On the Levy-Walk Nature of Human Mobility

Injong Rhee, *Member, IEEE*, Minsu Shin, *Student Member, IEEE*, Seongik Hong, *Student Member, IEEE*, Kyunghan Lee, *Associate Member, IEEE*, Seong Joon Kim, *Member, IEEE*, and Song Chong, *Member, IEEE*

***Abstract*—We report that human walk patterns contain statistically similar features observed in Levy walks. These features include heavy-tail flight and pause-time distributions and the super-diffusive nature of mobility. Human walks are not random walks, but it is surprising that the patterns of human walks and Levy walks contain some statistical similarity. Our study is based on 226 daily GPS traces collected from 101 volunteers in five different outdoor sites. The heavy-tail flight distribution of human mobility induces the super-diffusivity of travel, but up to 30 min to 1 h due to the boundary effect of people's daily movement, which is caused by the tendency of people to move within a predefined (also confined) area of daily activities. These tendencies are not captured in common mobility models such as random way point (RWP). To evaluate the impact of these tendencies on the performance of mobile networks, we construct a simple truncated Levy walk mobility (TLW) model that emulates the statistical features observed in our analysis and under which we measure the performance of routing protocols in delay-tolerant networks (DTNs) and mobile ad hoc networks (MANETs). The results indicate the following. Higher diffusivity induces shorter intercontact times in DTN and shorter path durations with higher success probability in MANET. The diffusivity of TLW is in between those of RWP and Brownian motion (BM). Therefore, the routing performance under RWP as commonly used in mobile network studies and tends to be overestimated for DTNs and underestimated for MANETs compared to the performance under TLW.**

***Index Terms*—Delay-tolerant network (DTN), human mobility, Levy walk, mobile ad hoc network (MANET), mobile network, mobility model.**

I. Introduction

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 mobile nodes strongly influence the performance of mobile network protocols. As wireless devices are often carried by 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 in computer networking research are random way point (RWP) [1]–[3] or random walk models such as Brownian motion (BM) [4]–[6] and Markovian mobility [7], [8]. 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, no empirical evidence exists to prove the accuracy of such models.

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 of a particle from one location to another without a directional change or pause. Einstein [9] 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. The mean squared displacement (MSD), which is a measure of the average displacement of a given object from the origin, is proportional to t. This mobility is said to have *normal diffusion*. Many objects in the physical world undergo normal diffusion. For example, when sugar dissolves in a cup of still water, sugar particles undergo normal diffusion. Physicists (e.g., [10]) have found that there are other objects in the physical world whose mobility cannot be characterized by normal diffusion. *Levy walks* are one of the random walk models that describe such atypical mobility undergoing super-diffusion: Their MSD is proportional to t^{γ}, where $\gamma > 1$. (When $\gamma = 1$, it is called normal diffusion, and when $\gamma < 1$, it is *subdiffusion*.) Typically turbulent flows are super-diffusive. For example, when sugar dissolves in a cup of stirred water, they undergo super-diffusion. The super-diffusive nature of Levy walks results from the heavy-tail distribution of their constituent flights. Intuitively, Levy walks consist of many short flights and occasionally long flights. RWP is invented primarily for mobile network simulation wherein a node chooses its next destination randomly within a mobility area and makes a straight line flight to that destination from the current destination. RWP is highly super-diffusive because of high probability of long flights. Sample trajectories of an object undergoing BM, Levy walks, and RWP are presented in Fig. 1.

The heavy-tail flight patterns are also found in animal foraging behaviors. Viswanathan *et al.* [11] show that the flight distribution found in the mobility of albatrosses follows a power-law distribution. Similar patterns are also discovered in jackals [12] and spider monkeys [13]. The authors in [13] conjecture that the heavy-tail flight 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 [14]. Unfortunately, some of

Manuscript received March 28, 2009; revised January 25, 2010; accepted July 19, 2010; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor R. Ramjee. Date of publication April 15, 2011; date of current version June 15, 2011. This work was supported in part by National Science Foundation (NSF) NeTS-NBD 0626850, NSF NeTSE 0910868, NSF NeTS-Small 1016216, the U.S. Army Research Office (ARO) under Grant W911NF-08-1-0105 managed by NCSU Secure Open Systems Initiative (SOSI), the Ministry of Knowledge Economy (MKE), Korea, under the ITRC (Information Technology Research Center) support program supervised by the National IT Industry Promotion Agency (NIPA) [NIPA-2010-(C1090-1011-0011)], and the Korean Government (MOEHRD) under Korea Research Foundation Grant KRF-2006-352-D00137. A preliminary version of this paper was published in the Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Phoenix, AZ, April 15–17, 2008.

I. Rhee, M. Shin, and K. Lee are with the Department of Computer Science, North Carolina State University, Raleigh, NC 27695 USA (e-mail: rhee@ncsu.edu; minsu.shin@gmail.com; khan@netsys.kaist.ac.kr).

S. Hong is with the Samsung Advanced Institute of Technology, Yongin 446-712, Korea (shong@ncsu.edu).

S. J. Kim is with the DMC Research Center, Samsung Electronics, Suwon 443-742, Korea (e-mail: sj2009.kim@samsung.com).

S. Chong is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: songchong@kaist.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2011.2120618

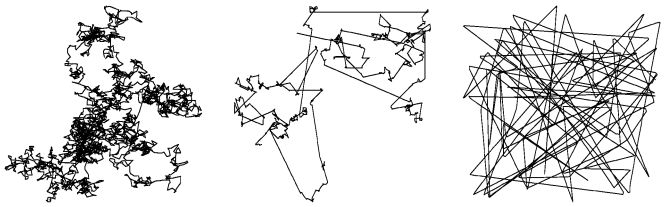


Fig. 1. Sample trajectories of (a) BM, (b) Levy walk, and (c) RWP.

these results are recently being disputed largely due to lack of accuracy in collected data and also in the processes collecting and analyzing them [15]. Because of the difficulty in collecting accurate trace data with high resolution from animals, such claims may not be easily proved or disproved.

In this paper, we study the mobility patterns of humans up to the scales of meters and seconds. We use mobility track logs obtained from over 100 participants carrying GPS receivers. The traces are obtained from five different sites: two university campuses (NCSU and KAIST), one metropolitan area (New York City), one theme park (Disney World), and one state fair. The participants walk most times in these sites and may also occasionally travel by bus, trolley, cars, or subway trains. These settings are selected because they are conducive to collecting GPS readings. The GPS receivers record their location information at every 10 s with accuracy of 3 m. The total number of participants is 101, the total duration of the traces taken is over 2228 h, and the total number of extracted flight samples is over 200 000.

Our data are by far the most detailed with high resolution and accurate traces of human mobility. Brockman *et al.* [16] show Levy-walk patterns in human travels over the scale of a few thousands of kilometers using bank note travel patterns. Gonzales *et al.* [17] use tracking information of 100 000 mobile phone users to show that human walks have heavy-tail flight distributions (note that our work [18], [19] precedes theirs). The location of a cell phone tower from which a user is initiating or receiving a call is registered as the location of the user whenever a call is made. Additionally, the locations of 206 users are sampled at every 2-h intervals for a one-week period. The resolution of location information is around 2–3 km^2. Both bank-note and phone-tracking data do not accurately record the flight information of humans. First, their resolution is at least three-orders-of-magnitude lower than ours (e.g., meters versus kilometers). Furthermore, any flights or travels that occurred between consecutive sampling points (e.g., a 2-h sampling interval or consecutive call establishments) are not tracked. Therefore, it is uncertain whether one can define every straight line between two consecutive sample locations (separated by up to a 2-h period) as a single flight. In fact, our data analysis suggests that the information lost within the two consecutive sample locations of people is very critical in understanding and recreating human walk patterns for mobile network simulation. It is hard to apply these statistical features to a detailed simulation of mobile networks, which requires resolutions of a few meters and a few seconds due to short radio ranges of mobile devices.

