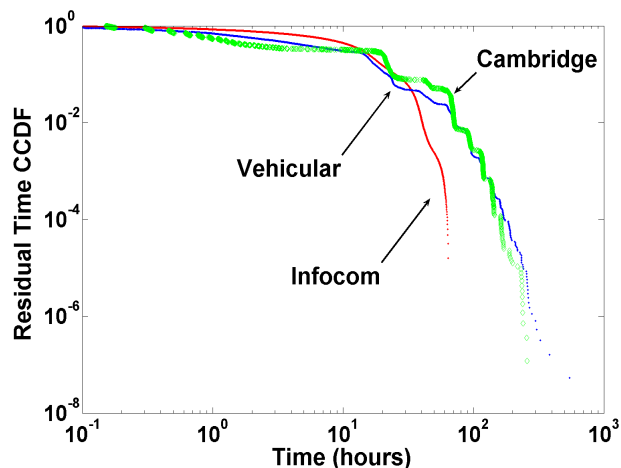


Figure 3: Residual inter-contact time CCDF.

follows:

$$F(t) = \lambda \int_t^{+\infty} F^0(s) ds \quad (2)$$

where $1/\lambda$ is the mean inter-contact time sampled at contacts, and is assumed to be finite. Under the assumptions of Chainterau et al, we have that Eq. (1) holds, and provided that $\alpha > 1$, it follows

$$F(t) = \left(\frac{t_0}{t}\right)^{\alpha-1}, \quad t \geq t_0,$$

thus, again a Pareto distribution but with scale parameter $\alpha - 1$. As a result, if we consider the empirical CCDF of the residual time, we should observe a power law that would manifest itself as a straight line in a log-log plot.

In contrast, if the empirical CCDF of inter-contact time exhibits the aforementioned dichotomy, we should rather observe that the rate of decrease in a log-log plot of the residual inter-contact time increases. This would

be the case provided that the exponent is not too large and would follow from the tail integration in Eq. (2). In accordance with the previous discussion, Fig. (3) shows the empirical CCDF of the residual inter-contact time for three datasets and confirms the increasing rate of the decay.

4. SIMPLE MOBILITY MODELS CAN SUPPORT THE DICHOTOMY

In this section, we show that already simple mobility models such as simple random walk on one-dimensional torus feature the dichotomy in the CCDF of inter-contact time in that it is close to a power law up to a characteristic time and beyond it has exponential decay. This model can be seen as an one-dimensional version of simple random walk on a two-dimensional torus, which was used as early as in [12] (Manhattan Street Network), and later used in recent studies (e.g. [6]), and can also be seen as a special case of random walk on torus model in [1]. We also show that random waypoint on a chain of discrete sites features the inter-contact CCDF that is close to a power law over an interval and has exponentially decaying tail.¹ These results contradict existing statements that current mobility models do not feature power law CCDF of inter-contact time. The results also show that for some mobility models the dichotomy in the inter-contact time CCDF is qualitatively precisely the same as observed in some empirical traces.

4.1 Simple random walk

We consider as mobility domain a circuit of m sites $0, 1, \dots, m-1$. (See Fig. 4.) A device moves according to a simple random walk: from a site i , it moves to either site $i-1 \bmod m$ or $i+1 \bmod m$ with equal probability. We denote with $X_k(n)$ the site on which a

¹The distribution of the inter-contact time under a random waypoint model was analyzed by Sharma and Mazumdar [17]. They showed that this distribution is exponentially bounded on both sides under assumptions that (a) mobility domain is a sphere and (b) any trip between two successive waypoints is of a fixed duration.

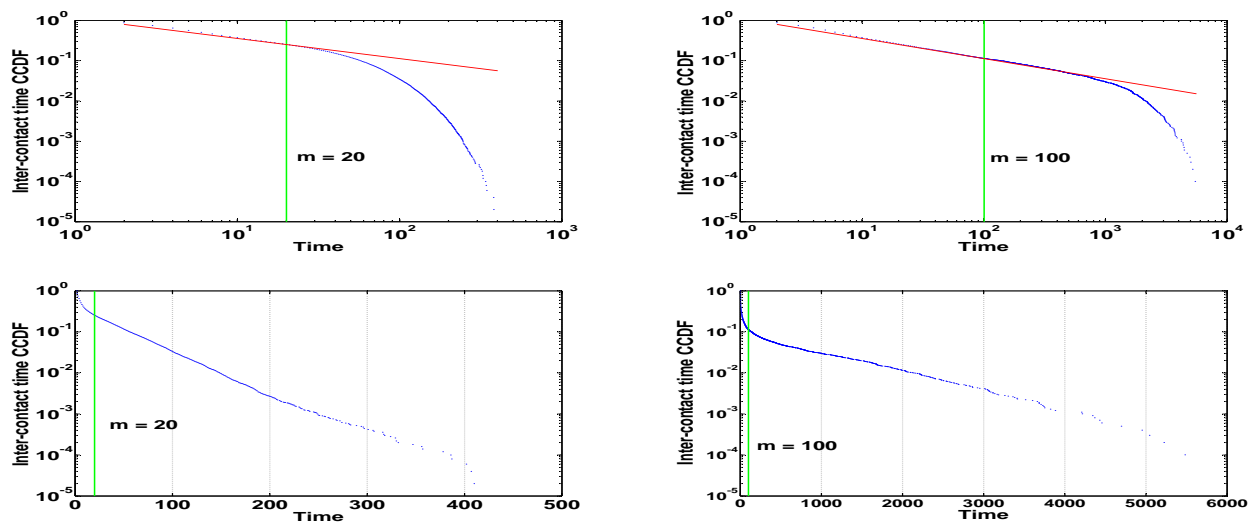


Figure 5: CCDF of inter-contact time for simple random walks on a circuit of m sites in log-log and lin-log scale.

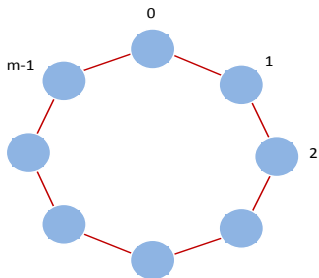


Figure 4: Circuit of m sites.

device k is at time $n \geq 0$.

4.1.1 Return time

We first consider the return time R of a single device to an arbitrarily fixed site. Without loss of generality, we may consider only the return time of device 1 to site 0, i.e. given that $X_1(0) = 0, X_1(1) \neq 0$,

$$R = \min\{n > 0 : X_1(n) = 0\}.$$

The result shows that already the return time to a site features the aforementioned dichotomy. We show later that inter-contact time CCDF features the same qualitative properties.

