

# On the Levy-walk Nature of Human Mobility: Do Humans Walk like Monkeys?

Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee and Song Chong

**Abstract**—We report that human walk patterns closely follow Levy walk patterns commonly observed in animals such as monkeys, birds and jackals. Our study is based on about one thousand hours of GPS traces involving 44 volunteers in various outdoor settings including two different college campuses, a metropolitan area, a theme park and a state fair. Important implications of this finding include that many statistical features of human walks are scale-invariant and bursty, and do not conform to the central limit theorem. None of commonly used mobility models for mobile networks captures these properties. Levy walks are more diffusive than Brownian motion (BM) while less diffusive than random way point (RWP). Based on these findings, we construct a simple Levy walk mobility model that emulates human walk patterns expected in outdoor mobile network environments. We demonstrate that the Levy walk model can be used to recreate the statistical patterns commonly observed in previous mobility studies such as the power-law distributions of human inter-contact times and that the simulation performance of mobile network routing protocols under the Levy walk model exhibits distinctive performance features unexplored under existing mobility models.

## I. INTRODUCTION

Do humans walk like monkeys? It is not about upright walks with two hind legs that are similar, but about the statistical patterns of their mobility. Why are they important? Although they may answer important biological and social scientific questions about human activities, our purpose of studying human mobility patterns is their use in simulating mobile networks formed by wireless devices carried by people. Mobile networks are inherently cooperative as mobile devices rely on nearby nodes to maintain network connectivity or relay messages. Therefore, the underlying mobility patterns of mobiles strongly influence the performance of mobile network protocols. As wireless devices are often attached to humans, understanding their mobility patterns leads to more realistic network simulation and more accurate understanding of the performance of the protocols therein. Commonly used mobility models are random way point (RWP) or random walk models such as i.i.d. mobility, Brownian motion, and Markovian mobility. These models are simple enough to be theoretically tractable and at the same time, to be emulated in network simulators in a scalable manner. However, there has been little statistical validation of such models for the accuracy in describing human mobility.

Now, back to the first question. Does human mobility have similar statistical patterns as monkeys? This paper provides statistical evidence that humans in outdoor settings produce similar mobility patterns as spider monkey, albatrosses (seabirds), jackals or even highly charged electron particles. What all these have in common is that their mobility patterns

are shown to resemble what physicists have long called *Levy Walks*. The term Levy walks was first coined by Schlesinger et al. [1] to explain atypical particle diffusion not governed by Brownian motion (BM). BM characterizes the diffusion of tiny particles with a mean free path (or flight) and a mean pause time between flights. A *flight* is defined to be a longest straight line trip from one location to another that a particle makes without a directional change or pause. Einstein [2] first showed that the probability that such a particle is at a distance  $r$  from the initial position after time  $t$  has a Gaussian distribution and thus is proportional to  $\sqrt{t}$ , i.e., the width or standard deviation of a Gaussian distribution. The mean squared displacement (MSD), which is defined to be the variance of the probability distribution, is proportional to  $t$ . It is a manifestation of the central limit theorem (CLT) as the sum of flight lengths follows a Gaussian distribution. (This work by Einstein later led to the discovery of atoms and Avogadro's numbers.)

However, when flight lengths do not have a characteristic scale - in other words, their second moment is not finite, the particles are making Levy walks and may undergo atypical diffusion. This implies that the mean squared displacement of particles making Levy walks is proportional to  $t^\gamma$  where  $\gamma > 1$  - thus CLT does not hold. Intuitively, Levy walks consist of many short flights and exceptionally long flights that eliminate the effect of such short flights. Sample trajectories of an object undergoing BM, Levy walks and RWP are presented in Fig. 1. We can clearly see the differences in the mobility patterns.

Levy walks are also found in animal foraging patterns. Viswanathan et al. [3] show that the foraging patterns of albatrosses can be described by Levy walks. The similar patterns are also discovered in jackals [4] and spider monkeys [5]. The authors conjecture that the Levy walk patterns of these animals are caused by the power-law distribution of prey and food sources. It is also known that Levy walks are an optimal way to find randomly dispersed objects [6]. Thus, these animals perform Levy walks for survival!

In this paper, we statistically establish that the mobility patterns of humans strongly resemble (truncated) Levy walks. We use mobility track logs obtained from 44 participants carrying GPS receivers from September 2006 to January 2007. The sample settings where traces are obtained are two university campuses (one in Asia and one in the US), one metropolitan area (New York city), one State fair and one theme park (Disney World). The participants walk most of times in these locations and may also occasionally travel by bus, trolley, cars, or subway trains. These settings are selected because they are conducive to collecting GPS readings as they

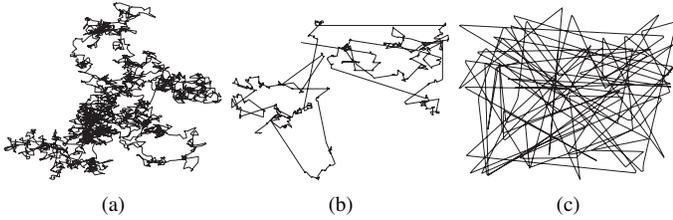


Fig. 1. Sample trajectories of (a) Brownian motion, (b) Levy walk and (c) Random way point

are outdoor. GPS receivers require line of sights with satellites to obtain position information.

Our analysis of the obtained traces reveals that the mobility patterns of the participants in these outdoor settings closely resemble those of Levy walks; their flight distributions and pause time distributions closely follow (truncated) power-law distributions. Their mean squared displacement shows the strong influence of these mobility patterns. We also confirm several other scale-free features of human mobility from these traces. To the best of our knowledge, this is the first study that proves the Levy walk nature of human walk mobility through real walk trace data, and none of the existing mobility models used for mobile network simulations captures the Levy walk characteristics of human walk mobility. Based on the statistical patterns obtained from the traces, we construct a simple Levy-walk model for use in mobile network simulations and show that the model can be used to create the power law inter-contact time distributions of human walks observed in [7]. We apply the Levy walk models to mobile ad hoc network (MANET) and delay-tolerant network (DTN) simulation and study the performance impact of Levy walks on routing performance in MANETs and DTNs. Our study reveals distinctive routing performance features manifested by Levy walks.

This paper is organized as follows. Section 2 provides preliminary background on Levy walks, Section 3 discusses our data collection and analysis techniques, Section 4 presents our main result – the statistical analysis of mobility traces to establish that human walks exhibit Levy walk characteristics, Section 5 presents a simple Levy walk model that can be used for mobile network simulations, and Section 6 contains our study on routing performance using the Levy walk model. Sections 7 and 8 contain related work and conclusion.

## II. PRELIMINARY

Consider a 2-dimensional random walk defined by a sequence of *steps* that a walker makes. A step is represented by a tuple  $S = (l, \theta, \Delta t_f, \Delta t_p)$  in which a walker makes a flight followed by a pause:  $\theta$  is the direction of that flight,  $l > 0$  is the length of the flight,  $\Delta t_f > 0$  is the time duration of the flight or *flight time*,  $\Delta t_p \geq 0$  is the time duration of the pause or *pause time*. At the beginning of each step, a walker chooses a direction randomly from a uniform distribution of angle within  $[0, 360]$ , a finite flight time randomly based on some distribution, and its flight length and pause time from probability distributions  $p(l)$  and  $\psi(\Delta t_p)$ , respectively. During a pause, a walker does not move from the location where the

current flight ends. The time elapsed during a step is called a *step time*  $\Delta t_s$ , which is the summation of its flight time and pause time. The walker starts its first step at the origin at time  $t = 0$ .

After some time  $t$ , the distance  $\text{dis}(t)$  of the random walker from the origin follows a Gaussian distribution with its width proportional to  $\sqrt{t}$  if the variance of flight lengths and the mean of step times are both finite. To see this, consider a one dimensional random walk where  $p(l)$  has a finite variance  $\sigma^2$  and the step time distribution has a finite mean  $\tau$ . The position  $x$  of a random walker after  $N$  steps can be described as a total sum of each flight. According to CLT, the scaled position  $y$  after making  $N$  steps,  $x/\sqrt{N}$ , obtains a Gaussian probability density function,  $f_Y(y, N)$ , in the limit  $N \rightarrow \infty$ .

$$\lim_{N \rightarrow \infty} f_Y(y, N) = f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} \quad (1)$$

From Eq. 1, we know that the position  $x$  of the random walker also follows a scaled Gaussian distribution. Translating  $N$  into  $t$  using  $N \approx t/\tau$ , we can see without much manipulation that the probability density function  $f_X(x, t)$  of position  $x$  after time  $t$  has also a Gaussian form known as a diffusion equation:

$$f_X(x, t) \sim \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \quad (2)$$

where  $D = \sigma^2/2\tau$ , known as a *diffusion constant*. Since the walker starts its walk from the origin,  $\text{dis}(t) = x$ . This type of diffusion is called *normal diffusion*.

Eq. 2, indicates that the MSD of normal diffusion, i.e., the variance of  $\text{dis}(t)$ , grows linearly with time  $t$ . However, CLT is no longer valid if the variance of flight lengths is infinite [8], [9]. One distribution for which the variance diverges is an inverse power-law distribution:  $p(l) \sim \frac{1}{l^{1+\alpha}}$ ,  $0 < \alpha < 2$ . The positions of the random walker with such a distribution of flight lengths converge to another distribution, called *Levy stable distribution* with a coefficient  $\alpha$  [9]. Such random walks are named as *Levy walks* [1]. The scale-free distribution of flight lengths leads to *super-diffusion* where MSD is proportional to  $t^\gamma$ ,  $\gamma > 1$ .

Levy walks are often accompanied by power law pause times. Such a random walk is called *Levy walk with trapping* where the motion can be either *super-diffusive* (i.e.,  $\gamma > 1$ ) or *sub-diffusive* (i.e.,  $\gamma < 1$ ), depending on the distributions of flight lengths and pause times. We denote a power law distribution of pause times by  $\psi(\Delta t_p) \sim 1/\Delta t_p^{1+\beta}$ .

## III. MEASUREMENT METHODOLOGY

### A. Data collection

Five sites are chosen for collecting human mobility traces. These are two university campuses (Campus I in the US and Campus II in Asia), New York City, Disney World, and one state fair (in the US). The total number of traces from these sites is over 150 daily traces. Garmin GPS 60CSx handheld receivers are used for data collection which are WAAS (Wide Area Augmentation System) capable with a position accuracy of better than three meters 95 percent of the time, in North



Fig. 2. Sample GPS traces from the Disney World scenario.