The analysis of our dataset indicates that the mobility of people contains similar statistical features to those found in Levy walks. In particular, their flight and pause-time distributions are best characterized by heavy-tail distributions such as Weibull, lognormal, Pareto, and truncated Pareto distributions, and their MSDs are characterized by super-diffusion up to 30 min to an hour and subdiffusion after that. These characteristics can be captured by Levy walkers moving within a confined area. The time threshold for super-diffusion is typically the time that our participants for data collection reach the boundary of their individually confined mobility area. Previous results from animal or human mobility studies show similar trends, but they are not as accurately observed as in our paper. These statistical features observed from our traces, however, contradict the mobility patterns found in commonly used mobility models for computer networking such as RWP, Random direction [20], and BM, whose mobility does not produce heavy-tail flights.

Typically, computer networks are studied using random mobility models or using a probabilistic model based on a particular distribution of intercontact times (ICTs), which are defined to be the time durations until two mobile objects meet again after meeting previously (e.g., [21]). While previous random mobility models lack the statistical features we found from our traces, the ICT-based simulation does not have essential positional information that might uniquely influence the performance of mobile networks. Since it is hard to define the underlying mobility uniquely from a given ICT distribution, the results of performance evaluation using only ICT distributions without knowing the exact underlying mobility is possibly misleading.

Empirical studies (e.g., [22]) show that the ICT distributions of human mobility have a power-law head followed by an exponential tail. It is also shown analytically that the exponential tail of the ICT distribution is caused by the homecoming nature of people [23] and also by the boundary effect [24]. However, what exact features of underlying mobility cause the power-law head of the ICT distributions is not known. Intuitively, when nodes do not move much, they tend not to meet with each other very often, thus having long ICT. In this paper, we find by simulation that BM and Levy walks produce power-law ICTs [18] as their mobility consists of many short flights, but Levy walks have much shorter ICTs than BM because of the frequency of long flights in Levy walks. On the other hand, RWP produces mostly short ICTs, and thus an exponential distribution of ICTs because of the very high frequency of long flights in RWP [25]. In summary, we find that Levy walks running in a confined area generates an ICT distribution with a power-law head followed by exponential tails whose average values are in between those of RWP and BM.

Based on the statistical patterns obtained from the traces, we construct a simple Levy-walk model called *truncated Levy walks* (TLW). TLW is a random walk that uses truncated Pareto distributions for flight and pause-time distributions to emulate mobility within a confined area. The main purpose of constructing TLW is to study the impact of heavy-tail statistical features on the performance of mobile networks. We do not claim that TLW is the most accurate human mobility model. As it is a simple random walk model, it cannot represent the important spatial, temporal, and social contexts that people live in. Despite these deficiencies, TLW can still provide

more realistic representations of statistical patterns found in human mobility than existing random mobility models while preserving the simplicity and analytical tractability of random mobility models.

We apply TLW to the performance evaluation of DTN and MANET routing protocols. We find that higher diffusivity induces shorter intercontact times in DTNs and shorter path durations with higher connection probability in MANETs. The diffusivity and ICTs of TLW are in between those of RWP and BM. Therefore, the routing performance reported using RWP as commonly used in mobile network studies tends to give at least an order of magnitude shorter routing delays than TLW in DTNs and an order of magnitude lower throughput than TLW in MANETs. Furthermore, since heavy-tail tendencies of TLW induce heavy-tail routing delays and throughput, reporting only a single performance number such as average and median is not very meaningful for understanding the performance of mobile networks.

As it is not the purpose of our paper to present TLW as an accurate mobility model, we do not attempt to study and compare the performance of TLW to many more sophisticated mobility models in the literature (e.g., [26]–[28]). However, clearly these existing models do not emulate heavy-tail statistical features. The main contribution of this paper remains at suggesting that the emulation of the heavy-tail features is an important new component that can be incorporated into these models to improve their realism. We leave the work of developing accurate human mobility models for future work. The readers are referred to [29] for our preliminary work on this topic, where the comparisons with various existing mobility models are presented.

This paper is organized as follows. Section II provides preliminary background on Levy walks and statistical analysis. Section III discusses our data collection and analysis techniques. Section IV presents our main result—the statistical analysis of mobility traces to establish that human walks exhibit Levy-walk characteristics. Section V presents a simple truncated Levy-walk model that can be used for mobile network simulations, and Section VI contains our study on routing performance using the truncated Levy-walk model. Sections VII and VIII contain related work and our conclusion.

II. Background

A. Levy Walks

Levy walks can be defined as continuous-time random walks whose turning points are visit points of the associated Levy flights, and unlike Levy flights [30], [31], they account for the time taken to complete each flight. The average displacement of a Levy walk is characterized by a time-dependent growth of displacement, as the cost of time for making each flight is explicitly given [31] from one location to another (i.e., flights) with the following features: 1) its MSD is infinite; and 2) the distribution of flight lengths follows a heavy-tailed distribution. We now consider a random walker and choose the joint space–time probability density function (PDF) $\Phi(r,t)$

$$\Phi(r,t) = \phi(t|r)p(r) \tag{1}$$

TABLE I
STATISTICS OF COLLECTED MOBILITY TRACES FROM FIVE SITES

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	0.77	2.83	10.57
Campus II (32)	92	4.21	12.21	23.32	0.31	1.83	13.31
NYC (12)	39	1.23	8.44	22.66	0.42	6.60	17.74
DW (18)	41	2.17	8.99	14.28	0.25	3.60	16.79
SF (19)	19	1.48	2.56	3.45	0.17	0.51	0.86

where $p(r)$ is the probability that a flight of length r occurs and $\phi(t|r)$ is the conditional probability density that such flight takes t time. t and r determine the speed of the flight. When $p(r)$ is a heavy-tailed distribution, a process specified by the probability density function $\Phi(r,t)$ is a Levy walk [30].

B. General Random Walk Model

Consider a two-dimensional random walk defined by a sequence of *steps* that a walker makes. A step is represented by a tuple $S = (l, \theta, \Delta t_{\mathrm{f}}, \Delta t_{\mathrm{p}})$ in which a walker makes a flight followed by a pause. θ is the direction of that flight, $l > 0$ is the length of the flight, $\Delta t_{\mathrm{f}} > 0$ is the time duration of the flight or *flight time*, and $\Delta t_{\mathrm{p}} \geq 0$ is the time duration of the pause time 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 $\psi(\Delta t_{\mathrm{p}})$, 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* Δt_{s}, which is the summation of its flight time and pause time. The walker starts its first step at the origin at time $t = 0$.

III. Measurement Methodology

A. Data Collection

Five sites are chosen for collecting human mobility traces. These are two university campuses [North Carolina State University, Raleigh, in the U.S. and Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in Asia]; New York City (NYC), NY; Disney World (DW), Orlando, FL; and one state fair (SF) in Raleigh, NC. The total number of traces from these sites is 226 daily traces. Garmin GPS 60CSx handheld receivers, which are Wide Area Augmentation System (WAAS)-capable with a position accuracy of better than 3 m 95% of the time, are used for data collection in North America [32]. Occasionally, track information has discontinuity when bearers move indoors where GPS signals cannot be received. The GPS receivers take readings of their current positions every 10 s and record them into a daily track log. The summary of daily trace is shown in Table I. We use Campuses I and II to indicate university campuses without revealing their locations.

The participants in Campus I were randomly selected students who took a course in the Computer Science Department. Every week, two or three randomly chosen students carried the GPS receivers for their daily regular activities. The Campus-II

traces are taken by 32 students who live in a campus dormitory. The New York City traces were obtained from 12 volunteers living in Manhattan or its vicinity. Their track logs contain relatively long-distance travels. Their means of travel include cars, buses, and walking. The Disney World traces were obtained from 18 volunteers who spent their Thanksgiving or Christmas holidays in Disney World. The participants mainly walked in the parks and occasionally rode trolleys. The state fair track logs were collected from 19 participants who visited the North Carolina State Fair, which includes many street arcades, small street food stands, and showcases. The event was very popular and attended by more than 1000 people daily for two weeks. The site is completely outdoors and is the smallest among all the sites. Each participant in the state fair scenario spent less than 3 h in the site.