THEOREM 1 (RETURN TIME). *For the return time R of a simple random walk to a fixed site on a circuit of m sites:*

1. Expected return time:

$$\mathbb{E}(R) = m$$

2. Power-law for infinite circuit:

$$\mathbb{P}(R > n) \sim \sqrt{\frac{2}{\pi}} \frac{1}{n^{1/2}}, \text{ large } n$$

where for two functions f and g , $f(n) \sim g(n)$ means that $f(n)/g(n)$ tends to 1 as n goes to infinity.

3. Exponentially decaying tail:

$$\mathbb{P}(R > n) \sim \varphi(n)e^{-\beta n}, \text{ large } n$$

where $\varphi(n)$ is a trigonometric polynomial in n and $\beta > 0$. We call trigonometric polynomial in n a function of the form $\varphi(n) = \sum_{k=1}^K [a_k \cos(n\omega_k) + b_k \sin(n\omega_k)]$ where a_k, b_k and ω_k are constants.

Item 1 shows that average return time to a site is equal to the time needed to circumvent the circuit. Item 2 shows that for a circle of infinitely many sites, the asymptotic of return time CCDF is precisely the power-law with exponent 1/2. The result suggests that the asserted asymptotic may be a good approximation of the CCDF for large but finite circuit (Fig. 5 shows that this holds already for as few as 20 sites). It is noteworthy that the power-law under item 2 holds more generally for any one-dimensional aperiodic recurrent random walk with finite variance $\sigma^2 < \infty$ such that we have [18, P3, p381]

$$\mathbb{P}(R > n) \sim \sqrt{\frac{2}{\pi}} \sigma \frac{1}{\sqrt{n}}, \text{ large } n.$$

Item 3 shows that for a circle of a finite number of sites, the CCDF of the return time has exponential decay.

PROOF. Item 1. Let r_i be the mean hitting time of

site 0 starting from site i . We have that $r_0 = 0$ and

$$r_i = 1 + \frac{1}{2}(r_{i-1} + r_{i+1}), \quad i = 1, \dots, m-1$$

where addition in indices is modulo m . It can be shown by induction on i that the solution is

$$r_i = i(m-i), \quad i = 0, 1, \dots, m-1.$$

The assertion of item 1 follows, noting that $\mathbb{E}(R) = 1 + r_1$.

Item 2. We consider the z -transform of the return time R to site 0 started from a site i , i.e.

$$f_i(z) := \mathbb{E}(z^R | X_0 = i), \quad i = 0, 1, \dots, m-1.$$

We have the following system of linear equations

$$\begin{aligned} f_0(z) &= f_m(z) = 1 \\ f_i(z) &= z \frac{1}{2} (f_{i-1}(z) + f_{i+1}(z)), \quad i = 1, \dots, m-1. \end{aligned}$$

It suffices to solve the following classical system of linear equations

$$f_0 = f_m(z) = 1 \quad (3)$$

$$f_i = x(f_{i-1} + f_{i+1}), \quad i = 1, \dots, m-1. \quad (4)$$

for a fixed $0 < x \leq 1/2$, which we will encounter again while considering inter-contact time. The system can be solved by noting that $f_i = u^i$ is a particular solution for some u . Plugging the particular solution in the system (4) that u is given as the solution of the quadratic characteristic equation for the recurrence (4) given by:

$$u^2 - \frac{1}{x}u + 1 = 0.$$

Thus, we have

$$u_{1,2} = \frac{1}{2x} \pm \sqrt{\left(\frac{1}{2x}\right)^2 - 1}.$$

The solution is

$$f_i = au_1^i + bu_2^i, \quad i = 1, \dots, m-1. \quad (5)$$

From the boundary conditions (3), we have

$$\begin{aligned} 1 &= a + b \\ 1 &= au_1^m + bu_2^m. \end{aligned}$$

It follows $b = 1 - a$ and

$$\begin{aligned} a &= \frac{1 - u_2^m}{u_1^m - u_2^m} \\ &= \frac{(x/2)^m - (1 - \sqrt{1 - (x/2)^2})^m}{(1 - \sqrt{1 + (x/2)^2})^m - (1 - \sqrt{1 - (x/2)^2})^m}. \end{aligned}$$

Finally, from (5), $x = z/2$ and the last display, we have

$$\begin{aligned} f_0(z) &= 0 \\ f_i(z) &= \frac{a(m,z)(1 + \sqrt{1 - z^2})^i + (1 - a(m,z))(1 - \sqrt{1 - z^2})^i}{z^i}, \\ a(m,z) &= \frac{z^m - (1 - \sqrt{1 - z^2})^m}{(1 + \sqrt{1 - z^2})^m - (1 - \sqrt{1 - z^2})^m}. \end{aligned}$$

Of our particular interest is $f_1(z)$ that can be expressed as:

$$f_1(z) = \frac{1 + (2a(m,z) - 1)\sqrt{1 - z^2}}{z}. \quad (6)$$

The assertion under item 2 follows by noting that

$$\lim_{m \rightarrow \infty} a(m,z) = 0, \quad \text{for } 0 < z < 1$$

and, hence, for an infinite circuit,

$$f_1(z) = \frac{1 - \sqrt{1 - z^2}}{z}.$$

Using the Binomial theorem for $(1 - z^2)^{1/2}$ and some elementary calculus, we have

$$f_1(z) = \sum_{n=1: n \text{ odd}}^{\infty} \binom{\frac{1}{2}}{n+1} (-1)^{\frac{n-1}{2}} z^n$$

where

$$\binom{\frac{1}{2}}{k} = \frac{\prod_{n=0}^{k-1} (\frac{1}{2} - n)}{k!}. \quad (7)$$

It thus follows that

$$\mathbb{P}(R > n) = \sum_{m \text{ odd}, m > n} \left| \binom{\frac{1}{2}}{\frac{m+1}{2}} \right| = \sum_{k \geq \lceil \frac{n}{2} \rceil + 1} \left| \binom{\frac{1}{2}}{k} \right| \quad (8)$$

Further,

$$\left| \binom{\frac{1}{2}}{k} \right| \sim \frac{1}{2\sqrt{\pi}k^{3/2}}, \quad \text{large } k. \quad (9)$$

To show this, note that Eq. (7) can be rewritten as

$$\binom{\frac{1}{2}}{k} = \frac{(-1)^{k+1}}{k!2^k} \prod_{n=1}^{k-1} (2n-1)$$

and then note

$$\prod_{n=1}^{k-1} (2n-1) = \frac{(2(k-2)+1)!}{2^{k-2}(k-2)!}.$$

Hence,

$$\binom{\frac{1}{2}}{k} = \frac{(-1)^{k+1} (2(k-2)+1)!}{k!2^k (k-2)!}.$$

The asymptotic (9) then follows by using Stirling's approximation.