Site (# of participants)	# of traces	Duration (hour)			Radius (km)		
		min	avg	max	min	avg	max
Campus I (20)	35	1.71	10.19	21.69	0.46	1.82	5.84
Campus II (4)	46	4.21	10.62	22.37	0.43	1.26	4.16
NYC (8)	30	1.23	9.34	22.66	0.37	4.18	6.98
DW (4)	15	4.43	8.68	13.20	0.39	1.67	4.43
SF (8)	8	1.81	2.57	3.12	0.22	0.28	0.34

TABLE I  
STATISTICS OF COLLECTED MOBILITY TRACES FROM FIVE SITES.

America [10]. Occasionally, track information has discontinuity mainly when bearers move indoor where GPS signals cannot be received. The GPS receivers take reading of their current positions at every 10 seconds and record them into a daily track log. The summary of daily traces is shown in Table I. The radius of each trace is a half of the maximum distance that a participant travels during a day. Fig. 2 shows sample GPS traces from the Disney World scenario.

The participants in Campus I were randomly selected students who took a course in the computer science department. Every week, 2 or 3 randomly chosen students carried the GPS receivers for their daily regular activities. The Campus II traces are taken by 4 students who live in a campus dormitory. Since the participants in Campuses I and II occasionally moved outside their campuses, we use only those logs recorded within a radius of 10 km from the center of each campus. The New York City traces were obtained from 8 volunteers living in Manhattan or its vicinity. Most of the participants have offices in Manhattan. Their track logs contain relatively long distance travels because of their long commuting paths. Their means of travel include subway trains, buses and mostly walking. The State fair track logs were collected from 8 volunteers who visited a local state fair that includes many street arcades, small street food stands and showcases. The event was very popular and attended by more than one thousand people daily for two weeks. The site is completely outdoor and is smallest among all the sites. Each participant in the State fair scenario spent less than three hours in the site. The Disney World traces were obtained from four volunteers who spent their thanksgiving or Christmas holidays in Disney World, Florida, USA. For our study, we use only the track logs from the inside of the theme parks. The participants mainly walked in the parks and occasionally rode trolleys.

## B. Trace analysis

From the traces, we extract the following data: flight length, pause time, direction, and velocity. To get these data from the traces, we map the traces into a two dimensional area (note that the GPS receivers produce three-dimensional positions), and to account for GPS errors, we clean the data as follows. We recompute a position at every 30 seconds by averaging three samples over that 30 second period (note GPS samples are taken at every 10 seconds). All the position information discussed below is based on the 30-second average positions.

As participants may move outside a line of sight from satellites or run out of battery, daily traces may contain discontinuities in time. For instance, if a participant disappears at time  $t$  (in seconds) at a position  $p$  from a trace and reappears at time  $t + \Delta t$  at another position  $p'$ , we use a similar method used in [11] to remove the discontinuity. If the next position recorded after the discontinuity is within a radius of 20 meters and the time to the next position is within a day boundary, then we assume that the participant walks to the next position from position  $p$  at a walking speed of 1 m/s from time  $t + \Delta t - k$  ( $k$  is the distance between  $p$  and  $p'$  in meters) just before he shows up again at position  $p'$  in the trace and the remaining time ( $\Delta t - k$ ) is recorded as a pause at the location where he disappeared. Otherwise, it is assumed that the trace has ended at time  $t$  and a new trace starts at time  $t + \Delta t$ .

We consider that a participant has a pause if the distance that he has moved during a 30 second period is less than  $r$  meters. To extract flights in a trace of one participant, we use three different methods, namely *rectangular*, *angle* and *pause-based* models. In the rectangular model, given two sampled positions  $x_s$  and  $x_e$  taken at times  $t$  and  $t + \Delta t$  ( $\Delta t > 0$ ) in the trace, we define the straight line between  $x_s$  and  $x_e$  to be a flight if and only if the following conditions are met. (a) the distance between any two consecutively sampled positions between  $x_s$  and  $x_e$  is larger than  $r$  meters (i.e., no pause during a flight), (b) when we draw a straight line from  $x_s$  to  $x_e$ , the sampled positions between these two end points are at a distance less than  $w$  meters from the line. The distance between the line and a position is the length of a perpendicular line from that position to the line, (c) for the next sampled position  $x'_e$  after  $x_e$ , positions and the straight line between  $x_s$  and  $x'_e$  does not satisfy conditions (a) and (b). An example of the rectangular model is shown in Fig. 3. In that figure, the straight line movement between positions sampled at times  $t(1)$  and  $t(4)$  is regarded as one single flight between the two positions because all the sampled positions between them are inside of the rectangle formed by the two end points. In this example, the flight time is 90 seconds because each sample is taken at every 30 seconds. By controlling  $w$ , we can obtain very “tight” flight information. Both  $r$  and  $w$  are model parameters.

The angle model allows more flexibility in defining flights. In the rectangular model, a trip can be broken into small flights even though consecutive flights have similar directions. This implies even a small curvature on the road may cause multiple short flights. To remedy this, the angle model merges multiple

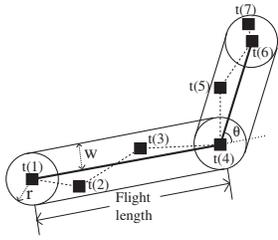


Fig. 3. The rectangular model used to extract flight information from traces.

successive flights acquired from the rectangular model into a single long flight if the following two conditions are satisfied: (a) no pause occurs between consecutive flights and (b) the relative angle ( $\theta$  as shown in Fig. 3) between any two consecutive flights is less than  $a_\theta$  degree. A merged flight is considered to be a straight line from the starting position of the first flight to the ending position of the last flight and its flight length is the length of that line.  $a_\theta$  is a model parameter.

The pause-based model can be viewed as an extreme case of the angle model. The pause-based model merges all the successive flights from the rectangular model into a single flight if there is no pause between the flights. A merged flight is defined in the same way as in the angle model. This model produces significantly different trajectories from the actual GPS trajectories, due to the abstraction. However, it represents more faithfully human intentions to travel from one position to another without much deviation caused by geographical features such as roads, buildings and traffic.

The rectangular and pause-based models can be viewed as special cases of the angle model with  $a_\theta = 0$  and  $a_\theta = 360$ , respectively. Fig. 4 presents sample traces produced by the above three flight models. The trajectories become more simplified as the flight model changes from the rectangular model to the pause-based model.

#### IV. HUMAN MOBILITY

In this section, we statistically establish that human walks resemble Levy walks. Below, we first examine the distributions of flight lengths and pause times, and their impact on the mean squared displacement and show that these features exhibit scale-free characteristics.

##### A. Flight length distribution

A power-law distribution of flight lengths is a hallmark of Levy walks. In this section, we study the distributions of flight lengths from our traces to verify that they follow a power-law distribution. In generating its distribution for each scenario, flight length samples from all the traces of the same site, regardless of their participants, are aggregated together and used in the same distribution. If we can accept that every participant in the same site has the same statistical mobility tendency, this ‘‘aggregation’’ is reasonable because every trace obtained from the same site is subject to the same or similar geographical constraints (i.e., roads, obstacles, traffic, and buildings). The same technique is used in other studies of Levy walks (e.g., [5]).

Fig. 5 shows the log-log distribution plots of flight lengths sampled according to the three different flight models ( $r = w = 5$ , and  $a_\theta = 30$ ) from the KAIST traces. Fig. 6 shows the same for the other scenarios under the angle model with  $a_\theta = 30$ . We perform line fitting on all the distributions using least squares matching and the figure shows their slopes. To see the effect of flight model parameters on the distribution patterns, we vary the values of  $r$ ,  $w$  and  $a_\theta$  from 2.5 meters to 10 meters and from 15 degrees to 90 degrees, respectively. We fit lines on the resulting log-log plots of flight length distributions over two ranges: (a) from the median to 99.9% quantile and (b) from the median to 99.99% quantile. Table II presents the average slope of the lines and their standard deviation of the fitted lines. All the scenarios have power law slopes as their slopes are larger than -3 (so  $\alpha < 2$ ). With

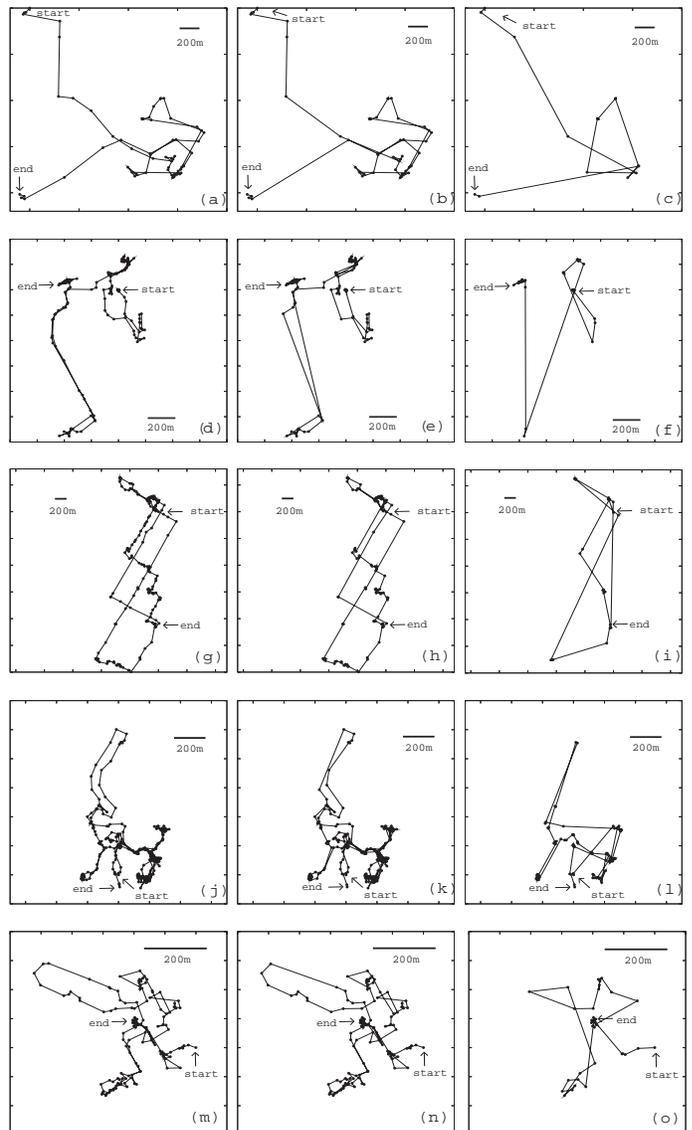


Fig. 4. Traces from Campus I (a)-(c), Campus II (d)-(f), New York City (g)-(i), Disney World (j)-(l) and State fair (m)-(o). The first column represents the rectangular model with  $r = w = 5$ , the second column the angle model with  $a_\theta = 30$ , and the third column the pause-based model.