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 every 30 s by averaging three samples over that 30-s period (note that GPS samples are taken every 10 s). All the position information discussed in this paper is based on the 30-s 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 the following method, which is a similar method used in [33], to remove the discontinuity. If the next position recorded after the discontinuity is within a radius of 20 m 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 $(\Delta t-k)$ 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-s period is less than r m. Extracting a flight from the GPS traces is not trivial because the definition of flight includes direction changes. However, people do not necessarily move in a perfect straight line although they may intend to do so. Therefore, we need to allow some margin of errors in defining the "straight" line. We use the following three different methods: namely *rectangular*, *angle*, and *pause-based* models. We differ only in the amount of the marginal angle errors. In the rectangular model, given two sampled positions $x_{\rm s}$ and $x_{\rm e}$ taken at time t and $t+\Delta t$ $(\Delta t>0)$ in the trace, we define the straight line between $x_{\rm s}$ and $x_{\rm e}$ to be a flight if and only if the following conditions are met.

1) The distance between any two consecutively sampled positions between $x_{\rm s}$ and $x_{\rm e}$ is larger than r m (i.e., no pause during a flight).
2) When we draw a straight line from $x_{\rm s}$ to $x_{\rm e}$, the sampled positions between these two endpoints are at a distance less than w m from the line. The distance between the line and a position is the length of a perpendicular line from that position to the line.
3) For the next sampled position $x'_{\rm e}$ after $x_{\rm e}$, positions and the straight line between $x_{\rm s}$ and $x'_{\rm e}$ do not satisfy conditions 1) and 2).

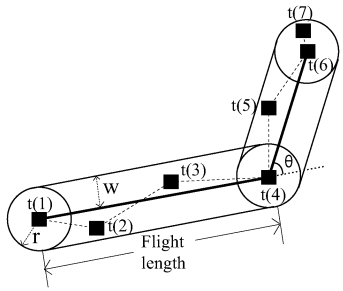


Fig. 2. Rectangular model used to extract flight information from traces.

An example of the rectangular model is shown in Fig. 2. In the 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 endpoints. In this example, the flight time is 90 s because each sample is taken every 30 s. 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 successive flights acquired from the rectangular model into a single long flight if the following two conditions are satisfied: 1) no pause occurs between consecutive flights; and 2) the relative angle (θ as shown in Fig. 2) between any two consecutive flights is less than a°_{θ}. 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_{θ} 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 $a_{\theta}=0^{\circ}$ and $a_{\theta}=360^{\circ}$, respectively. Fig. 3 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.

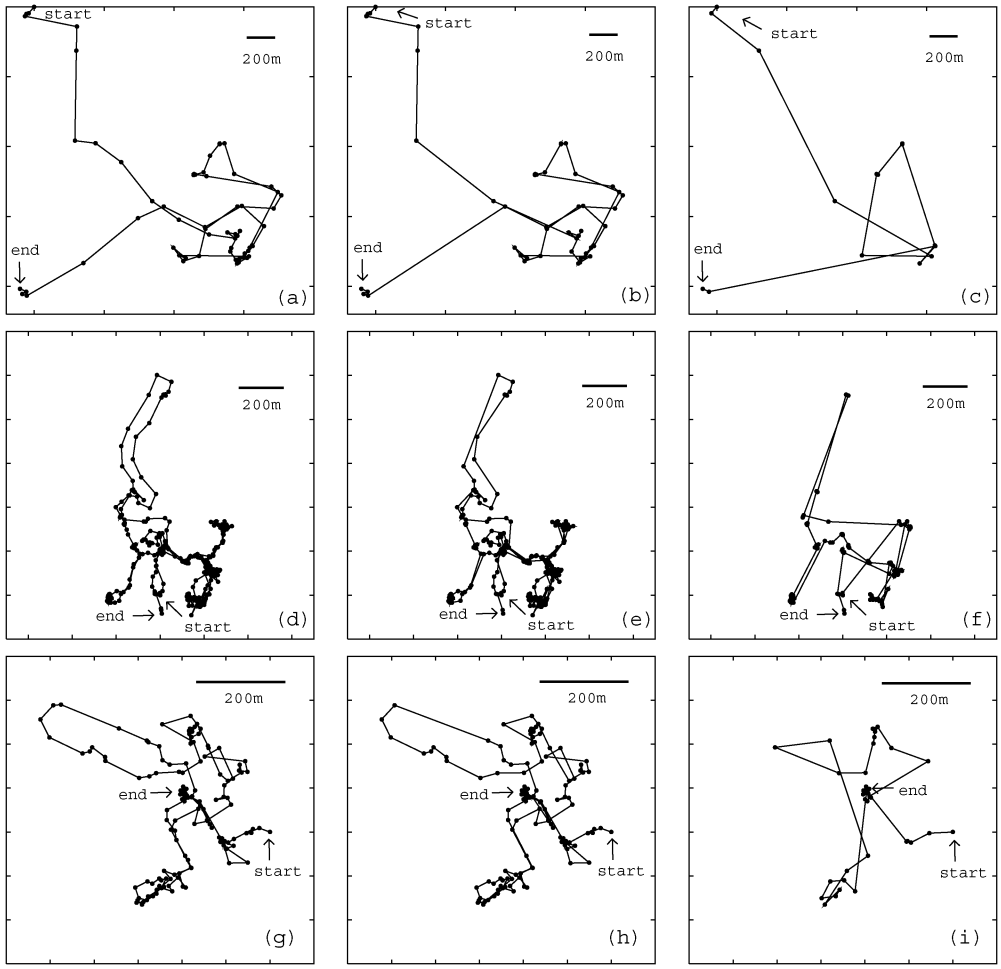


Fig. 3. Traces from (a)–(c) Campus I, (d)–(f) Disney World, and (g)–(i) State Fair. The first column represents the rectangular model with $r = w = 5$ m, the second column represents the angle model with $a_\theta = 30^\circ$, and the third column represents the pause-based model.

TABLE II
WELL-KNOWN HEAVY-TAILED DISTRIBUTIONS

Distribution	Probability density function (pdf)
Weibull[1]	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{x}{\lambda}\right)^k\right]$
lognormal	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right]$
Pareto[2]	$\frac{\alpha a^\alpha}{x^{\alpha+1}}$
Truncated Pareto[3]	$\frac{\alpha a^\alpha}{x^{\alpha+1}(1-(\frac{a}{b})^\alpha)}$

[1] A Weibull distribution is heavy-tailed when $k < 1$.
[2] For $0 < a \leq x$.
[3] For $0 < a \leq x \leq b < \infty$.

C. Fitness Metrics

Table II shows well-known heavy-tailed distributions. In practice, all commonly used heavy-tailed distributions such as Pareto, truncated Pareto, lognormal, or Weibull with decreasing failure rate (i.e., $k < 1$) belong to the subexponential class [34]. To quantitatively find the best fitting distributions, we apply Akaike's information criterion (AIC) [35] after fitting various distributions to our GPS traces by maximum likelihood estimation (MLE).

AIC [15], [35] is a model (distribution) selection criterion and is used in combination with MLE. MLE is a popular method used to fit a mathematical probability distribution f_θ parameterized by an unknown parameter θ (which could be vector-valued) to empirical data. MLE finds an estimator $\hat{\theta}$ that maximizes the likelihood function

$$\mathrm{AIC} = -2\log\left(L(\hat{\theta}|\mathrm{data})\right) + 2K \tag{2}$$

where $L(\cdot)$ is the likelihood function and K is the number of estimable parameters (the value of θ) in the approximating model (probability distribution).

The AIC test can be applied only when there are a sufficient number of samples (n), more specifically when $n/K > 40$ [35]. Since the numbers of estimated parameters of the probability distributions used in this paper are less than three, our datasets have enough numbers of samples to be qualified for AIC as shown in Table III.