We now use the fact that if $u_n \sim \frac{1}{n^\alpha}$ then $\sum_{m=n}^{\infty} u_m \sim \frac{1}{\alpha-1} \frac{1}{n^{\alpha-1}}$ n large, for $\alpha > 1$. The asserted result follows from (9) and (8).

Item 3: It follows from Lemma 1 (see below) with the subset Δ reduced to the element 0. \square

The following lemma is used in the proof of Theorem 1 and in other places in this paper, so we give it in a fairly general form.

LEMMA 1 (RETURN TIME FOR FINITE MARKOV CHAIN).
Let X_n be an irreducible Markov chain on some finite state space S and let Δ be a subset of S ($\Delta \neq \emptyset$ and $\Delta \neq S$). Let R be the return time to Δ . The stationary distribution of R is such that

$$\mathbb{P}(R > n) \sim \varphi(n)e^{-\beta n}, \text{ large } n$$

where $\varphi(n)$ is a trigonometric polynomial and $\beta > 0$.

Proof is given in appendix. Note that R is the time between leaving Δ and returning to Δ . The hypotheses imply that the chain is positive recurrent and thus R is finite. The proof relies on spectral decomposition of non-negative matrices, using results in [16, 3].

4.1.2 Inter-contact time

We consider mobility of two devices according to two independent simple random walks on a circuit of m sites. We assume that m is even. The inter-contact time between two devices is defined as, given $X_1(0) = X_2(0)$ and $X_1(1) \neq X_2(1)$,

$$T = \min\{n > 0 : X_1(n) = X_2(n)\}.$$

We next examine the CCDF of inter-contact time T .

THEOREM 2 (INTER-CONTACT TIME).
Consider two independent simple random walks on a circuit of m sites, where m is assumed to be even. The inter-contact time T between the two random walks has the following properties.

1. Expected inter-contact time:

$$\mathbb{E}(T) = m - 1$$

2. Power-law for an infinite circuit:

$$\mathbb{P}(T > n) \sim \frac{2}{\sqrt{\pi}} \frac{1}{n^{1/2}}, \text{ large } n$$

3. Exponentially decaying tail:

$$\mathbb{P}(T > n) \sim \varphi(n)e^{-\beta n}, \text{ large } n$$

where $\varphi(n)$ is a trigonometric polynomial in n and $\beta > 0$.

With regard to the asserted properties, the inter-contact time is qualitatively the same as the return time to a site for a single simple random walk on a circuit considered in Theorem 1. In particular, item 2 asserts the same asymptotic CCDF as for the return time in Theorem 1 except only for different multiplicative constant. Similarly, item 3 is the same property as holding for the return time in Theorem 1.

PROOF. We prove by reduction to considering return time to a site for a single simple random walk as described next. The two independent random walks

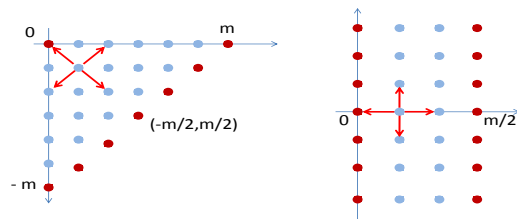


Figure 6: Reduction to return time of an one-dimensional simple random walk.

amount to a random walk on a two-dimensional lattice, with transition probabilities

$$\mathbb{P}_{(i,j)}((X_1(1), X_2(1)) = (i \pm 1, j \pm 1)) = \frac{1}{4}.$$

Without loss of generality, we assume $(X_1(0), X_2(0)) = (0, 0)$ and $(X_1(1), X_2(1)) = (1, -1)$. The inter-contact time is the hitting time plus 1 of the two-dimensional random walk (X_1, X_2) with the hitting set $\{i \cdot m + j, i, j = \dots, -1, 0, 1, \dots\}$, starting from the point $(1, -1)$. This is equivalent to considering hitting time of the boundaries $(0, i)$ and $(m/2, i)$, $i = \dots, -1, 0, 1, \dots$, for a simple two-dimensional random walk started at the point $(0, 1)$. See Fig. 6 for an illustration. This hitting time can be represented as

$$T = 1 + \sum_{i=1}^H V_i \quad (10)$$

where H is the number of transitions along the x axis until hitting of the boundaries and V_i is the number of transitions along the vertical axis between the $(i-1)$ st and i th horizontal transition. Note that H is return time to site 0 of a simple random walk on a circuit of $m/2$ sites started at site 1. We have that H and (V_1, V_2, \dots, V_H) are independent and that for any given H , (V_1, V_2, \dots, V_H) is a sequence of independent and identically distributed random variables with distribution

$$\mathbb{P}(V_i = k) = \frac{1}{2^k}, \quad k = 1, 2, \dots$$

Item 1: From (10) and noted independency properties, we can use Wald's lemma to assert

$$\mathbb{E}(T) = 1 + \mathbb{E}(H)\mathbb{E}(V_1).$$

The random variable H is the return time for simple random walk on a circuit of $m/2$ sites, so $\mathbb{E}(H) = m/2 - 1$. Note also that $\mathbb{E}(V_1) = 2$. It follows

$$\mathbb{E}(T) = m - 1.$$

Item 2. We consider the z -transform of the inter-contact time T . We have

$$\mathbb{E}(z^T) = z g_H \left(\frac{z}{2-z} \right) \quad (11)$$

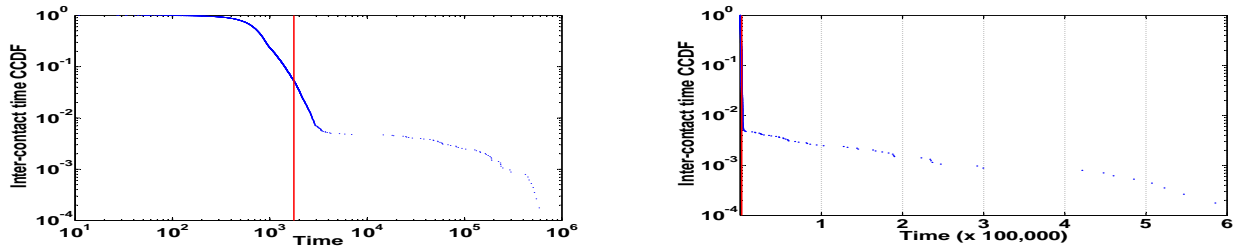


Figure 7: Inter-contact time CCDF for random waypoint on a chain of $m = 1000$ sites on log-log and lin-log scale.

where $g_H(\cdot)$ is the z -transform of the random variable H . To see this, first note

$$\mathbb{E}(z^{V_i}) = \frac{z}{2-z}.$$

The asserted identity Eq. (11) is direct by following simple calculus

$$\begin{aligned} \mathbb{E}(z^T) &= z\mathbb{E}\left(\prod_{i=1}^H z^{V_i}\right) \\ &= z\mathbb{E}\left(\mathbb{E}(z^{V_1})^H\right) = zg_H\left(\frac{z}{2-z}\right) \end{aligned}$$

where $g_H(z) := \mathbb{E}(z^H)$.