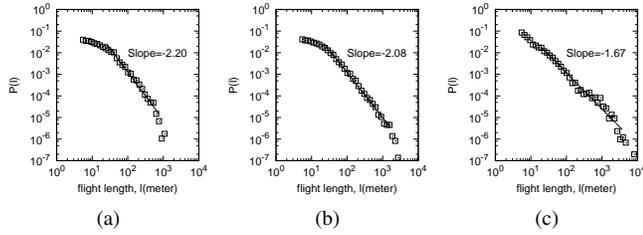


Fig. 5. The flight length distribution of KAIST in a log-log scale with logarithmic bin sizes. (a) rectangular model ( $r = w = 5$ ), (b) angle model ( $a_\theta = 30$ ) and (c) pause-based model.

	Rectangular	Angle	Pause-based
NCSU	-1.81 (0.19)	-1.84 (0.05)	-1.59 (0.08)
KAIST	-2.31 (0.13)	-2.10 (0.10)	-1.67 (0.09)
NYC	-1.75 (0.10)	-1.66 (0.06)	-1.35 (0.09)
Disney World	-2.31 (0.07)	-2.15 (0.10)	-1.82 (0.06)
State fair	-2.40 (0.34)	-1.93 (0.21)	-1.58 (0.27)

TABLE II

THE AVERAGE SLOPE OF THE LINE (WITH STANDARD DEVIATION) FITTED TO THE LOG-LOG PDF OF FLIGHT LENGTHS OBTAINED BY VARYING FLIGHT PARAMETERS:  $r$  AND  $w$  FROM 2.5 METERS TO 10 METERS AND  $a_\theta$  FROM 15 DEGREES TO 90 DEGREES. NOTE THAT  $\alpha$  IS ONE LESS THAN THE ABSOLUTE VALUE OF THE SLOPE.

exception of State fair, all distributions have a heavy tail up to a few kilometers. Furthermore, varying the flight models does not alter significantly the power-law tendency of these distributions. With State fair, although the slopes indicate a power-law tendency, the distributions show a sign of excessive truncations because of the small area of the site and also heavy people traffic in the area. We discuss this more later.

We find that the slope of the fitted line also decreases as the flight definition changes from the rectangular model to the pause-based model, indicating the heaviest tails for the pause-based model. This is a natural consequence of truncations due to geographical constraints. As the rectangular model is most sensitive to changes in moving directions, its flights contain a smaller number of long flights – in most cases, the truncation occurs around a few hundred meters to 1 km. On the other hand, the pause-based model is least affected by these constraints since any directional changes without pauses are considered as part of a single flight. Therefore, when truncations have less impact on flight lengths, the mobility of our participants has a stronger power-law tendency. It implies that human intentions and activities for mobility, independent of geographical constraints, are still scale-free. This point is important for mobility simulation since it is human intentions, not geographical obstacles, that have likely caused the power-law tendency. Thus although geographical constraints may vary in different scenarios, this power-law tendency is invariant. For network simulations involving human-assisted mobile networks, while human navigation around obstacles and road shapes is relatively easy to program, the power-law tendency of human intentions must be built in the mobility generation model to accurately depict human walk patterns.

To further confirm the power-law tendency of the measured

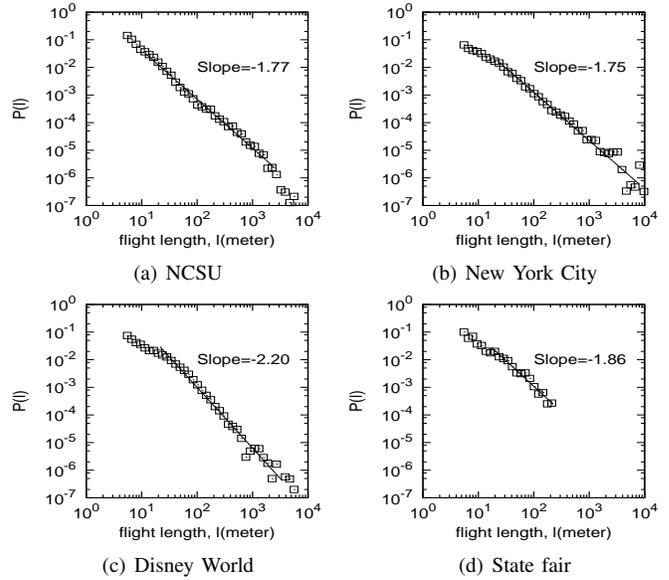


Fig. 6. The flight length distributions of human walks in a log-log scale with logarithmic bin sizes, using the angle model ( $a_\theta = 30$ ).

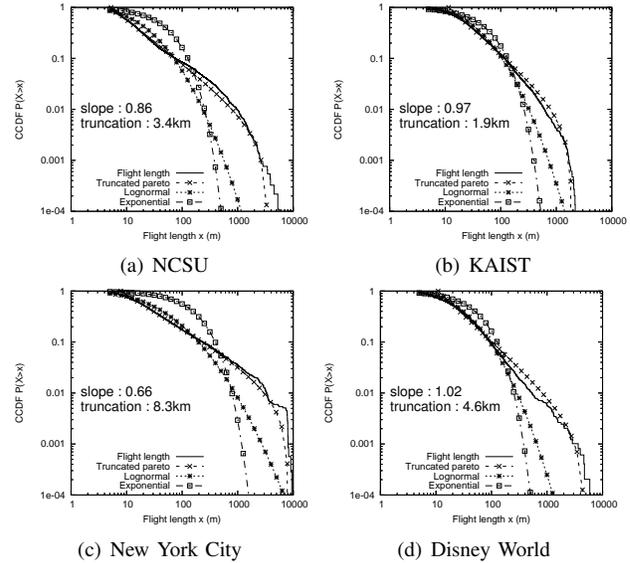


Fig. 7. The CCDF of flight lengths. Various known distributions are fitted to the measured data distribution using maximum likelihood estimation.

flight length distributions, we use Maximum Likelihood Estimation (MLE) to fit three known distributions, exponential, log-normal, and truncated Pareto distributions [12] to the CCDF (complementary cumulative density function) of the measured flight lengths. Fig. 7 shows the result of MLE according to the angle model. The MLE of the truncated Pareto is performed over the  $x$ -axis range between 50 meter and the 99.9% quantile of each distribution to isolate only the tail behavior. We observe that truncated Pareto has the best fit among the three distributions in all cases. Table II shows that the power-law slopes from the MLE of truncated Pareto are within the error ranges of those obtained directly from their PDFs. (Note that the slopes of CCDF and PDF naturally differ by one).

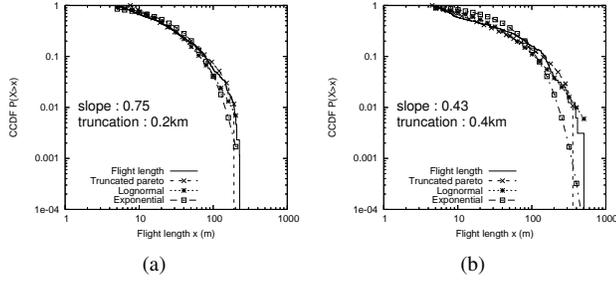


Fig. 8. The CCDF of flight lengths from the State fare traces. We use the angle model with  $a_\theta = 30$  in (a) and the pause-based model in (b).

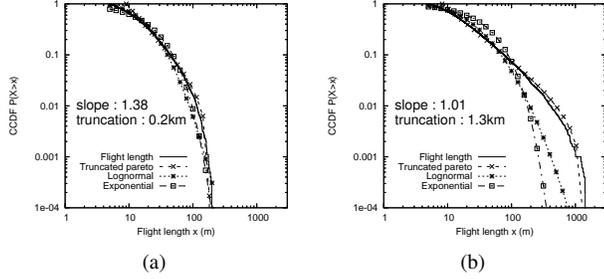


Fig. 9. The CCDF of flight lengths obtained from Levy walk simulations in two squares: one with width 200 meters and the other with 2 km. The Levy walk in the smaller area appears like Brownian motion.

The flight length distribution of State fair in Fig.8 appears close even to a short-tailed distribution such as exponential or lognormal distributions. We conjecture the followings for this. The State fair traces are obtained from a highly confined area of less than 350 meter radius (it is smallest among the five sites). Thus, it is subject to more truncations. Furthermore, as the area is highly crowded, people have to make frequent changes of direction during their walks. These factors induce short flights and discourage long flights. Even in the pause-based model (Fig. 8(b)), truncated Pareto appears to have only slightly better fitting than the other distributions. To verify that Levy walks may appear like BM in a small confined area of square, we simulate two instances of Levy walks, one with width 200 meters and the other with 2 km. Fig. 9 shows that the Levy walk in the small area has the same truncation problem as State fair and the flight length distribution can fit well even to a short-tailed distribution. But when we increase the area, the same Levy walk has a heavy tail.