As AIC values contain arbitrary constants and are greatly affected by the sample size, they do not represent an absolute metric and cannot be used directly. The following transformation makes the result an interpretable metric:

$$\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min} \tag{3}$$

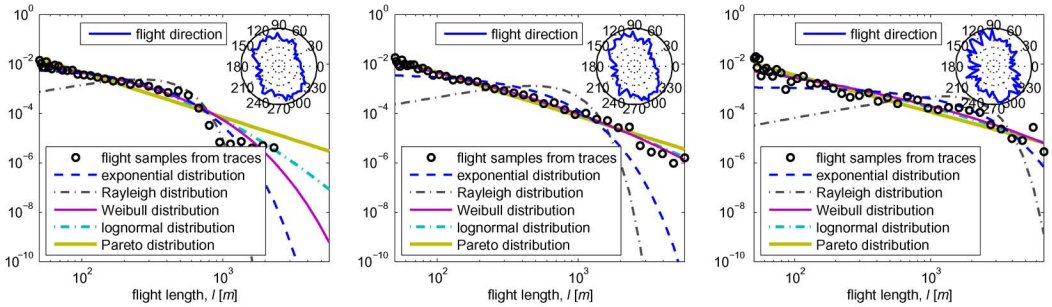


Fig. 4. Flight length distribution from the Campus-I traces in a log-log scale with logarithmic bin sizes. (a) Rectangular model ($r = w = 5$ m and $a_\theta = 0°$). (b) Angle model ($a_\theta = 30°$). (c) Pause-based model.

TABLE III
NUMBER OF FLIGHT SAMPLES

Site	number of flight (l) samples	
	$l > 0\ m$	$l > 50\ m$
Campus I	4286	297
Campus II	9416	1563
New York City	1161	325
Disney World	4548	1178
State Fair	685	222

where $\mathrm{AIC}_{\min}$ is the minimum of different AIC values. The Akaike weights w_i are useful as the weight of evidence [35]

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_{r=1}^{R} \exp(-\Delta_r/2)} \tag{4}$$

where R is the size of a set of the approximating models (distributions). Akaike weights are considered as normalization of the model likelihoods. We use MLE as an estimation method and AIC as a model (distribution) selection criterion.

IV. HUMAN MOBILITY

In this section, we analyze our GPS traces and closely examine the statistical characteristics of mobility features including flights, pause times, MSD, and absolute flight directions. We first examine the distribution of flight lengths (flight distribution, in short) taken from our traces. By fitting with various well-known distributions, we find that the flight distributions from the traces fit best to heavy-tailed distributions such as Weibull, lognormal, Pareto, and truncated Pareto. This result is unique as it is first to show such characteristic of human mobility in the scale of meters. Our work precedes the work by [17], which shows a similar tendency of human mobility using cell phone logs. Our work also contrasts with the earlier finding by [16] that human travel patterns have heavy-tail, based on bank-note tracking on the scale of 1000 km. All these results collectively confirm that human mobility is characterized by*scale-freedom* in which, in any scale, human movement has similar patterns. We then show other evidence of scale-freedom in human mobility. The distribution of pause times, the time durations that a walker spends in each stopping point before directional changes, also follows a heavy-tail distribution. The consequence of heavy-tail flights is the super-diffusivity of human mobility. We confirm this property by analyzing the mean squared distribution of people from the traces. Due to truncation caused by the movement boundaries, we can observe the super-diffusivity only up to 1 h (mostly to 30 min). These characteristics coincide with those from Levy walks although human walks are not random walks like Levy walks. To complete our data analysis, we examine the other statistical aspects of human mobility from the traces such as the distribution of flight directions and speed of travels, which are useful in constructing realistic human mobility models.

A. Flight Length Distribution

We examined both individual and aggregated flight distributions. The aggregated flight distributions aggregate flight samples from all the traces of the same site, regardless of their participants. The individual flight distributions show similar patterns as the aggregated ones, so we present only the aggregated distributions. Fig. 4 shows the log–log distribution plots of flight lengths sampled according to the three different flight models ($r = w = 5$ m, and $a_\theta = 30°$) from the Campus-I traces. We use MLE to fit aggregated flight lengths to well-known distributions such as exponential, Rayleigh, Weibull, lognormal, and Pareto distributions. MLE is performed over the x-axis range over 50 m of each distribution to isolate only the tail behavior. We find that as a_θ increases, the distribution becomes flatter with a heavier tail.

Fig. 5 shows the same for the other scenarios under the pause-based model. Fitting to truncated Pareto is shown separately in Fig. 6. Each figure shows the distribution and MLE fitting results over the flight samples over the tail section of the distribution, that is, samples between 50 m and the 0.999-quantile of each distribution. The Akaike test [35] also quantitatively measures the best fit among the tested distributions. Tables IV and V show the results without and with truncated Pareto, respectively. We separate the results for truncated Pareto mainly for clarity of presentation and also, more importantly, to isolate the effect of truncation in proving the heavy-tail tendency of the data. We did not fit truncated versions of other distributions as their analytical definitions are not well defined and they are also not used commonly. By visual inspection, heavy-tail distributions, including Weibull, lognormal, Pareto, and truncated Pareto, show the better fit than the other distributions. The Akaike comparison test shows that without truncated Pareto, that also quantitatively measures the best fit among the tested distributions. Table IV shows Weibull has the best fit for Campus I, Campus II, and New York City, and lognormal has the best fit for Disney World and the North Carolina State Fair. When a Weibull distribution

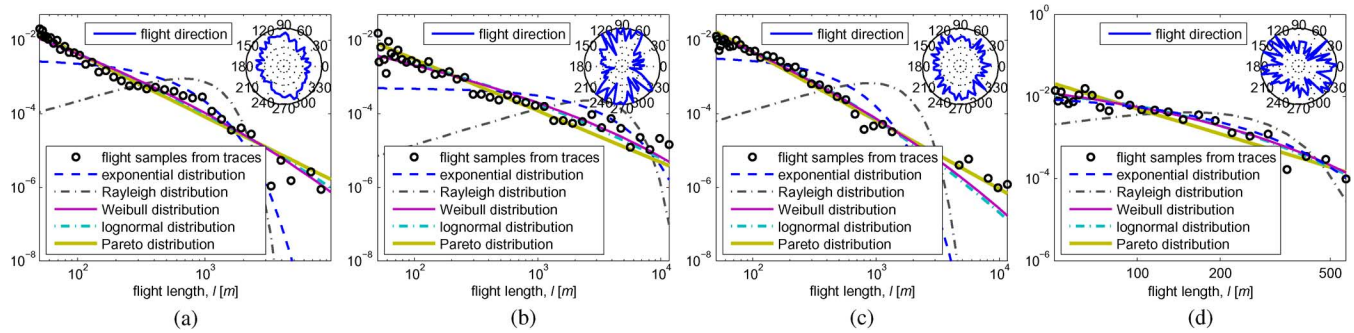


Fig. 5. Flight length distribution in a log–log scale with logarithmic bin sizes, using the pause-based model. (a) Campus II. (b) New York City. (c) Disney World. (d) State Fair.

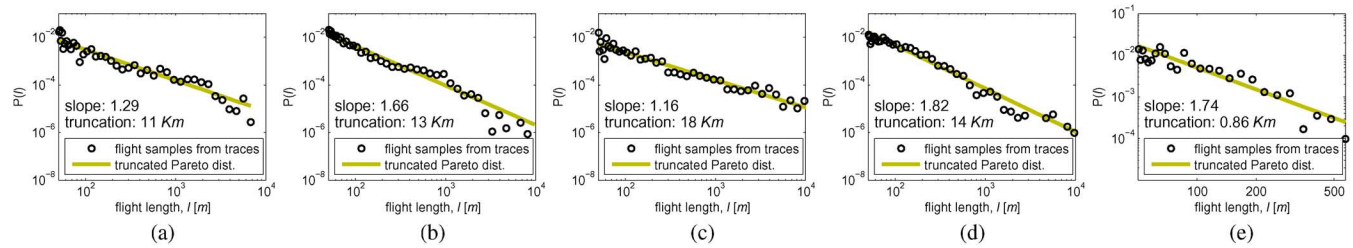


Fig. 6. Flight length distribution fitted with truncated Pareto distributions using the pause-based model. (a) Campus I. (b) Campus II. (c) New York City. (d) Disney World. (e) State Fair.