Now, in Eq. (6), we already derived the z -transform of the return time to site 0 for simple random walk on a circuit of m sites. From Eq. (11) and Eq. (6), we obtain

$$\mathbb{E}(z^T) = 2 - z + 2(2b(m/2, z) - 1)\sqrt{1-z} \quad (12)$$

with

$$b(n, z) = \frac{z^n - (2 - z - 2\sqrt{1-z})^n}{(2 - z + 2\sqrt{1-z})^n - (2 - z - 2\sqrt{1-z})^n}.$$

The assertion under item 2 now follows from (12) by mimicking the proof of Theorem 1.

Item 3. This follows from Lemma 1 with Markov chain $X(n) = (X_1(n), X_2(n))$, S the set of reachable states, and subset $\Delta = \{(i, i), i = 0, \dots, m-1\}$. \square

4.2 Random walk on a 2-dim torus

Similarly one may consider mobility of a device defined as a random walk on a two-dimensional torus of m and k sites in the respective two dimensions. One may extend the one-dimensional mobility of a device to two dimensions by assuming that device mobility in each of the dimensions is a simple random walk and the two random walks are independent. This model resembles the well known Manhattan-grid model [12]. From Lemma 1, we know that also for this model, the CCDF of inter-contact times between two devices has exponential tail. A detailed analysis of the CCDF of inter-contact time is beyond the scope of this paper. Such

an analysis may follow the same steps as in this section, but for the difference random walk describing the difference between coordinates of the two independent random walkers. It is not clear that the same dichotomy would hold. For example, we know from [18, E1, p167] that the CCDF of the return time to a site for a simple random walk in two dimensions is $\pi/\log(n)$, for large n . The interested reader may refer to [11] for simulation estimates of the inter-contact time in two dimensions.

4.3 Random waypoint on a chain

We consider random waypoint on a chain of m sites. This is a discrete time, discrete space version of well known random waypoint [9]. Each device is assumed to move stochastically independently. A movement of a device is specified by its current site and next waypoint site. The device moves to its next waypoint site by one site per time instant. When it reaches the next waypoint, it updates the next waypoint to a sample drawn uniformly at random on the set of sites constituting the chain and the movement continues as described. Two devices are assumed to be in contact at a time t , if at this time they reside in the same site. We analyze this model by simulations. In Fig. 7, we show the empirical estimate of the CCDF of inter-contact time, both in log-log and lin-log scale. The results demonstrate that random waypoint can feature a power law like decay of the CCDF over an interval that covers the mean inter-contact time. We also observe the presence of short and long inter-contact times. Fig. 8 suggests that long inter-contact times occur due to the assumptions that two devices are in contact only if in the same site.

5. SPATIO-TEMPORAL BREAKDOWN

Here, we breakdown device contacts along several dimensions. Our goal is to better understand individual elements that contribute to the aggregate measures reported in preceding sections. Note that our findings thus far have been obtained by aggregating over individual device pairs and also time.

First, we breakdown device inter contacts by analyzing return times of individual mobile devices to their re-

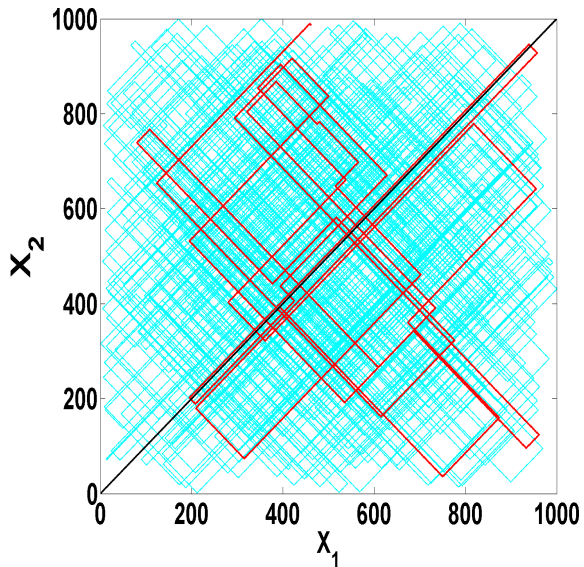


Figure 8: Mobile positions moving according to random waypoint on a chain of $m = 1000$ sites. The thick trajectory corresponds to a long inter-contact time.

spective most frequently visited sites. A site here refers to a location region such as a circular area for the vehicular data or an AP/cell for UCSD/MIT data. Our analysis suggests that return times exhibit the same dichotomy in the distribution as the one found for the inter-contact times between device pairs. We then pose and confirm the hypothesis that devices are in contact at a small set of distinct sites. These two findings suggest that the dichotomy in the distribution of the return time may already explain the observed dichotomy in the distribution of inter-contact time between devices.

Second, we discuss how the aggregate CCDF of inter-contact time between devices as obtained from aggregate samples of inter-contact time over *all* device pairs over a measurement period relates to the CCDF of inter-contact time for individual device pairs. Further, we ask the question what this aggregate CCDF yields when used to characterize the inter-contact time between devices observed from an arbitrary point in time.

Third, we examine the extent of time nonstationarity in device inter contacts and non surprisingly confirm the presence of strong time of day dependencies.