### B. Pause time distribution

The definition of Levy walks does not require a power-law distribution of pause times. But we find from our traces that the pause times of our walkers have scale-free characteristics. Fig. 10 shows the CCDF of the pause-time distributions extracted from our traces. The flight definitions do not make impact on the shape of pause time distributions because they differ mostly in the number of zero pause time. Even when we vary  $r$  in the pause time definition, we do not see much difference in the pause time distribution patterns. In the plots, we use the pause-based models. All the pause time distributions, except that from State fair, show the best fit with truncated Pareto. In most scenarios, truncations for pause times are less emphatic

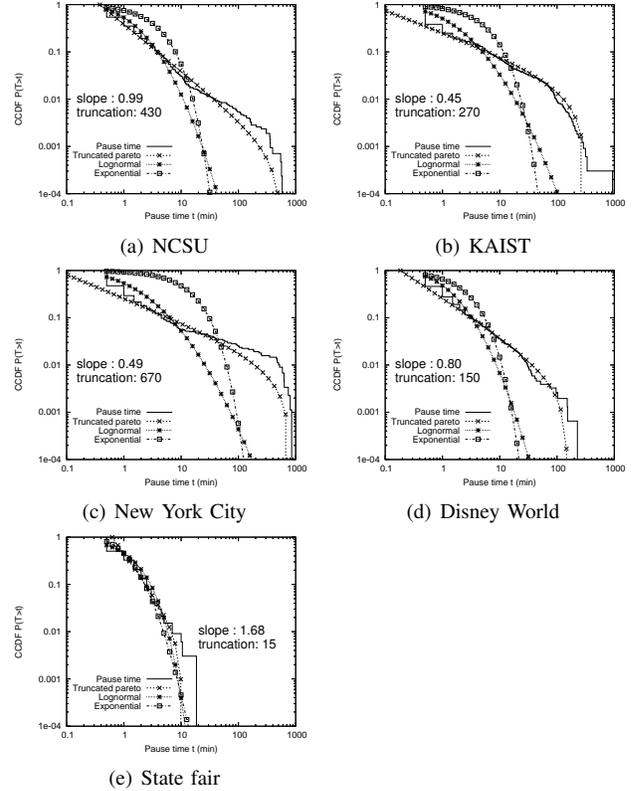


Fig. 10. The pause time CCDF of human walkers in various scenarios along with the MLE of various known distributions.

than for flight lengths. However, State fair does not follow this and its distribution shows a good fit even with an exponential distribution. We conjecture that this is because of the setting that consists of many small shopping and game arcades close to each other. In this setting, participants tend to make many short stops, and furthermore, big crowds and many showcases in the setting prevented them from staying at one location for a long time.

Power-law pause time distributions affect the MSD of walkers as shown in [13], [14] – long trapping caused by heavy-tail pause-time distributions makes the mobility less diffusive, sometimes causing sub-diffusion. We have more details on this next section.

### C. Mean Squared Displacement

Scale-free mobility leads to abnormal diffusion where MSD does not grow linearly with time. Measuring MSD from real mobility traces is not straightforward because it is hard to define the “origin” from the traces. A common technique to handle this is to take average of MSD values measured by varying the origin among all locations that the walker has been at [15], [16]. Specifically, for each scenario, we compute the following. Given each trace  $T$  from that scenario that consists of an ordered sequence of location samples  $(t_0, \text{pos}_T(t_0))$  where  $\text{pos}_T(t_0)$  is the two dimensional position of the walker at time  $t_0$  in trace  $T$ , the MSD( $t$ ) of that scenario in terms of

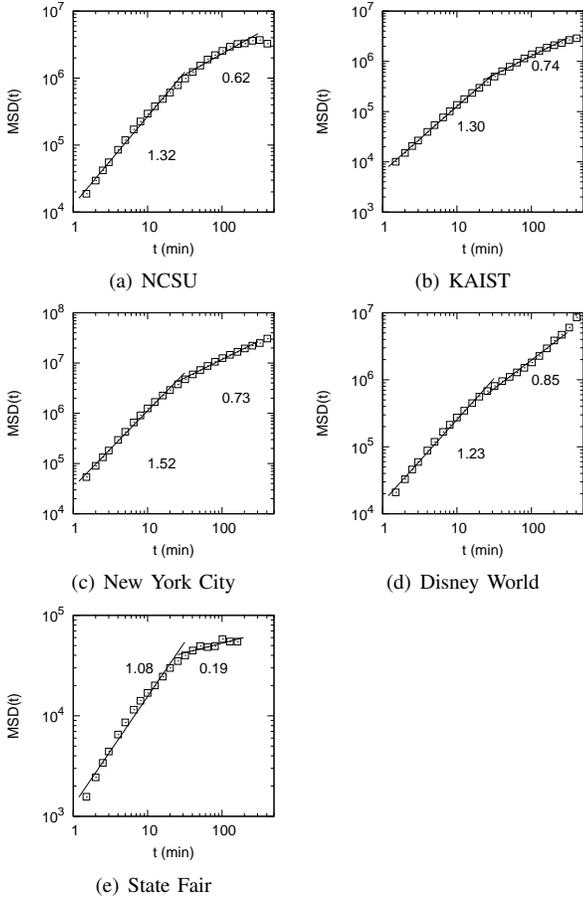


Fig. 11. MSD from various settings.

time interval  $t$  is:

$$\text{MSD}(t) = \frac{\sum_T \sum_{t_0} |\text{pos}_T(t+t_0) - \text{pos}_T(t_0)|^2}{N} \quad (3)$$

$\text{pos}_T(t+t_0) - \text{pos}_T(t_0)$  is a vector subtraction and  $|\cdot|$  is the norm operator.  $N = \sum_T n(T)$  where  $n(T)$  is the total number of *eligible* samples  $t_0$  from trace  $T$ . A sample taken at time  $t_0$  is eligible if  $t_0 + t < t_e^T$  where  $t_e^T$  is the time that the last sample of trace  $T$  is taken. If  $t+t_0 > t_e^T$ , the contribution of  $t_0$  to  $\text{MSD}(t)$  is zero. We compute  $\text{MSD}(t)$  directly from the GPS traces mapped to the two dimensional space.

Fig. 11 plots the  $\text{MSD}(t)$  for all settings. The shape of  $\text{MSD}(t)$  in a log-log scale can be fitted by two lines using least squares matching. From the plots, we can see that up to about 30 minutes, our participants make super-diffusion ( $\gamma > 1.2$ ) and after that, they make sub-diffusion ( $\gamma < 0.9$ ). We can explain this behavior as follows. As we increase time  $t$ , we are increasing the scale of aggregation (note that  $\text{pos}_T(t+t_0) - \text{pos}_T(t_0)$  is a result of summing all the displacement vectors over the trace segment between the two positions). When the scale is small, the effect of truncations does not appear so the flight lengths follow a power-law distribution very well. However, as we increase the scale, the truncation takes effect and the flights become close to Gaussian. As we look at flights from a far distance, the number of long flights visible

	Rectangle	Angle	Pause-based
NCSU	1.32 (0.002)	1.39 (0.043)	1.58 (0.032)
KAIST	1.31 (0.002)	1.36 (0.037)	1.60 (0.040)
NYC	1.52 (0.001)	1.59 (0.041)	1.77 (0.019)
Disney World	1.24 (0.003)	1.31 (0.047)	1.61 (0.047)
State fair	1.10 (0.008)	1.15 (0.047)	1.34 (0.081)

TABLE III  
THE AVERAGE VALUE OF  $\gamma$  (WITH STANDARD DEVIATION) OBTAINED BY VARYING FLIGHT PARAMETERS ( $r$  AND  $w$  FROM 2.5 METERS TO 10 METERS, RESPECTIVELY, AND  $a_\theta$  FROM 15 TO 90).

at that distance decreases very fast because of truncations. Thus, when  $t$  is small (in our case, less than 30 minutes), the effect of heavy tailed distributions shows up and the mobility appears super-diffusive. But when  $t$  is large, the flight lengths follow Gaussian and the mobility is close to that of BM. This point was observed in [16]. When Gaussian flight lengths are combined with power-law pause times, it is shown in [13] that the mobility appears sub-diffusive. Another significant factor causing the sub-diffusion is the human tendency to return to the original starting points. Humans are not truly making random walks and they come home in the end of day or come back to one point (like entrance and exit in Disney world). This “homecoming” tendency slows down diffusion excessively, resulting in sub-diffusion.

We also find that when flight lengths and pause times are close to Gaussian,  $\gamma$  is close to one. This is observed in the State fair traces.

[13], [14] show a theoretical relation between the shapes of flight length and pause-time distributions (i.e.,  $\alpha$  and  $\beta$ ) and MSD (i.e.,  $\gamma$ ). With some manipulation of MLE fitting ranges, we are able to verify the relation with our trace statistics. However, it is not clear whether that result has any significance because (a) it is hard to measure  $\alpha$  and  $\beta$  accurately due to various factors including truncations, the range of power-law fitting, and differing flight length definitions, and (b) the theoretical relation assumes constant velocity which is not the case with our traces. So we do not present the result here. Instead, we verify further that for the period before the inflection point all traces exhibit a super-diffusive characteristic, by measuring MSD for that super-diffusive regime while varying the flight definitions on our traces. Table III presents the average MSD with standard deviation. In this table, we can also see a clear tendency that human mobility is more super-diffusive as we reduce the effect of geographical constraints: in the pause-based model,  $\gamma$  tends to be higher.

#### D. Mean squared fluctuation of displacement

The scale free behaviors of animal mobility can also be explained by long range correlations in the movement time series data such as the number of times per hour a jackal moves its position [4], and the number of intervals per hour albatrosses are stationary [3]. This metric is called the *root mean squared fluctuation of displacement* (RMS). RMS is computed as follows.  $u(i)$  is defined to be the number of flights per the  $i$ -th unit interval. The unit interval is arbitrary chosen to

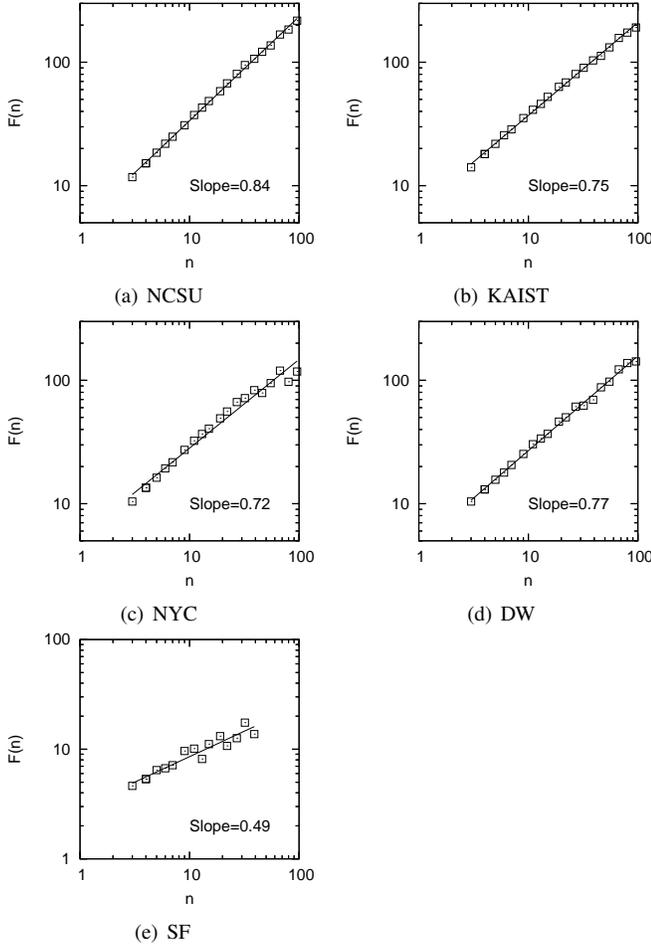


Fig. 12. RMS on the number of flights.

include some number of flights. The total displacement  $w(n)$  is defined by a running sum  $w(n) = \sum_{i=1}^n u(i)$ . Then, RMS  $F(n)$  is calculated as follows.

$$F(n) = \sqrt{\langle (\Delta w(n))^2 \rangle - \langle \Delta w(n) \rangle^2} \quad (4)$$

where  $\Delta w(n) = w(n_0+n) - w(n_0)$ , and  $\langle \cdot \rangle$  denotes averaging over all possible  $n_0$ 's.