TABLE IV
AKAIKE WEIGHTS OF EXPONENTIAL, RAYLEIGH, WEIBULL, AND LOGNORMAL DISTRIBUTIONS AND MLE OF k FOR WEIBULL DISTRIBUTION, UNDER THE PAUSE-BASED MODEL

Site	Akaike weights					Weibull
	exp	Rayleigh	Weibull	logn	Pareto	k
Campus I	0.0000	0.0000	0.9987	0.0013	0.0000	0.39
Campus II	0.0000	0.0000	0.9997	0.0003	0.0000	0.23
NYC	0.0000	0.0000	0.9997	0.0003	0.0000	0.35
DW	0.0000	0.0000	0.0028	0.9972	0.0000	0.21
SF	0.0005	0.0000	0.4811	0.5183	0.0000	0.67

TABLE V
AKAIKE WEIGHTS OF EXPONENTIAL, RAYLEIGH, WEIBULL, LOGNORMAL, AND TRUNCATED PARETO DISTRIBUTIONS UNDER THE PAUSE-BASED MODEL

Site	Akaike weights					
	exp	Rayleigh	Weibull	logn	Pareto	T.Pareto
Campus I	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Campus II	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
NYC	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
DW	0.0000	0.0000	0.0028	0.9970	0.0000	0.0002
SF	0.0005	0.0000	0.0035	0.0038	0.0000	0.9928

is fitted, the estimated value of parameter k is less than 1. Note that a Weibull distribution with $k < 1$ is heavy-tailed, and lognormal and Pareto distributions are heavy-tailed by definition. Table V shows that all traces except Disney World have the best fit with truncated Pareto. Disney World still has the best fit with the lognormal distribution.

In the insets of Figs. 4 and 5, we plot the normalized frequency of each turning angle. Their distributions are close to uniform in general, although the New York trace seems to have some biases to particular directions.

TABLE VI
AVERAGE OF SLOPES (WITH STANDARD DEVIATION) FROM THE MLE FOR TRUNCATED PARETO TO FIT-TO-FLIGHT LENGTHS OBTAINED BY VARYING FLIGHT EXTRACTION PARAMETERS: r AND w FROM 2.5 TO 10 m AND a_θ FROM 15° TO 90°

	Rectangular $a_\theta = 0°$	Angle	Pause-based $a_\theta = 360°$
Campus I	-1.53 (0.03)	-1.64 (0.03)	-1.22 (0.11)
Campus II	-2.27 (0.02)	-2.15 (0.04)	-1.63 (0.11)
NYC	-1.62 (0.02)	-1.57 (0.04)	-1.17 (0.10)
Disney World	-2.20 (0.04)	-2.16 (0.08)	-1.85 (0.09)
State fair	-2.81 (0.45)	-2.11 (0.18)	-1.76 (0.15)

Table VI shows the average of slopes from the MLE of truncated Pareto and their standard deviation. All the scenarios have slopes larger than -3 (so $\alpha < 2$).

The flight-length distribution of State Fair in Fig. 5 appears close even to a short-tailed distribution such as exponential. This seems inconsistent with the other data as they show clear separation from short-tailed distributions. To see if this disparity comes from heavy truncations due to the small size of the state fair site (less than 860-m radius), we simulate two instances of a Levy walk, with width 200 m and the other with 2 km. Fig. 7 shows the complementary cumulative distribution function (CCDF) of flight lengths obtained from Levy-walk simulations in two squares. The Levy walk in the small area has the same truncation problem (phenomenon) as the state fair, and we find that the flight distribution can fit well even to a short-tailed distribution. However, when we increase the area, the same Levy walk has a heavy tail. This indicates that the state fair data may not be inconsistent with the other data.

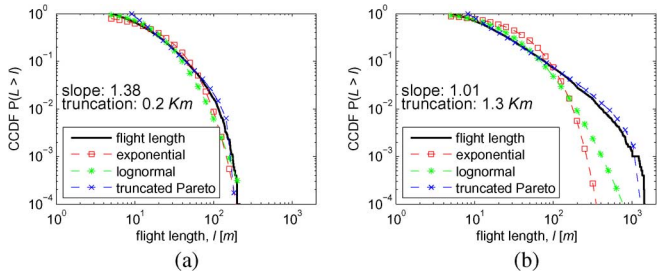


Fig. 7. CCDF of flight lengths obtained from Levy-walk simulations in two squares: (a) one with width 200 m and (b) the other with 2 km. The Levy walk in the smaller area appears like Brownian motion.

B. Pause-Time Distribution

We find from our traces that the pause times of our walkers can be fitted to truncated Pareto well and they have a heavy-tailed distribution. Fig. 8 shows the pause-time distributions extracted from our traces. The flight definitions do not influence 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.

Power-law pause-time distributions affect the MSD of walkers as shown in [36] and [37]: Long trapping caused by heavy-tailed pause-time distributions makes the mobility less diffusive, sometimes causing subdiffusion. We have more details in Section IV-C.

C. Mean Squared Displacement (MSD)

The average distance of a walker at time t from the origin (i.e., the position at time 0) is called *displacement*. The MSD is the square of the average displacement of a random walker. If a random walker follows Brownian-motion patterns (i.e., with finite mean and variance of flights and pause times), then its MSD is proportional to t [9] as governed by the central limit theorem (CLT). We say such walks have *normal diffusion*. However, if flights do not have characteristic scales (e.g., power law) as in Levy walks, then the MSD of the random walker is proportional to t^γ, $\gamma > 1$. This is called *super-diffusion*. If Brownian-motion flights are combined with power-law pause times, then their MSD becomes proportional to t^γ, $\gamma < 1$, which is called *subdiffusion*. It is also shown that when random walkers with power-law flights and pause times are confined within a fixed area, because of truncations of flights, their MSD shows a dichotomy in which it is super-diffusive up to some time limits and then becomes normal-diffusive or subdiffusive [36], [38].

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 the average of MSD values measured by varying the origin among all locations that the walker has been at [38], [39]. Specifically, for each scenario, we compute the following. Given each trace T from that scenario that consists of an ordered sequence of location samples $(t_0, \mathrm{pos}_T(t_0))$, where $\mathrm{pos}_T(t_0)$ is the two-dimensional position of the walker at time t_0 in trace T, the $\mathrm{MSD}(t)$ of that scenario in terms of time interval t is

$$\mathrm{MSD}(t) = \frac{\sum_T \sum_{t_0} |\mathrm{pos}_T(t+t_0) - \mathrm{pos}_T(t_0)|^2}{N}. \quad (5)$$

$\mathrm{pos}_T(t+t_0) - \mathrm{pos}_T(t_0)$ 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 $t_0 + t < t_{\mathrm{e}}^T$, where t_{e}^T is the time that the last sample of trace T is taken. If $t_0 + t > t_{\mathrm{e}}^T$, the contribution of t_0 to $\mathrm{MSD}(t)$ is zero. We compute $\mathrm{MSD}(t)$ directly from the GPS traces mapped to the two-dimensional space.

Fig. 9 plots the $\mathrm{MSD}(t)$ for Campus I, Campus II, New York City, Disney World, and State Fair. The shape of $\mathrm{MSD}(t)$ in a log–log scale can be fitted by two lines using the least-squares method. From the plots, we can see that up to about 1 h, our participants make super-diffusive mobility ($\gamma > 1.15$), and after that, they make subdiffusive mobility ($\gamma < 0.80$).

The dichotomy of super-diffusion and subdiffusion observed in our traces is consistent with the finding in [36] and [38]. Since our GPS traces are daily traces and each trace is obtained within a specific area, flight lengths are limited and areas have boundaries. When t is small (in our case, less than 30 min), the effect of truncations does not appear. The effect of heavy-tailed distributions shows up, and the mobility appears super-diffusive. As we increase time t, the truncation takes effect and a walker reaches a boundary. We also confirmed this pattern using simulation (the result is not shown). Another significant factor causing the subdiffusion 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 the entrance and exit in Disney World). This "homecoming" tendency slows down diffusion excessively, resulting in subdiffusion as t increases.