5.1 Return versus inter-contact time

Our evaluation in the previous sections reveals a characteristic time in the order of half a day, which could be attributed to the daily periodicity of human behavior. Our goal here is to capture features of human mobility

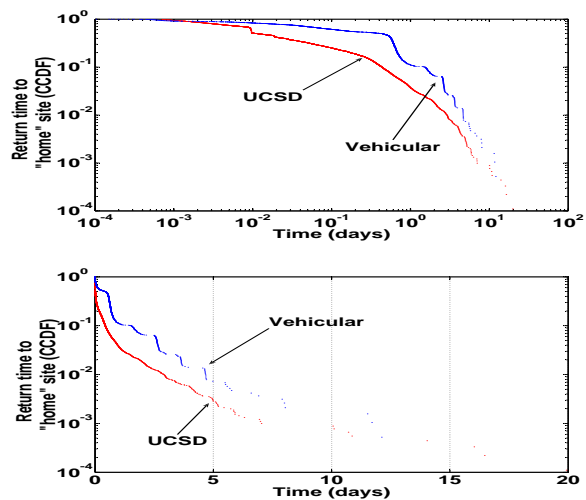


Figure 9: Return time exhibits the same dichotomy as the inter-contact time (in log-log and lin-log scale).

that could play a role in the observed dichotomy and in the particular decay of the inter-contact time distribution within the two timescales.

To this end, we examine the return time of a device to a particular location or site. Note that the return time characterizes mobility of a single human and thus may be regarded as a more elementary characterization of human mobility than inter-contact time. Having established in the previous section that for simple mobility models (e.g., independent random walks on a finite circuit) the CCDF of inter-contact time between two random walkers and the CCDF of the return time to a specific site for a single random walker feature precisely the same dichotomy, we now examine whether this observation holds in real mobility cases.

In a hypothetical scenario where two mobile devices would almost always meet at a particular site, the inter-contact time between the two devices would be stochastically larger than the return time of any of the two devices to that given site. Supposing further that two devices are synchronized in time, then the return time to a site would closely characterize the inter-contact time. In this section, we demonstrate that return times of a device to a specific site feature the observed dichotomy.

The dichotomy characterizes the return times of individual devices to their “home” sites. In order to test the hypothesis of the dichotomy in the CCDF of device return time to a specific site, we conducted the following analysis. For each device, we infer a “home” site defined as the location region where the device spends most of its time. A location region is either circular area of some radius r for the vehicular trace, or an AP/cell for the UCSD/MIT trace. In Fig. 9, we show the CCDF

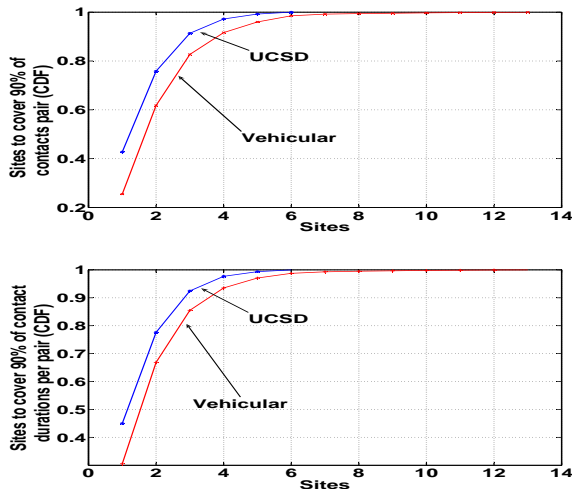


Figure 10: CDF of the number of sites that cover 90% of contacts (top) and contact durations (bottom) per device pair.

of the inferred return time of a device to its home site over all devices. The figure shows remarkable qualitative similarity to the corresponding CCDF of device inter-contact time (Fig. 1). As we argued over the previous paragraphs, this is an interesting property as the return time is a more basic characterization of human mobility than inter-contact time.

We further test the hypothesis that typically two devices meet at a few sites. To that end, we counted the number of sites per each device pair ranked with respect to their frequency of contacts. Then for each device pair, we examined how many sites cover either 90% of their contacts or 90% of their total contact duration. The two corresponding CDFs are shown in Fig. 10, where the median number of sites is less than 2 and the 90% quantile is less than 4 sites in all the considered cases.

The main theme of this paper is around the time dimension of device mobility. In this paragraph, we detour slightly to briefly consider the spatial aspect of the return time to the home site. In Fig. 11, we show the CCDF of the trip distance incurred on the return trips to the home site of a device. We present results only for the vehicular dataset since this is the only trace with precise location information for each device. The CCDF is well approximated by a straight line in the log-log scale over a wide range of distances spanning 40 to 200 kilometers. For smaller distances, the distribution appears to decay exponentially. While the spatial aspect of human mobility is itself an interesting topic, it is beyond the scope of this paper to pursue this further in more detail.

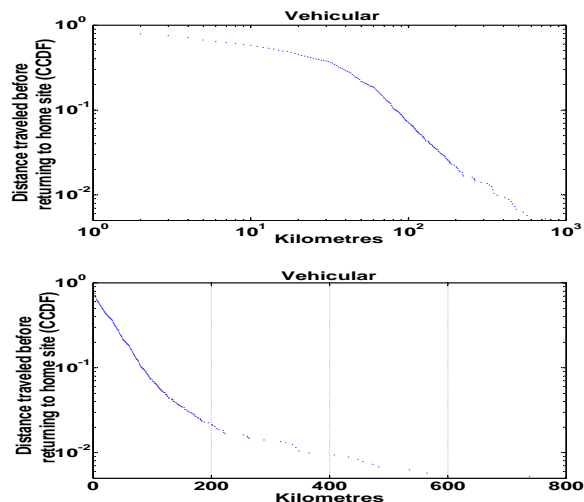


Figure 11: Travel distance on a return trip to device home site in log-log and lin-log scale.

5.2 Contacts across different viewpoints

In this section we consider different viewpoints on device inter contacts and their interpretations from the packet forwarding perspective. We address the following questions:

- Is the aggregate CCDF of inter-contact time, derived from samples aggregated over all device pairs over a measurement interval, representative of the CCDF of inter-contact time for a specific pair of devices?*
- What metric does the aggregate CCDF correspond to when used to evaluate the delay of an opportunistic forwarding scheme?*
- How does the inter-contact time statistic depend on the time of day?*

5.2.1 Aggregate vs per device-pair viewpoint

Previous studies and the analysis in Section 3 considered the CCDF of inter-contact time obtained from samples aggregated over all device pairs over a measurement period. *We call this the aggregate CCDF.* We examine here how the aggregate CCDF of inter-contact time relates to the CCDF of inter-contact time for a device pair. In general, the two are different and the bias is such that the aggregate CCDF gives more weight to devices that meet more frequently.