$F(n)$  is significant because only in the presence of long range correlations with no characteristic time scales,  $F(n) \sim n^d$ ,  $d \neq 1/2$  holds [3]. Uncorrelated or short-term correlated data gives  $d = 1/2$  because of CLT. In this paper, we measure  $u(i)$  as the number of flights per 10 minutes, and analyze whether the long range correlation exists in the time series of flight samples.

Fig. 12 shows the RMS on the number of flights (defined based on the angle model with  $a_\theta = 30$ ) from NCSU and NYC traces. They show RMS coefficient  $d$  larger than 0.7 except State fair. State fair does not show much long-term correlation in the bursty nature of flights mainly due to truncations. These results (except that of State fair) indicate that like albatrosses and jackals, human mobility is very bursty and contains a similar type of scale-freedom as those animals.

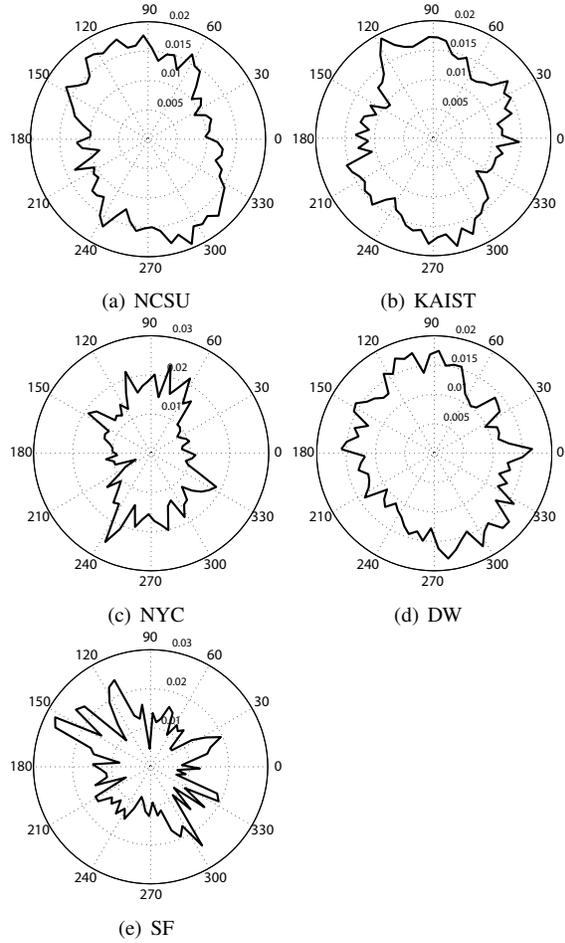


Fig. 13. Direction distribution from human walk traces.

### E. Directions, velocity and auto-correlation

We also study other statistics that are relevant to generating human mobility models. Fig. 13-15 shows statistics on direction, velocity and correlation of flight lengths and directions over time series. These statistics are not explicitly specified in Levy walk models, but are useful in generating human mobility tracks for simulation.

From our data, we find that while most scenarios produce close to a uniform distribution of turning angles similar to Disney world's, the New York City traces have more bias in particular directions mostly in 90 and 270 degrees. This pattern is related to geographical artifacts since Manhattan tends to induce more perpendicular directional changes.) Fig. 13 shows the turning angle distributions from all traces produced based on the angle model with  $a_\theta = 30$ . The angle distributions show the effect of the shapes of geographical constraints. The speed of human mobility has high correlation with flight lengths: velocity increases as flight lengths increase. Constant velocity is a common assumption in Levy walks. Fig. 14 depicts the correlation between flight lengths and velocity. We also measure auto-correlation of flight lengths and turning angles over the time series of flight length and turning angle samples. We find some auto-correlation of flight lengths over up to 10

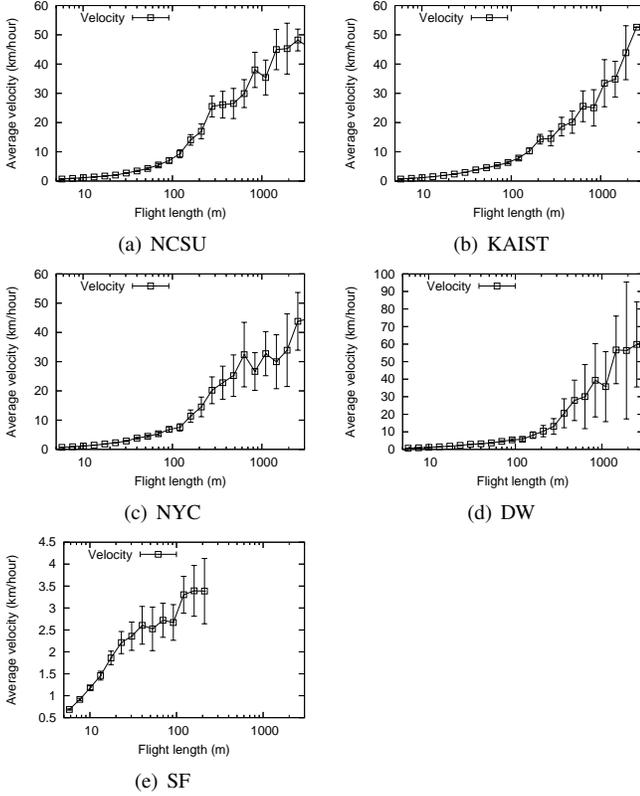


Fig. 14. Velocity distribution from human walk traces.

sample lags while almost no auto-correlation of turning angles (in some cases, we find some negative correlation around one or two lags). We did not find any significant difference of these statistics over different scenarios. Fig. 15 shows representative auto-covariance coefficients. The significant auto-correlation of flight lengths indicate that when small flights are made, there are non-zero preference for similar sizes near future. This pattern cannot be described by random walks (including Levy walks) as they produce flights randomly without any dependency on the past history of flights.

## V. LEVY-WALK MOBILITY MODEL

In this section, we discuss a simple Levy-walk mobility model for simulating human carried mobile networks that generates synthetic mobility tracks reflecting the statistical patterns of human mobility that we find in our study. We use the same random walk model discussed in the section 2. A step is represented by four variables, flight length ( $l$ ), direction ( $\theta$ ), flight time ( $\Delta t_f$ ), and pause time ( $\Delta t_p$ ). Our model picks flight lengths and pause times randomly from their PDFs  $p(l)$  and  $\psi(\Delta t_p)$  which are Levy distributions with coefficients  $\alpha$  and  $\beta$ , respectively. The following defines a Levy distribution with a scale factor  $c$  and exponent  $\alpha$  in terms of a fourier transformation,

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx - |ct|^\alpha} dt \quad (5)$$

For  $\alpha = 1$ , it reduces to a Cauchy distribution and for  $\alpha = 2$ , a Gaussian with  $\sigma = \sqrt{2}c$ . Asymptotically, for  $\alpha < 2$ ,  $f_X(x)$

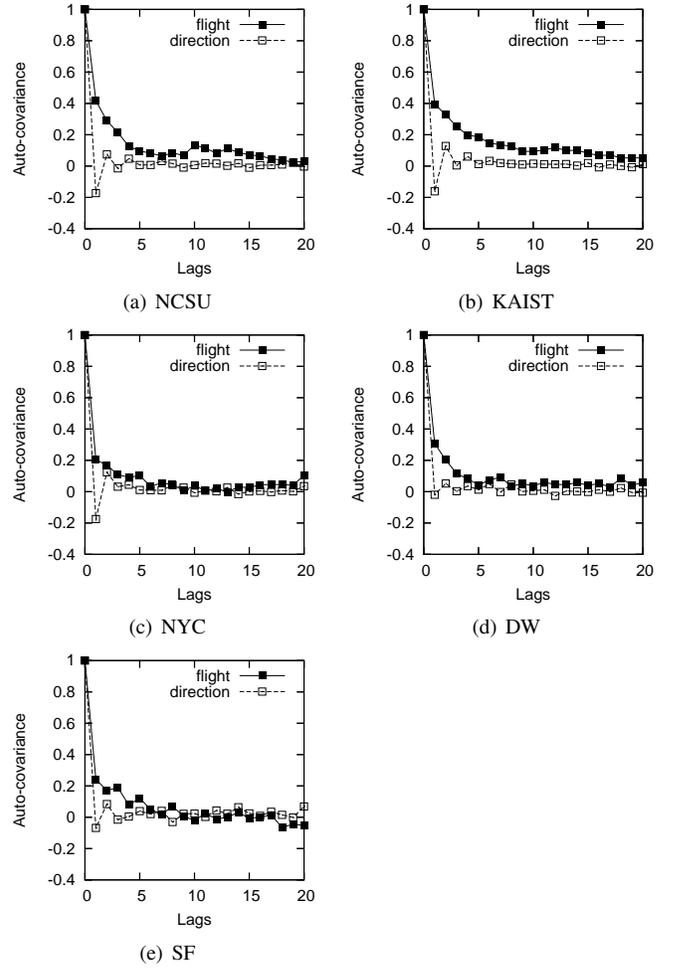


Fig. 15. Flight and direction autocorrelations over a time series

can be approximately by  $\frac{1}{|x|^{1+\alpha}}$ . We allow  $c$ ,  $\alpha$  and  $\beta$  to be simulation parameters.