D. Flight Speed

Fig. 10 shows the velocity and flight times in terms of flight lengths, extracted from all the five scenarios. Flight times and lengths are highly correlated. From Fig. 10(a), 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: $\Delta t_f = k l^{1-\rho}$, $0 \le \rho \le 1$, where k and ρ are constants. In one extreme, when ρ is 0, flight times are proportional to flight lengths, and it models the constant velocity movement. In another extreme, when ρ 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 $k = 30.55$ and $\rho = 0.89$ when $l < 500$ m, and with $k = 0.76$ and $\rho = 0.28$ when $l \ge 500$ m.

V. Levy-Walk Mobility Model

We are interested in the effect of mobility patterns we observed from our data analysis on the performance of mobile networks. To this end, we construct a simple TLW model that generates synthetic mobility tracks reflecting the statistical patterns of human mobility. We use the same random walk model discussed in Section II. A step is represented by four variables:

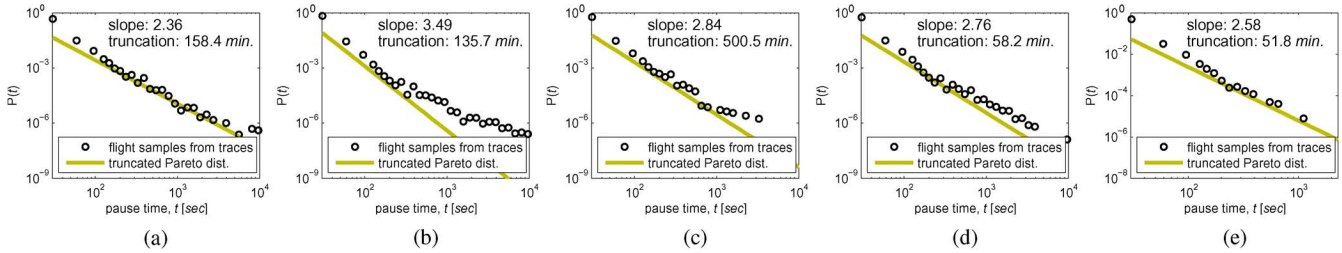


Fig. 8. Pause-time distribution of human walks in a log–log scale with logarithmic bin sizes, using the pause-based model. (a) Campus I. (b) Campus II. (c) New York City. (d) Disney World. (e) State Fair.

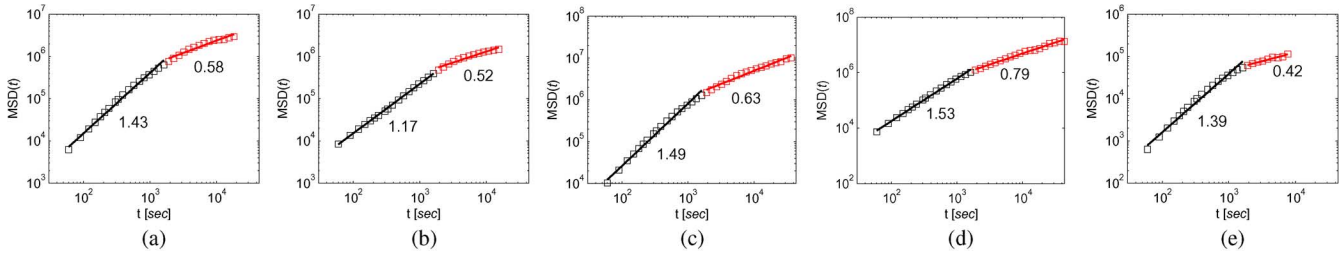


Fig. 9. MSD from various settings. (a) Campus I. (b) Campus II. (c) New York City. (d) Disney World. (e) State Fair.

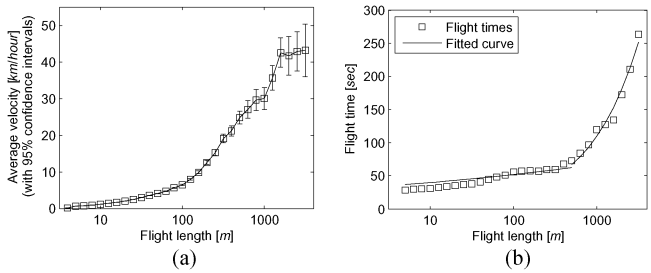


Fig. 10. (a) Average velocity depending on flight lengths with 95% confidence intervals, obtained from all the traces. (b) Flight times and the corresponding fitted curve.

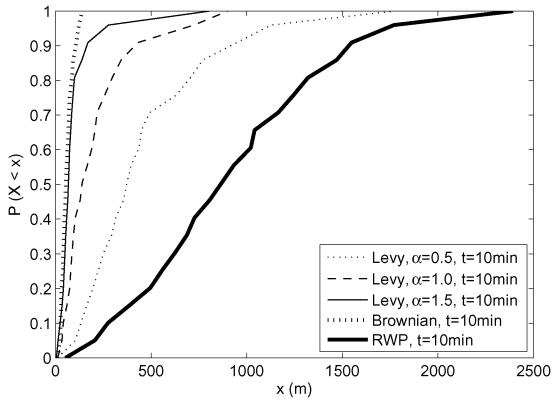


Fig. 11. CDF of node displacement from its initial position after 10-min travel. RWP is most diffusive, while BM is least diffusive. The diffusion rates of truncated Levy walks are in between.

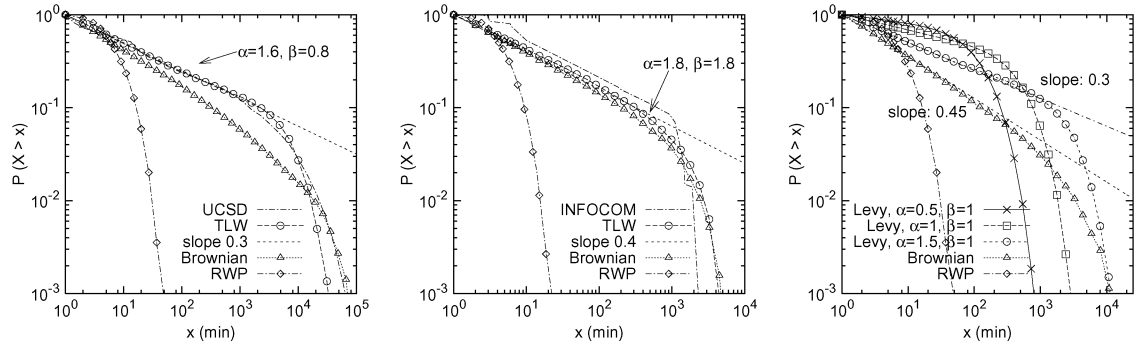
flight length l, direction θ, flight time Δt_{f}, and pause time Δt_{p}. Our model picks flight lengths and pause times randomly from their PDFs $p(l)$ and $\psi(\Delta t_{\mathrm{p}})$, which are Levy distributions with coefficients α and β, respectively. The following defines a Levy distribution with a scale factor c and exponent α in terms of a Fourier transformation:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx - |ct|^{\alpha}}. \tag{6}$$

For $\alpha = 1$, it reduces to a Cauchy distribution, and for $\alpha = 2$, a Gaussian with $\sigma = \sqrt{2}c$. Asymptotically, for $\alpha < 2$, $f_X(x)$ can be approximated by $1/|x|^{1+\alpha}$. We allow c, α, and β to be simulation parameters. We use a uniform flight direction distribution. We use the speed model used in Section IV.

Using the above model, we generate synthetic (truncated) Levy-walk mobility traces with truncation factors τ_{l} and τ_{p} for 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 $(l, \theta, \Delta t_{\mathrm{f}}, \Delta t_{\mathrm{p}})$ is generated randomly from their corresponding distributions. If l and Δt_{p} are negative or $l > \tau_{\mathrm{l}}$ or $\Delta t_{\mathrm{p}} > \tau_{\mathrm{p}}$, then we discard the step and generate a new step. We repeat this process after the step time $\Delta t_{\mathrm{f}} + \Delta t_{\mathrm{p}}$. Until the end of the simulation, we generate the tuples repeatedly. When a flight crosses a boundary of the predefined area, we allow the flight to be reflected off the wall.