Consider a mobility trace over a measurement interval of duration T and let the time origin 0 be defined as the beginning of the measurement interval. We denote with \mathcal{P} the set of all distinct device pairs that were in contact at least twice over the measurement interval. Let T_n^p the time of the n th contact for a device pair p , with $n = 1, 2, \dots$ and let for this pair, $N_p(t)$ be the number of contacts on $[0, t]$. The empirical aggregate

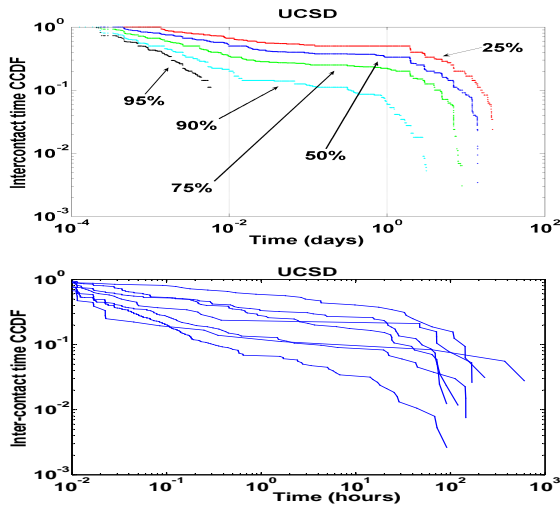


Figure 12: Aggregate vs. per device-pair CCDF of inter-contact time (top) and 7 individual device-pair CCDFs (bottom).

CCDF for a value $t \geq 0$ is defined as the fraction of inter-contact times over all device pairs in \mathcal{P} that are larger than t , i.e.

$$\hat{F}^0(t, T) = \frac{1}{N(T)} \sum_{p \in \mathcal{P}} \sum_{n=1}^{N_p(T)-1} 1(T_{n+1}^p - T_n^p > t) \quad (13)$$

where $1(A)$ is the indicator whether the condition A holds² and $N(t) = \sum_{p \in \mathcal{P}} N_p(t)$ is the number of contacts over all pairs in \mathcal{P} on $[0, t]$. We rewrite the above identity as:

$$\hat{F}^0(t, T) = \sum_{p \in \mathcal{P}} \frac{N_p(T) - 1}{N(T)} \hat{F}_p^0(t, T) \quad (14)$$

where $\hat{F}_p^0(t, T)$ is the empirical CCDF of inter-contact time for a device pair p given by

$$\hat{F}_p^0(t, T) = \frac{1}{N_p(T) - 1} \sum_{n=1}^{N_p(T)-1} 1(T_{n+1}^p - T_n^p > t).$$

Eq. (14) tells us that the aggregate CCDF is a *weighted* sum of the CCDFs over device pairs, with the weight for a device pair proportional to the number of contacts observed for this device pair. This is indeed intuitive as we expect to observe a larger number of inter-contact samples for pairs of devices that meet more frequently.

For the sake of discussion, suppose for a moment that contacts between mobile devices occur at instances of a point process that is assumed to be time stationary and ergodic, but not necessarily stochastically identical over pairs of devices. From Eq. (13), it follows that as

² $1(A) = 1$ if A true, else $1(A) = 0$.

T tends to be large, $\hat{F}^0(t, T)$ converges to

$$F^0(t) = \sum_{p \in \mathcal{P}} \frac{\lambda_p}{\lambda} F_p^0(t) \quad (15)$$

where $1/\lambda_p$ is the mean of inter-contact times sampled at contact instances of the device pair p and $F_p^0(\cdot)$ is the corresponding CCDF, and $\lambda = \sum_{p \in \mathcal{P}} \lambda_p$ is the total rate of contacts over all device pairs. We note that the aggregate CCDF, $F^0(t)$, exactly matches each of the CCDFs $F_p^0(t)$, $p \in \mathcal{P}$, only if contacts for distinct pairs are *stochastically identical*. The aggregate CCDF of inter-contact time is equal to the weighted sum of individual CCDFs given in Eq. (15) with weight for a device pair p proportional to the rate of contacts λ_p .

The preceding discussion raises the question of how representative the aggregate CCDF of inter-contact time is for an arbitrarily chosen device pair. To address this question, we explore how the aggregate CCDF differs from the CCDF of a pair of devices in the different datasets. Fig. 12-top shows the aggregate CCDF of inter-contact time along with percentiles of the CCDF over all device-pairs for the UCSD dataset. We observe that for each given time, more than half of node pairs have a CCDF of the inter-contact time in a reasonably narrow neighborhood around the aggregate CCDF. In Fig. 12-bottom, we show 7 distinct CCDFs of individual device-pairs, which on the other hand present some variability that could be hidden at the aggregate viewpoint. We have examined the discrepancy of the aggregate CCDF and the arithmetic mean of individual CCDFs and observed that the former lower bounds the latter but their difference was not substantially large.

5.2.2 Time-average viewpoint

In performance analyses of forwarding schemes, the CCDF of residual time until contact between two devices from an observation time is often derived from the CCDF of inter-contact time sampled at contact instants of this device pair. The latter is often estimated by the aggregate CCDF of inter-contact time. We would like to understand what does this residual time CCDF correspond to when we use the aggregate CCDF of inter-contact time. We will see that this residual inter-contact time distribution, in fact, corresponds to an observation time sampled uniformly at random on the measurement interval and for a device pair sampled uniformly at random. Hence, the resulting viewpoint is that of time averaging and averaging over device pairs.

We revisit the earlier setting and now consider the fraction of device pairs for which the residual time until next contact is larger than $t \geq 0$ as observed from a time instant s , i.e.

$$\hat{F}(t, s) = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} 1(T_{N_p(s)+1}^p - s > t). \quad (16)$$

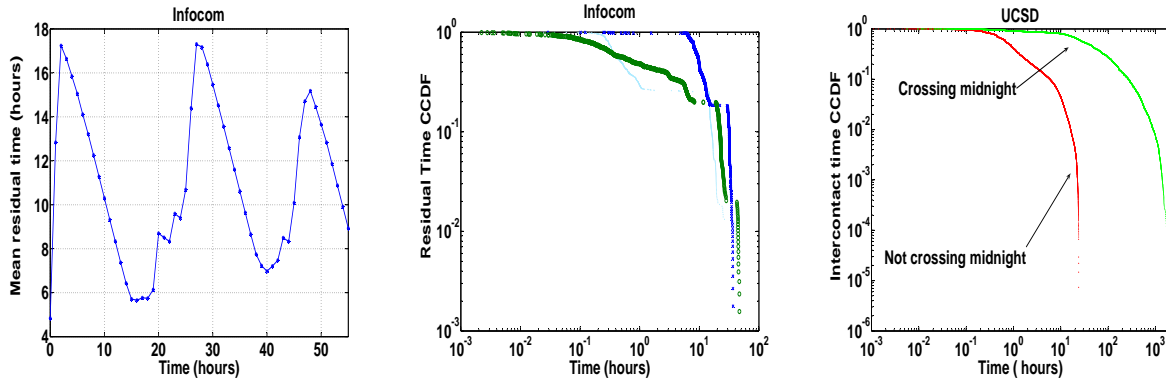


Figure 13: LEFT: Mean residual time at various times of the day shows nonstationarity effects. MIDDLE: Residual time CCDF at different times of day. RIGHT: CCDFs of inter-contact times crossing midnight vs. inter-contact times during the day.