Remaining parameters to complete the step definition are flight directions and times. Fig. 16 shows the angle distribution of directions, and velocity in terms of flight lengths, both extracted from all the five scenarios. The angle distribution is close to a uniform distribution which our model adopts. Unlike flight directions, flight times and lengths are highly correlated. From Fig. 16(b), we verify that the average velocity is not constant, but increases as flight lengths increase because long flights are usually generated when participants use a transportation rather than walking. To reflect this tendency, our model uses the following relation between flight times and flight lengths:  $\Delta t_f = kl^{1-\rho}$ ,  $0 \leq \rho \leq 1$  where  $k$  and  $\rho$  are constants. In one extreme, when  $\rho$  is 0, flight times are proportional to flight lengths and it models the constant velocity movement. In another extreme, when  $\rho$  is 1, flight times are constant and flight velocity is linearly proportional to flight lengths. In our measurement data, the relation is best fitted with  $k = 18.72$  and  $\rho = 0.79$  when  $l < 500m$ , and with  $k = 1.37$  and  $\rho = 0.36$  when  $l \geq 500m$ .

Based on the above model, we generate synthetic Levy-walk mobility tracks with truncation factors  $\tau_l$  and  $\tau_p$  for

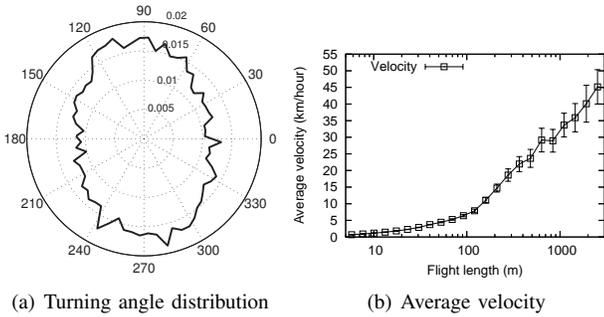


Fig. 16. (a) Circular distribution of turning angle and (b) average velocity depending on flight lengths with 95% confidence intervals, obtained from all the traces.

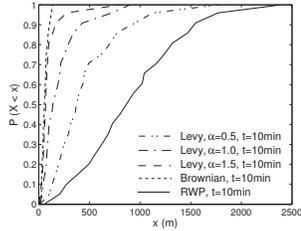


Fig. 17. The CDF of node displacement from its initial position after 10 minute travel. RWP is most diffusive while BM is least diffusive. The diffusion rates of Levy walks are in-between.

flight lengths and pause times respectively in a confined area as follows. First, the initial location of a walker is picked randomly from a uniform distribution in the area. At every step, an instance of tuple  $(l, \theta, \Delta t_f, \Delta t_p)$  is generated randomly from their corresponding distributions. If  $l$  and  $\Delta t_p$  are negative or  $l > \tau_l$  or  $\Delta t_p > \tau_p$ , then we discard the step and regenerate another step. We repeat this process after the step time  $\Delta t_f + \Delta t_p$ . Until the end of the simulation, we generate the tuples repeatedly.

Using the above model, we can generate various types of Levy walk models. When  $\alpha$  and  $\beta$  are 2, then the model becomes Brownian motion. Furthermore by adjusting these parameters between 0 and 2, we can control the diffusive nature of mobility. For instance, Fig. 17 shows the CDF of the distance that a mobile is away from its initial position after the first 10 minute travel. The data are extracted from MATLAB simulation of the mobility models. The simulation area is set to 2 km by 2 km. The Levy walk models are constructed by setting the pause time factor ( $\beta$ ) 0.5 but varying the flight length factor ( $\alpha$ ) from 0.5 to 1.5. We set the truncation points  $\tau_l = 1$  km and  $\tau_p = 1000$  seconds and set the scale factors ( $c$ ) of flight length and pause time distributions to 10 and 1, respectively. The BM model uses the same simulation setup and parameter setting as the Levy walk model but sets  $\alpha = 2$  and  $\beta = 2$ . We use a typical setup for RWP: (a) for each step, it chooses a random destination uniformly within the simulation area, thus implicitly selecting a flight length by choosing the next destination and (b) the pause time is uniformly distributed between 0 and 60 seconds. All models use the same velocity model discussed above and 100 nodes are simulated at the same time. The figure shows that RWP is most diffusive while

BM is least diffusive. The diffusion rates of the Levy walk models are in-between these two extremes. As we reduce  $\alpha$ , the mobility becomes more diffusive. This disparate diffusion rate of mobility in each model has distinctive effects on the performance of routing in mobile networks. We will discuss these effects in the ensuing sections.

### A. Model verification

In this section, we verify whether LW can synthetically generate the statistical features we have observed in our traces. Figs. 19 (a) and (b) show statistical distributions of flights and pause-time matching each scenario (we do not show the matching of NCSU data as it is similar to that of KAIST). To produce these traces, we set the simulation area by the same size of each corresponding scenario. We then vary the values of  $\alpha$  and  $\beta$  to find synthetic traces that have similar flight length and pause-time distributions of each scenario. We do not add any geographical constraints other than the simulation area (i.e., we set  $\tau_l$  to infinity) and any flight that goes outside the area is truncated to that boundary and flights at the boundary are always set toward the inside of the area. We set the truncation of pause time ( $\tau_p$ ) using the same values we obtained from the traces. Our synthetic traces show strikingly similar flight and pause time distributions seen from our real traces. This shows the versatility of our model. Our model does not exactly mimic human walk patterns. This can be seen from the MSD values measured from the synthetic traces. Figs. 19 (c) and (d) show the MSD values from the synthetic traces that matches the New York city and the Disney World trace, respectively. While it shows a similar pattern of super and then sub-diffusion, we cannot match MSD values (while simultaneously matching flight and pause time distributions). This is because human walks are not truly random as our model is and contains various factors that only humans control including context, home-coming tendency, flight auto-correlation, etc. Furthermore, since we do not model the geographical constraints, they can also make a difference. These are weaknesses of our mobility model that requires more refinements.

## VI. ROUTING PERFORMANCE

In this section, we apply the mobility model developed in Section V to the simulation of DTN and MANETs and measure routing performance in these networks and compare resulting routing patterns with those generated from existing models such as RWP and BM.

### A. Routing in Delay Tolerant Networks

In delay tolerant networks (DTN), mobile nodes may establish on and off connectivity with their neighbors and the rest of the network. Therefore, store-and-forward is the main paradigm of routing in such networks where communication transpires only when two devices are in a radio range. We call the time period that two nodes are in a radio range the *contact time* of the two nodes. One of the most widely studied routing algorithms in DTN is *two-hop relay routing* [17] where

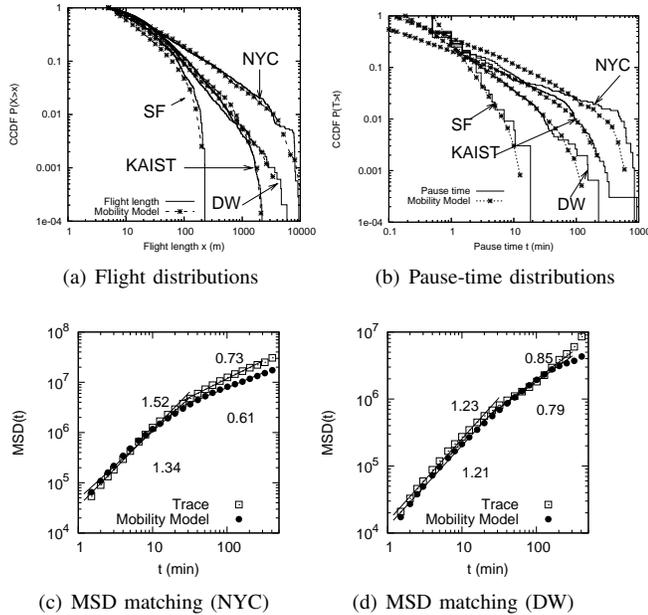


Fig. 18. Our Levy walk model can generate synthetic traces that match the flight and pause time distributions seen from real human walk traces. However, it also shows some discrepancy in the MSD values although its overall shape is similar.

a source node sends a message (or a sequence of data packets) to the first node it contacts and then that first node acts as a relay and delivers the message when it contacts the destination node of the message. Here the period between the time that the message has originated and the time that the message is delivered to the relay node is called *first contact time* (FCT) and the period after that to the time the message is delivered to the destination is called *remaining inter-contact time* (RICT). In a dense network, FCT is typically negligible and RICT dominates the message delay. One way to characterize RICT is to measure the *inter-contact time* (ICT), the time period between two successive contact times of the same two nodes. Since it is difficult to measure RICT from real mobility traces, ICT has been used to characterize RICT [7].

It is known that the ICT of human mobility exhibits a strong power-law tendency [7]. The result is interesting because [18] showed by simulation that RWP produces exponentially decaying ICT, implying human mobility cannot be modeled by RWP. What's not obvious is the type of mobility patterns that gives rise to the power-law tendency of ICT distributions. In this section, we explore this problem using the mobility model from Section V.

The earlier measurement studies on ICT (e.g., [7]) report power-law distributions of ICT with human mobility with slopes in the range of [0.3,0.4]. By varying the parameters of  $\alpha$  and  $\beta$  of our mobility model, we are able to generate ICT distributions with the similar characteristics as in [7] by MATLAB simulation. [7] reports power-law slopes of 0.3 from the INFOCOM trace [19] and 0.4 from the UCSD trace [20]. Fig. 19(a) shows the result. In the UCSD simulation, we fix the simulation area to 3.5 km by 3.5 km,  $\tau_l$  to 3 km and  $\tau_p$  to 28 hours. These values are chosen based on the data

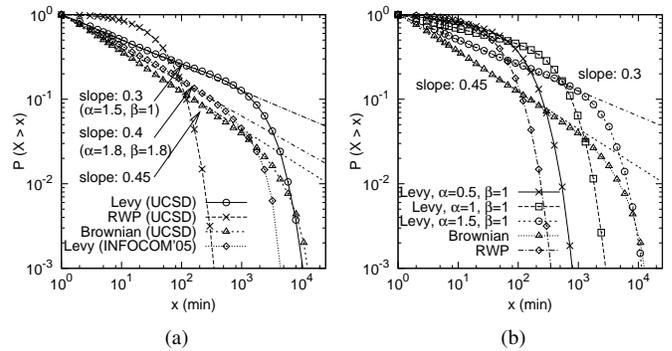


Fig. 19. The ICT distributions of mobility models. Levy walks recreate the ICT distributions seen in the INFOCOM and UCSD traces.

from [20]. The remaining parameters are the same as in the experiment for Fig. 17. The transmission range of each node is set to 250 meter radius (which is typical for IEEE 802.11b). For the INFOCOM simulation, we set the area to 1.5 km by 1.5 km,  $\tau_l$  to 200 m,  $\tau_p$  to 1 hour and the transmission range of each node to 100 m – the maximum transmission range for the Bluetooth devices used for taking the original traces. 40 nodes are simulated in both scenarios for 300 hours. For all the simulations, we assume infinite buffer and that message transfers occur instantaneously. These assumptions are used to isolate the effect of mobility patterns on the performance of DTN routing.