By adjusting α and β, we can generate mobility traces with different diffusivity. Note that when α is 2, then the model becomes BM. Fig. 11 shows the cumulative distribution function (CDF) of the distance that a mobile is away from its initial position after the first 10 min of travel. The simulation area is set to 2×2 km^2. The truncated Levy-walk models are constructed by setting the pause-time factor (β) to 0.5, but varying the flight length factor (α) from 0.5 to 1.5. We set the truncation points $\tau_{\mathrm{l}} = 1$ km and $\tau_{\mathrm{p}} = 1000$ s 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 $\alpha = 2$. All models use the same velocity and pause-time model discussed, 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 truncated Levy-walk models are in between these two extremes. As we reduce α, the mobility becomes more



(a) UCSD (b) INFOCOM (c) TLW

Fig. 12. ICT distributions of mobility models. Truncated Levy walks recreate the ICT distributions seen in the (a) UCSD and (b) INFOCOM traces. The measured ICT distributions of UCSD and INFOCOM are obtained from [22]. ICT distributions from various mobility models in the same setting as the UCSD traces. (c) ICT distributions of TLW with different α.

diffusive. We shall see that this disparate diffusion rate of mobility in each model has distinctive effects on the performance of routing in mobile networks.

VI. Impact of Heavy-Tail Flights on Routing Performance

In this section, we apply TLW to the simulation of DTNs and MANETs and study the impact of heavy-tail flights on routing performance in these networks.

A. Routing in Delay-Tolerant Networks

In DTNs, mobile nodes may establish ON and OFF connectivity with their neighbors and the rest of the network. Store-and-forward is the main paradigm of routing in such networks where communication transpires only when two devices are in a radio range and the time duration between two consecutive contacts of the same two nodes (called *intercontact time*) is an important metric.

It is known that the ICT distribution of human walks exhibits a power-law tendency up to some time after which it shows exponential decay [23]. The result is interesting because [25] showed by simulation that RWP produces exponentially decaying ICT. What is not obvious is the type of mobility pattern that gives rise to the power-law tendency of ICT distributions.

The earlier measurement studies on ICT (e.g., [22]) report power-law distributions of ICT with human mobility with slopes of 0.3 from the University of California, San Diego (UCSD) and Dartmouth, Hanover, NH, traces [40] and 0.4 from the INFOCOM trace [22]. TLW can generate the similar ICT distributions in the similar settings as UCSD and INFOCOM by adjusting α and β. Fig. 12 shows the result. In the UCSD simulation, we fix the simulation area to 3.5×3.5 km^2, τ_1 to 3 km, and τ_p to 28 h. The transmission range of each node is set to a 250-m radius (which is typical for IEEE 802.11b) for UCSD. For the INFOCOM simulation, we set the area to 1.5×1.5 km^2, τ_1 to 200 m, τ_p to 1 h, and the transmission range of each node to a 50-m radius to fit to the transmission range of the Bluetooth devices used for taking the original traces. These values are chosen based on the corresponding real traces. In both simulations, 40 nodes are simulated, and we set the scale factors (c) of flight lengths and pause-time distributions to 10 and 1, respectively.

We also simulate RWP and BM in the same setup as the UCSD and INFOCOM environments. The BM model uses $\alpha = 2$, and RWP chooses a random destination uniformly within the simulation area. All the models use the same pause-time distribution and velocity model as TLW. All the simulation runs are ensured to be in their stationary regimes as all the mobility models have finite pause time and trip durations, and we discard the first 100 h of simulation results to avoid transient effects. BM and TLW produce a heavy-tail ICT distribution, while RWP's shows an exponential decay. In the UCSD experiment, TLW produces a better fitting ICT than BM, while both BM and TLW produce the similar ICT patterns for INFOCOM. In both cases, TLW can fit both power-law head and exponentially decaying tail. In the INFOCOM setting, the area is very small compared to the radio range so that there are a lot of truncations. Thus, in such a setting, TLW may look like BM (as shown in the state-fair results of Section IV). This result indicates that TLW is much more flexible than the other models in generating more realistic statistical patterns observed in real traces.

To study the impact of diffusivity on the ICT patterns, we run TLW with various α while fixing β to 1. Fig. 12(c) shows the result. This indicates that the ICT distribution patterns of various mobility models are closely related to their diffusion rates. In RWP, the mobility is the most diffusive, and in BM it is the least. In TLW, the diffusivity is in between, and with a smaller value of α, it becomes more diffusive and its ICT has a shorter tail.

To see the effect of Levy-walk features on routing performance, we simulate one of the most widely studied routing DTN algorithms called *two-hop relay routing* [41], where 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. We run the protocol under RWP, BM, and TLW with various α-values. For all the simulations, we assume an 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. The area of the simulation is set to the size of UCSD.

Fig. 13 shows the performance of the protocol with one relay and multiple relays. BM has the heaviest tail distribution of routing delays, and RWP has the shortest. BM tends to have much longer delays than any other models because of their slow

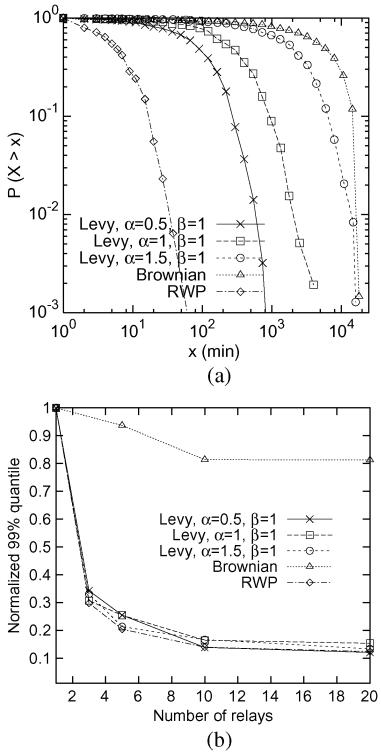


Fig. 13. DTN delay distributions of various mobility models and normalized 99% quantile delay with multiple relays. (a) One relay. (b) Multiple relays.

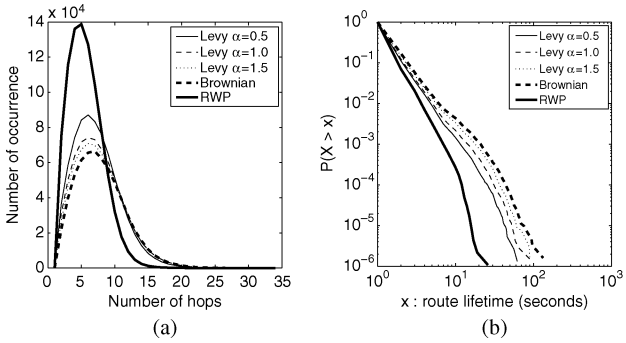


Fig. 14. (a) Hop-count distributions of the shortest path between two randomly selected nodes undergoing various mobility patterns. The numbers inside the parenthesis represent the average hop counts. (b) CCDFs of their corresponding path durations.

diffusion rate, while RWP, as expected, shows the smallest delays because of high mobility of nodes. TLW shows performance in between the two extremes: As we increase α, its delays get closer to BM's, and as we reduce α, they get closer to RWP.

We simulate a multiple-copy protocol where the source distributes the message to the first m relays that it contacts. The routing delay is the time until any copy of the message is delivered to the destination. Fig. 13(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 a 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 RWP as we add more relays. This implies that while most nodes travel long distances frequently in RWP, 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 routing protocols. We first focus on the features of mobility that affects the performance of MANET routing protocols such as hop counts and the duration of routing paths being connected. These features strongly influence the routing performance of MANETs. For instance, [42] shows that data throughput is proportional to path durations within the limit of link capacity in the network.

Fig. 14(a) 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 m. We run the simulation for 3000 s. Four hundred pairs of nodes are selected, and the hop count of each pair is measured and sampled once whenever they establish a new path. RWP tends to have very short paths because RWP nodes tend to cluster around the center of the simulation area [1]. Levy walks and BM tend to have longer paths than RWP. Because of the less diffusive nature of these models, nodes tend to stay longer in one location than RWP.