By averaging over the measurement interval, we have

$$\hat{F}(t, T) = \frac{1}{T} \int_0^T \hat{F}(t, s) ds. \quad (17)$$

After some straightforward but tedious calculus, it follows that we can rewrite Eq. (17) as

$$\hat{F}(t, T) = \frac{N(T)}{|\mathcal{P}|T} \int_t^{+\infty} \hat{F}^0(s, T) ds + e(T) \quad (18)$$

where $e(T)$ is a term that captures the boundary effects and in all regular cases (e.g. stationary ergodic) diminishes with the length of the measurement interval T , so we ignore it for the sake of our discussion.

From Eq. (16) and Eq. (18), we note that by using the aggregate CCDF of inter-contact time to estimate the CCDF of the residual inter-contact time, this in fact corresponds to time averaging and averaging over device pairs. This viewpoint may differ substantially from the viewpoint at a specific time of day due to time nonstationarity of device contacts. We explore this non stationarity in the following section.

5.2.3 Time of day viewpoint

We now confirm from our datasets that device inter contacts exhibit strong time-of-day nonstationarity. It is important to note presence of this nonstationarity as a claim based on the time-average viewpoint may not hold for the viewpoint of a particular time of day. Fig. 13 presents three sets of results highlighting the effects of time nonstationarity.

In Fig. 13-left, we show the mean residual inter-contact time over all device pairs versus the time for the Infocom trace. The figure shows strong dependency on the time of day, with day and night periods resulting in the mean residual time ranging from about 6 to 17 hours. This effect is also evident in the aggregate CCDF of the residual inter-contact time in Fig. (13)-middle, where

we plot the CCDF for three distinct times within the same day (midday, early evening, after midnight). We observe a significant variation across the three curves in accordance with the mean variation (Fig. 13-left). We further looked at the aggregate CCDF of inter-contact time conditional on whether inter-contact time cross over midnight or not for the UCSD trace. Fig. (13)-right shows the discrepancy of the respective conditional distributions. In summary, the results confirm the intuition that device contacts would typically exhibit strong time nonstationarity and particular time of day viewpoints may differ much from the time-average viewpoint.

We subsequently demonstrate the time of day dependence by examining the contact durations for the vehicular trace. Fig. 14-top shows samples of contact durations per device pair. These samples suggest a dichotomy of contact durations consisting of (a) short contacts in the order of half a minute, and (b) long contacts in the order of 10 hours. Examining the trace, we found that the short contacts occur while two vehicles drive by each other, while long contacts take place for spatially collocated vehicles during working hours. Fig. 14-bottom further confirms the previous discussion by showing two CDFs of samples of contact durations for device-pairs that initiated contacts within the hours of 9AM and 4PM. As previously mentioned, these distributions suggest that long contact durations occur during working hours while at other times short contacts may be more frequent.

6. CONCLUDING REMARKS

The dichotomy hypothesis—power law decay of inter-contact time distribution up to a point and exponential decay beyond—which we observed to hold across diverse mobility traces, implies that existing predictions

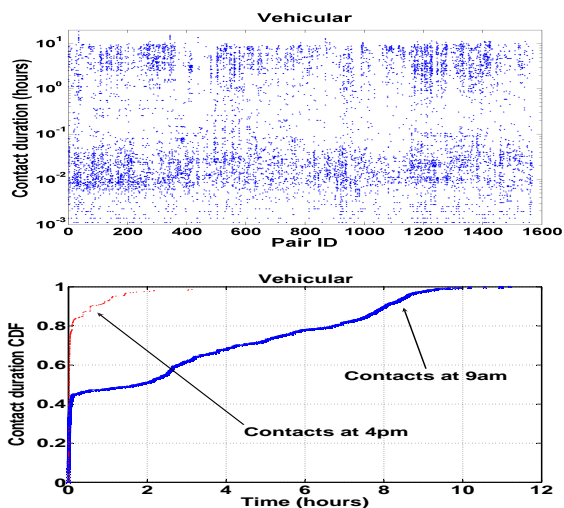


Figure 14: TOP: Samples of contact durations over all device pairs indicate dichotomy of contact durations. BOTTOM: CDFs of contact duration at two distinct times of day.

on the performance of forwarding schemes based on the power-law tail might be overly pessimistic. The dichotomy is not at odds with current mobility models since we show that already simple models support it. The empirical results suggest the dichotomy to hold also for inter-contact time between a mobile device and its frequently visited site, which may inform design of opportunistic communication systems provisioned with stationary infrastructure nodes. The diversity of viewpoints such as per device pair and at a time of day may deviate from the average viewpoint derived from the inter-contact time characterization widely used in previous studies and also considered in this paper. Future work may study further the underlying mobility patterns to understand better the first principles that induce the observed aggregate behavior of contact opportunities.

Acknowledgements

We are grateful to those who made their mobility traces available to public that include MIT [5, 4], UCSD [13], Infocom [2, 15]. We also thank our colleagues Eric Horvitz and John Krumm who generously provided us with the vehicular traces that they had collected and we used for analysis in this paper.