We also simulate RWP and BM in the same setup as the UCSD environment to compare the results. BM's ICT distribution shows 0.45 power-law slope while RWP's shows an exponential decay. Although there could be other types of mobility patterns that could generate the same ICT distributions as INFOCOM's and UCSD's, this result allows us to conjecture that the actual mobility that generates these characteristics in these settings might have been Levy walks. Furthermore, the ICT distribution patterns of various mobility models are closely related to their diffusion rates examined in Fig. 17. The more diffusive the mobility is, the shorter tail its ICT distribution becomes. To confirm this pattern, we run Levy walks with different values of  $\alpha$  while fixing  $\beta$  to one. Fig. 19(b) shows that as  $\alpha$  gets smaller, the tail distribution of ICT becomes shorter.

ICT directly impacts routing delays in DTN. To see their relation, we measure the ICT and routing delays from simulation setups that mimic the environments of our five scenarios. In this simulation, we set  $\alpha$  and  $\beta$  to the average values extracted from the angle model (with  $a_\theta = 30$  and  $r = 5$ ), set the scale factors for flight lengths and pause times to 10 and 1, respectively, and match the simulation areas to the same as those of the five sites and the flight length and pause time truncation points are set to those measured from the traces. The routing delays of RWP which uses the same environment as Campus II are also measured for comparison. In all scenarios, we simulate 40 nodes. The resulting routing delay distributions along with their ICT's are shown in Fig. 20. The figure shows that all the ICT distributions from our

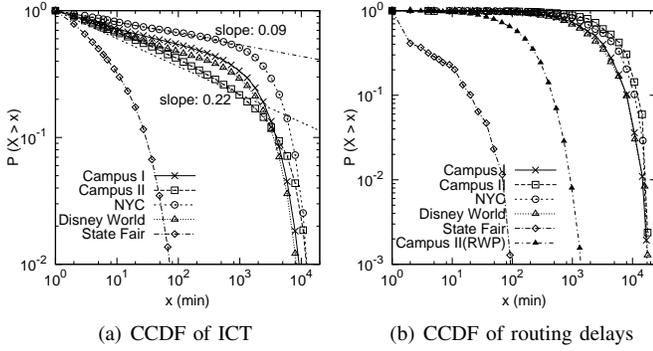


Fig. 20. ICT and DTN routing delays under various models including those constructed based on the statistics from our traces.

models, except that from State fair, follow strong power-law for the duration up to several hours. The truncations in the ICT distributions occur because the simulation is cut off around 300 hours and we consider only those contacts made within the simulation time. The ICT distribution from State fair exhibits exponential decay. This is because the area of State fair (a radius of 340 m) is much smaller than the others, and given the transmission range of 250 m, nodes can make contacts with each other without traveling a long distance. The routing delays of the corresponding Levy walk models tend to have high delays because, among many factors influencing the delays, the simulation area of our models is particularly large: on average, our models have at least four times larger an area than the UCSD simulation area. However, the routing delays of RWP still show a short tail distribution. To see the effect of flight length distributions on routing delays more clearly, we measure routing delays in the simulation runs used for Fig. 19 (b). Figure 21 (a) shows the result from which the following can be observed. BM tends to have much larger delays than any other models while RWP, as expected, shows the smallest delays because its probability of long flights is highest. The Levy walk models show their patterns in between the two extremes: as we increase  $\alpha$ , their delays get closer to BM's and as we reduce  $\alpha$ , they get closer to RWP.

The heavy tail distribution of routing delays may intuitively imply that many nodes experience similar long routing delays and that use of more relays (or copies of messages) may not necessarily improve the performance drastically. In a generalized relaying algorithm, the source distributes the message to the first  $m$  relays that it contacts. The routing delay is the time till any copy of the message is delivered to the destination. Fig. 21 (a) shows the DTN routing delays of various models when one relay is used, and Fig. 21 (b) shows the 99% quantile delays of the same models normalized by their corresponding one-relay delays as we add more relays. As expected, BM hardly achieves this goal; the delay does not improve so much as the number of relays increases, since every relay takes long time to meet the destination. However, we are surprised to find that all our Levy walk models including the one with  $\alpha = 1.5$  which shows fairly similar delay patterns as BM for one relay case, show almost the same improvement ratio as

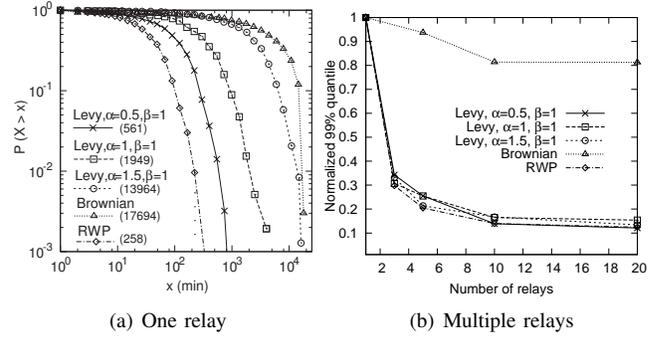


Fig. 21. The DTN delay distributions of various mobility models and normalized 99% quantile delay with multiple relays. The numbers in the parenthesis represent the actual delays in minutes at the 99% quantile of the distributions.

RWP as we add more relays. This implies that while in RWP, most nodes travel long distances frequently, in Levy walks, although not all nodes make such long trips, there exist with high probability some nodes within the mobility range of the source nodes that make such long trips. This contributes to the great reduction of the delays even with a small number of relays.

### B. Routing in MANETs

In this section, we examine the impact of Levy walks on the performance of MANET routings. There exist many MANET protocols in the literature. It is impractical to evaluate all the protocols, but instead, we first focus on the features of mobility that affects the performance of MANET routing such as hop counts and path durations. These features strongly influence the routing performance of MANETs. For instance, [21] shows that data throughput is proportional to path durations within the limit of link capacity in the network.

Fig. 22 shows the hop count distributions of the shortest path between two randomly picked nodes in the simulation of various mobility models, and the CCDF of their corresponding path durations. We use the same simulation setup as discussed in Section V. The radio range of each mobile is set to 250 meter. We run the simulation for 3000 seconds. 400 pairs of nodes are selected and the hop count of each pair is measured and sampled once at each time they establish a new path. Most of RWP hop counts are less than 15 hops and their distribution is peaked around 7 hops. This occurs because RWP nodes tend to cluster around the center of the simulation area [22]. The hop count samples of RWP are also much larger than those of the other models because as we can see in Fig. 22(b), RWP tends to maintain much shorter routing paths than the other models because of its high mobility. On the other hand, Levy walks tend to have longer paths than RWP. Because of the less diffusive nature of Levy walks, Levy walk nodes tend to stay longer in one location than RWP. Therefore, nodes are more spread out in the simulation area than RWP. Since the path durations of Levy walks are longer than those of RWP, the numbers of hop count samples of these models are much smaller. BM shows an extreme case of inactivity as its average hop count is longest. However, its path duration is

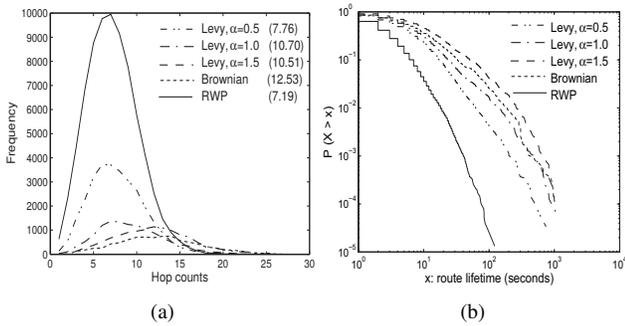


Fig. 22. (a) The hop count distributions of the shortest path between two randomly selected nodes undergoing various mobility patterns. The numbers inside the parenthesis represent the average hop counts. (b) the CCDF of their corresponding path durations.

the second longest to the Levy walks model with  $\alpha = 1.5$ . This is because most paths with long durations are from short paths, and BM nodes are more spread out and tend to have less chance of short paths as we can see from Fig. 22(a). In addition, because BM has  $\beta = 2$ , it has more occurrences of short pause times than the Levy walk models. These factors collectively contribute to reducing the path durations of BM below that of the Levy walk model with  $\alpha = 1.5$  although BM is slightly less diffusive than the Levy walk model.

To see the effects of the above-discussed factors on routing performance, we simulate DSR [23], a source-based MANET routing, in the same simulation setup as the above using GloMoSim [24]. In this simulation, we measure the data throughput of FTP connections over 300 node pairs randomly selected. The link bandwidth in these simulations is set to 2 Mbps. Figs. 23 (a) and (b) show the CCDF of throughput measured in low and high node density network environments for various mobility models. For the high density environment, we use 100 nodes in 1 km by 1 km area with  $\tau_l = 500$  m and for the low density environment, 2 km by 2 km area with  $\tau_l = 1$  km. We use the same values for the other parameters as in the simulation run for Fig. 22.

In general, both hop counts and path durations have significant impact on routing throughput. Typically, the influence of hop counts itself on data throughput gets less emphatic as hop counts increase because each simulation run contains one connection so there is only self-interference, and self-interference is limited only within a few hops. However, it is clear that as the number of hops of a path increases, its path duration is likely to reduce. Path durations seem to be a significant determinant of data throughput in our simulation. This can be seen from the similarity of Figs. 22(b) and 23(a).

In the low density simulation, the node pairs with the best throughput around the tail of the throughput CCDF tend to have long path durations. In the simulation BM and Levy walks have an order of magnitude higher maximum throughput than RWP. However, around the top of the CCDF in the figure, BM and Levy walks show a significantly less number of node pairs. This is because the number of successful path connections is much less for BM and Levy

walks. In Fig. 23(c), we plot the connection probability of node pairs, the probability that two randomly selected nodes successfully establish a routing path between them, in our simulation runs. The connection probabilities of BM and Levy walks are around 30% and 60%, respectively. This is because the difference in the diffusion rates of mobility has influenced their clustering behavior. As mentioned above, BM and Levy walk nodes tend to be more spread out, likely incurring more disconnected islands. On the other hand, while RWP nodes do not have any connectivity problem, their throughput tends to be much lower than that of BM and Levy walks. These factors collectively cause BM and Levy walks to have heavier tail throughput distributions while causing RWP to have a short tail. Thus, when examining network performance under realistic mobility models, we need to examine the entire distribution of performance instead of single numbers such as average or median values which are much less meaningful under power-law distributions of performance metrics of interest. Under the high density network simulation, all mobility models achieve 100% connection probability. Even in this environment, the data throughput under BM and Levy walks is much higher than that of RWP because of their longer path durations.

## VII. RELATED WORK

Recently, measurement studies of detailed human mobility patterns have been conducted. At Dartmouth [11] and UCSD [20], mobility traces of users are collected based on the association information of mobile handheld devices (e.g., PDAs and VoIP phones) that access wireless LAN access points (APs). However, these traces are inherently restricted by the locations of the deployed APs and thus, estimated movements in between access points might be incorrect because of relatively long distance among APs. Due to the coarse granularity of the measurement methodology, these traces are not adequate to describe detailed human mobility trajectories. In other groups, human contact patterns are studied by using iMotes [7] or information of class schedules and class rosters [25], but they do not generate detailed or accurate mobility trajectories suitable for our study. Recently, Brockmann et al. [26] analyze human traveling patterns from the circulation patterns of bank notes, in the scale of several hundred to thousand kilometers, and prove that human long-distance traveling patterns at a macro scale show Levy walk patterns. However, considering real mobile network deployments, the mobility patterns over several hundred kilometers are too large to apply to the mobility modeling for mobile network simulations. However, combined with our results that show the same result but within a much smaller scale, we can confirm the self-similar nature of human mobility. Regarding the scale-free nature of human activity, Barabasi [27] reports that various human-initiated activities including communications and work patterns are better approximated by a heavy tailed distribution, but his work does not include human mobility.

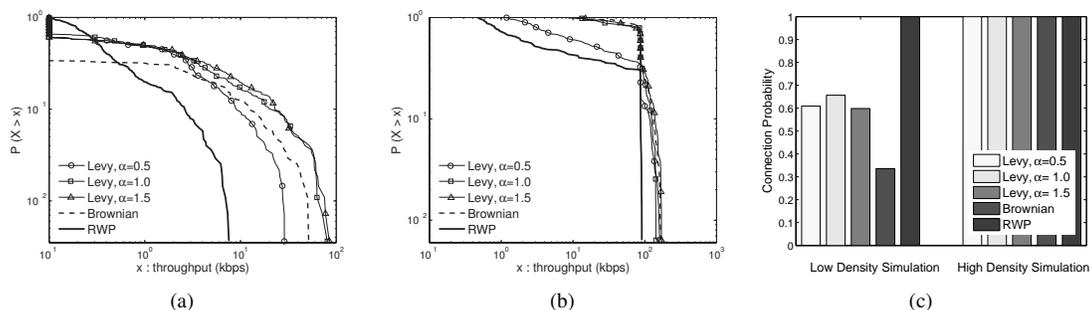


Fig. 23. (a) CCDF of FTP throughput in a low node density simulation (b) CCDF of FTP throughput in a high node density simulation (c) the probability of establishing a route between two randomly selected nodes under the low and high node density simulations

## VIII. CONCLUSION

In summary, this paper establishes statistically that human mobility at outdoor settings within a scale of less than 10 km resembles Levy walks. This is shown by the power-law tendency of flight lengths and its impact on MSD. We find that MSD shows both super-diffusive and sub-diffusive characteristics because of the power-law tendency of flight lengths and pause times. Combined with the result from [26], our result shows a self-similar nature of human mobility even beyond the scale of a few thousand kilometers. Using a simple Levy walk model we constructed, we are able to recreate the power-law distribution of inter-contact times which earlier studies have observed from human mobility. Routing performance in a mobile network undergoing Levy walks has distinctive features. In DTN, while its routing delay distribution is heavy tailed, use of multiple relays for two-hop relay routing results in drastic performance improvement. This is because there exist, with high probability, some nodes within the mobility range of a source node that make long trips. The performance of MANET routing is a complicated function of various parameters such as hop counts, connection probability and path durations. Levy walks tend to have more hop counts and longer path durations (or path survivability) than RWP. However, with Levy walks, the network is more likely to be disconnected. Again, we observe a heavy tail distribution of throughput so the performance of MANET routing cannot be easily characterized by single numbers such as average or median.

Because of space constraints, there are many research issues that are not addressed in this paper. In particular, it is interesting to further explore the cause of scale-free human mobility. From our study, we find human intentions instead of geographical artifacts play a major role. Barabasi's result [27] supports this argument. We also conjecture that this is also caused by the power law tendency of human interests or popularity of locations people visit. More studies to confirm this conjecture are required. Our treatment on the impact of Levy walks (or mobility) on network performance is limited because of space constraints. Some omitted results include the impact of  $\alpha$  and  $\beta$  on routing performance, and a study on delay and throughput tradeoffs caused by Levy walks, both of which are very interesting. Characterizing inter-contact time analytically

using Levy parameters is also intriguing. In addition, our mobility characteristics ignore the inter-dependency of humans (or nodes) such as grouping. Thus, it would be interesting to explore techniques to characterize this property and develop a model that captures it.

## REFERENCES

- [1] M. F. Shlesinger, J. Klafter, and Y. M. Wong, "Random walks with infinite spatial and temporal moments," *J. Stat. Phys.*, vol. 27, pp. 499–512, 1982.
- [2] A. Einstein, "On the motion, required by the molecular-kinetic theory of heat, of particles suspended in a fluid at rest," *Ann. Phys.*, vol. 17, pp. 549–560, 1905.
- [3] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, and H. E. Stanley, "Levy flights search patterns of wandering albatrosses," *Nature*, vol. 381, pp. 413–415, 1996.
- [4] R. P. D. Atkinson, C. J. Rhodes, D. W. Macdonald, and R. M. Anderson, "Scale-free dynamics in the movement patterns of jackals," *OIKOS*, vol. 98, no. 1, pp. 134–140, 2002.
- [5] G. Ramos-Fernandez, J. L. Morales, O. Miramontes, G. Cocho, H. Laralde, and B. Ayala-Orozco, "Levy walk patterns in the foraging movements of spider monkeys (ateles geoffroyi)," *Behavioural Ecology and Sociobiology*, vol. 55, pp. 223–230, 2004.
- [6] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. E. da Luz, E. P. Raposo, and H. E. Stanley, "Optimizing the success of random searches," *Nature*, vol. 401, pp. 911–914, October 1999.
- [7] 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. of IEEE INFOCOM 2006*, Barcelona, Spain, April 2006.
- [8] M. F. Shlesinger, G. M. Zaslavsky, and J. Klafter, "Strange kinetics," *Nature*, vol. 363, pp. 31–37, May 1993.
- [9] M. F. Shlesinger, G. M. Zaslavsky, and U. Frisch, *Levy Flights and Related Topics in Physics. In Lecture Notes in Physics*. Berlin: Springer Verlag, 1995, no. 450.
- [10] "Garmin GPSMAP 60CSx User's manual," <http://www.garmin.com/products/gpsmap60csx/>.
- [11] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. of IEEE INFOCOM 2006*, Spain, April 2006.
- [12] I. B. Aban, M. M. Meerschaert, and A. K. Panorska, "Parameter estimation for the truncated pareto distribution," *Journal of the American Statistical Assoc.*, vol. 101, no. 473, pp. 270–277, March 2006.
- [13] A. Vazquez, O. Sotolongo-costa, and F. Brouers, "Diffusion regimes in levy flights with trapping," *Physica A*, vol. 264, pp. 424–431, 1999.
- [14] G. Zumofen and J. Klafter, "Laminar localized phase coexistence in dynamical systems," *Physical Review E*, vol. 51, no. 3, pp. 1818–1821, March 1995.
- [15] P. A. DiMilla, J. A. Stone, J. A. Quinn, S. M. Albelda, and D. A. Lauffenberger, "Maximal migration of human smooth muscle cells on fibronectin and type iv collagen occurs at an intermediate attachment strength," *J Cell Biol*, vol. 122, pp. 729–737, 1993.
- [16] Y. Maruyama and J. Murakami, "Truncated levy walk of a nanocluster bound weakly to an atomically flat surface: Crossover from superdiffusion to normal diffusion," *Physical Review B*, vol. 67, no. 8, pp. 085 406–085 410, February 2003.

- [17] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. on Networking*, vol. 10, no. 4, pp. 477–486, 2002.
- [18] G. Sharma and R. R. Mazumdar, "Scaling laws for capacity and delay in wireless ad hoc networks with random mobility," in *Proc. of IEEE ICC 2004*, Paris, France, June 2004.
- [19] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proc. of ACM WDTN '05*, Philadelphia, PA, August 2005, pp. 244–251.
- [20] M. McNett and G. M. Voelker, "Access and mobility of wireless pda users," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, no. 2, pp. 40–55, 2005.
- [21] N. Sadagopan, F. Bai, B. Krishnamachari, and A. Helmy, "Paths: analysis of path duration statistics and their impact on reactive manet routing protocols," in *Proc. of ACM MobiHoc '03*, June 2003, pp. 245–256.
- [22] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," *IEEE Trans. Mobile Computing*, vol. 2, no. 3, pp. 257–269, July 2003.
- [23] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, Imielinski and Korth, Eds. Kluwer Academic Publishers, 1996, vol. 353.
- [24] "Glomosim from UCLA homepage," <http://pcl.cs.ucla.edu/projects/glomosim/>.
- [25] V. Srinivasan, M. Motani, and W. T. Ooi, "Analysis and implications of student contact patterns derived from campus schedules," in *Proc. of ACM MobiCom 2006*, Sept. 2006, pp. 86–97.
- [26] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, pp. 462–465, January 2006.
- [27] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, pp. 207–211, May 2005.