7. REFERENCES

- [1] J.-Y. L. Boudec and M. Vojnović. The Random Trip Model: Stability, Stationary Regime, and Perfect Simulation. *IEEE/ACM Trans. on Networking*, 14(6):1153–1166, Dec 2006.
- [2] 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 *INFOCOM*, 2006.
- [3] E. Cinlar. *Introduction to Stochastic Processes*. Prentice Hall, 1 edition, 1975.
- [4] N. Eagle and A. Pentland. CRAWDAD data set mit/reality (v. 2005-07-01), July 2005.
- [5] N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. In *Journal of Personal and Ubiquitous Computing*, 2005.
- [6] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah. Optimal Throughput-delay Scaling in Wireless Networks – Part I: The Fluid Model. *IEEE Trans. on Information Theory*, 52(6):2568–2592, June 2006.
- [7] M. Grossglauser and D. Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Trans. on Networking*, 10(4):477–486, 2002.
- [8] Intelligent Transportation Systems Standards Program. Dedicated Short Range Communications, April 2003. http://www.standards.its.dot.gov/Documents/...advisories/dsrc_advisory.htm.
- [9] D. B. Johnson and D. A. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. In *Mobile Computing*. 1996.
- [10] J. Krumm and E. Horvitz. The Microsoft Multiperson Location Survey, August 2005. Microsoft Research Technical Report, MSR-TR-2005-13.
- [11] M. Vojnović. On the origins of power laws in mobility systems. Workshop on Clean-Slate Network Design, Cambridge, UK, Sept 2006, <http://research.microsoft.com/~milanv/...powerlaw.pps>.
- [12] N. F. Maxemchuk. Routing in the Manhattan Street Network. In *IEEE Trans. on Comm.*, volume 35, pages 503–512, 1987.
- [13] M. McNett and G. M. Voelker. Access and mobility of wireless pda users. In *Mobile Computing Communications Review*, 2005.
- [14] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau. CRAWDAD data set cambridge/haggle (v. 2006-01-31), Jan. 2006.
- [15] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau. CRAWDAD trace cambridge/haggle/ imote/infocom (v. 2006-01-31), Jan. 2006.
- [16] E. Seneta. *Non-Negative Matrices*. Wiley and Sons, 1 ed., 1973.
- [17] G. Sharma and R. R. Mazumdar. Delay and Capacity Trade-off in Wireless Ad Hoc Networks with Random Waypoint Mobility, 2005. Preprint,

School of ECE, Purdue University, 2005.

- [18] F. Spitzer. *Principles of Random Walk, Graduate Texts in Mathematics*. Springer, 2nd edition, 1964.
- [19] W. Zhao, Y. Chen, M. Ammar, M. D. Corner, B. N. Levine, and E. Zegura. Capacity Enhancement using Throwboxes in DTNs. In *Proc. IEEE Intl Conf on Mobile Ad hoc and Sensor Systems (MASS)*, Oct 2006.

APPENDIX

Proof of Lemma 1

Let $\vec{f}(z)$ be the vector whose s th entry is

$$f_s(z) = \mathbb{E}(z^R | X_0 = s)$$

and let $p_{s,j}$ be the transition probability from state s to state j . Mimicking the proof of item 2 of Theorem 1, we have, using the backward equation of Markov chains:

$$\begin{aligned} f_s(z) &= 1 \text{ if } s \in \Delta \\ f_s(z) &= z \left[\sum_{j \notin \Delta} p_{s,j} f_j(z) + \sum_{j \in \Delta} p_{s,j} \right] \end{aligned}$$

In matrix form, this gives

$$\vec{f}(z) = z \left(Q \vec{f}(z) + \vec{b} \right) \quad (19)$$

where $\vec{f}(z)$ is the column vector whose s th entry, $s \in \Delta$, is $f_s(z)$, Q is the square matrix with entries $p_{s,j}$, $s \in \Delta$, $j \in \Delta$, and \vec{b} is the column vector whose s th entry, $s \in \Delta$, is $b_s = \sum_{j \in \Delta} p_{s,j}$.

We derive from (19) that for $z \in (0, \rho)$, where ρ is the spectral radius of Q :

$$\vec{f}(z) = \sum_{n \geq 0} z^{n+1} Q^n \vec{b}$$

and thus $\mathbb{P}(R = n | X_0 = s)$ is the s th entry of $Q^{n-1} \vec{b}$.

Let π^0 be the row vector whose s th entry, $s \in \Delta$, is the stationary probability that the Markov chain is in state s just after leaving Δ . It follows that

$$\mathbb{P}(R = n) = \pi^0 Q^n \vec{b}$$

We show later that the spectral radius ρ of Q is > 1 , thus the following manipulations are legitimate:

$$\mathbb{P}(R > n) = \sum_{m > n} \pi^0 Q^m \vec{b} = \pi^0 Q^n (Id - Q)^{-1} \vec{b}$$

Now we study the spectral decomposition of Q . Assume temporarily that Q is irreducible (this is not necessarily true, even if the Markov chain X_n is irreducible, as Q does not include transitions to states in Δ , but we will relax this assumption later). In this case, the spectral structure of Q follows from the Perron-Frobenius theorem [3]. There is an integer d (the periodicity of the underlying graph) such that all eigenvalues have a modulus $< \rho$ (the spectral radius, here < 1), except for

exactly d eigenvalues, equal to ρu^m , $m = 0, \dots, d-1$, where u is the complex number of modulus equal to 1 $u = e^{\frac{2i\pi}{d}}$. Further, these latter eigenvalues are simple. It follows from this and (6) that

$$\mathbb{P}(R > n) \sim \sum_{m=0}^{d-1} c_m (\rho u^m)^n, n \text{ large} \quad (20)$$

A few manipulations show that this is equivalent to $\mathbb{P}(R > n) \sim \varphi(n) e^{-\beta n}$, where φ is the trigonometric polynomial $\varphi(n) = \sum_{m=0}^{d-1} c_m u^{mn}$ and $\beta = -\log(\rho) > 0$.

Now we relax the assumption that Q is irreducible. From the general structure of non-negative matrices in [3], Eq. Appendix (4.2), we can relabel the states such that Q has the form

$$Q = \begin{pmatrix} P_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k & 0 & \cdots & 0 \\ T_{k+1,1} & T_{k+1,2} & \cdots & T_{k+1,k} & P_{k+1} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ T_{m,1} & T_{m,2} & \cdots & T_{m,k} & T_{m,k+1} & \cdots & P_m \end{pmatrix}$$

where the letters represent blocks, and the diagonal blocks P_k are square and irreducible non-negative matrices. Let ρ_k be the spectral radius of P_k ; the hypothesis that the chain X_n is irreducible implies that the min row sum of P_k , for any k , is less than 1, and the max is ≤ 1 . It follows [16] (2.13 p. 49) that $\rho_k < 1$. The spectral radius of Q is $\rho = \max_k \rho_k$ thus this shows in passing that $\rho < 1$. Let also d_k be the periodicity of the graph underlying P_k and $u_k = e^{\frac{2i\pi}{d_k}}$. The spectral decomposition of Q is entirely defined by that of the blocks P_k , thus:

$$\mathbb{P}(R > n) \sim \sum_k \sum_{m=0}^{d_k-1} c_{m,k} (\rho_k u_k^m)^n, n \text{ large} \quad (21)$$

Further, let $K_0 = \{k : \rho_k = \rho\}$. We can simplify (21) by removing all indices k not in K_0 . so that all ρ_k 's other than ρ disappear. This shows the asserted lemma.