A *path duration* is the time period that a path stays connected. Fig. 14(b) shows the path duration distributions measured under various mobility models. RWP has the shortest path durations due to its high mobility; BM has the longest because of slow diffusivity. The path durations of TLW are in between the two extremes.

To see the effects of the above-discussed factors on routing performance, we simulate the dynamic source routing protocol (DSR) [43] in the same simulation setup as the above using GloMoSim [44]. In this simulation, we measure the data throughput of FTP connections over 600 node pairs randomly selected. In each run, one source-and-destination pair is selected. The link bandwidth in these simulations is set to 2 Mb/s. Fig. 15(a) and (b) shows 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 a 1×1 km^2 area with $\tau_1 = 500$ m, and for the low-density environment, a 2×2 km area with $\tau_1 = 1$ km. We use the same values for the other simulation parameters as in the simulation run for Fig. 14.

In general, both hop counts and path durations have significant impact on routing throughput. Since each run contains only one connection, there is no effect of interference other than self-interference—the interference caused by the nodes in the same path. Typically, the influence of hop counts itself on data throughput gets less emphatic as hop counts increase because self-interference is limited only within a few hops. However, it is clear that as a path gets longer, its path duration is likely to

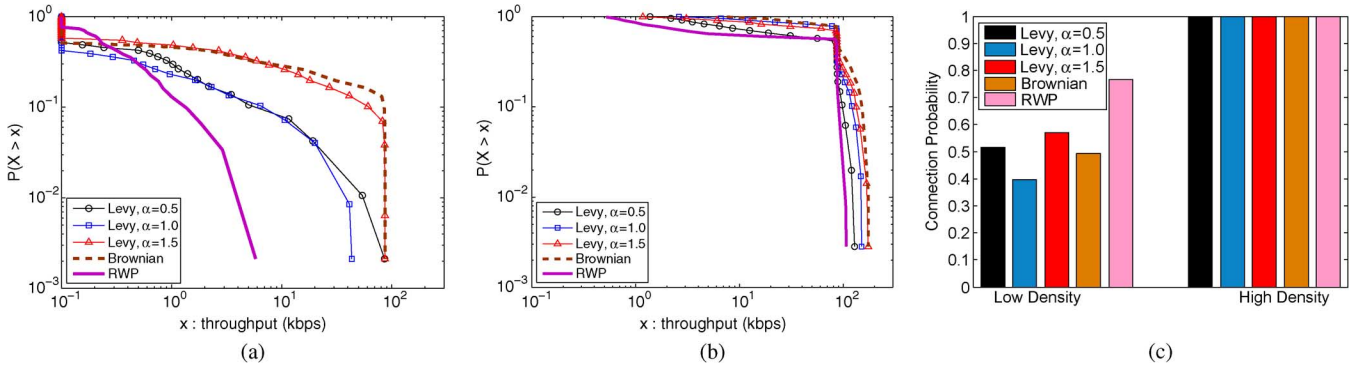


Fig. 15. (a) CCDF of FTP throughput in a low-node-density simulation. (b) CCDF of FTP throughput in a high-node-density simulation. (c) Probability of establishing a route between two randomly selected nodes under the low- and high-node-density simulations.

reduce because of higher probability of disconnection. Path durations are a significant determinant of data throughput in our simulation, which can be seen from the similarity of Figs. 14(b) and 15(a).

In the low-density simulation, we confirm that the node pairs with the best throughput around the tail of the throughput CCDF tend to have long path durations. BM and TLW have an order of magnitude higher throughput than RWP. However, around the head of the CCDF in the figure, BM and TLW show a significantly less number of node pairs with nonzero throughput. This is because the number of successful path connections is much less for BM and TLW. In Fig. 15(c), we plot the connection probability of node pairs, the probability that two randomly selected nodes successfully establish a path between them. The connection probabilities of BM and TLW are around 50%. The nodes of BM and TLW with large α likely incur more disconnected islands because of low mobility. On the other hand, while RWP nodes have better connectivity probability around 80%, their throughput tends to be much lower than that of the other models. These factors collectively cause BM and TLW to have a heavier tail throughput distribution. Thus, when examining network performance under more 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 TLW are 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 [33] and UCSD [40], 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 [22] or information of class schedules and class rosters [45], but they do not generate detailed or accurate mobility trajectories suitable for our study. Recently, Brockmann *et al.* [16] analyzed human traveling patterns from the circulation patterns of bank notes, in the scale of several hundred to thousand kilometers, and proved that human long-distance traveling patterns at a macro scale show Levy-walk patterns. It was also reported by Gonzales *et al.* [17] that humans tend to perform Levy walks, in the scale of hundreds of meters, within heterogeneously bounded areas from the analysis of location information that is taken when mobile phone calls are made or received or sampled at every 2 h. However, considering real mobile network deployments, the mobility patterns over large scales covering several hundred meters to several hundred kilometers are too large to apply to the mobility modeling for mobile network simulations. Combined with our results that show the same result but within a much smaller scale, we can confirm the scale-free nature of human mobility. Regarding the scale-free nature of human activity, Barabasi [46] 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.

VIII. Conclusion and Discussion

In summary, this paper finds that human walks at outdoor settings within less than 10 km contain statistically similar features as Levy walks including heavy-tail flight and pause-time distributions and the super-diffusion followed by subdiffusion, which is an indication of heavy-tail flights in a confined area. Combined with the results from [16] and [17], our results show a scale-free nature of human mobility even beyond the scale of a few thousand kilometers. Using a simple truncated Levy-walk model we constructed, we are able to recreate the power-law distribution of intercontact times that earlier studies have observed from human mobility. Routing performance in a mobile network undergoing Levy walks has distinctive features. In DTNs, while their 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. TLW tends to have more

hop counts and longer path durations (or path survivability) than RWP. However, with TLW, the network is more likely to be disconnected. We also 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 in producing heavy-tail tendencies. We also conjecture that this is 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 α and β on routing performance and a study on delay and throughput tradeoffs caused by Levy walks, both of which are very interesting. Characterizing intercontact time analytically using parameters of a Levy-walk model is also intriguing. In addition, our mobility characteristics ignore the interdependency 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] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 2, no. 3, pp. 257–269, Jul.–Sep. 2003.

[2] E. Hyytia, P. Lassila, and J. Virtamo, "Spatial node distribution of the random waypoint mobility model with applications," *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 680–694, Jun. 2006.

[3] J. Yoon, M. Liu, and B. Noble, "Random waypoint considered harmful," in *Proc. IEEE INFOCOM*, San Francisco, CA, 2003, vol. 2, pp. 1312–1321.

[4] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Commun. Mobile Comput.*, vol. 2, no. 5, pp. 483–502, Aug. 2002.

[5] R. Groenevelt, E. Altman, and P. Nain, "Relaying in mobile ad hoc networks: The Brownian motion mobility model," *Wireless Netw.*, vol. 12, no. 5, pp. 561–571, Sep. 2006.

[6] S. Ioannidis and P. Marbach, "A Brownian motion model for last encounter routing," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006, pp. 1–12.

[7] C. Bettstetter, "Mobility modeling in wireless networks: Categorization, smooth movement, and border effects," *Mobile Comput. Commun. Rev.*, vol. 5, no. 3, pp. 55–66, July 2001.

[8] I. F. Akyildiz, Y.-B. Lin, W.-R. Lai, and R.-J. Chen, "A new random walk model for pcs networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1254–1260, Jul. 2000.

[9] 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.

[10] M. F. Shlesinger, J. Klafter, and Y. M. Wong, "Random walks with infinite spatial and temporal moments," *J. Stat. Phys.*, vol. 27, no. 3, pp. 499–512, Mar. 1982.

[11] 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.

[12] R. P. D. Atkinson, C. J. Rhodes, D. W. Macdonald, and R. M. Anderson, "Scale-free dynamics in the movement patterns of jackals," *OIKOS, J. Ecol.*, vol. 98, no. 1, pp. 134–140, 2002.

[13] G. Ramos-Fernandez, J. L. Mateos, O. Miramontes, G. Cocho, H. Larralde, and B. Ayala-Orozco, "Levy walk patterns in the foraging movements of spider monkeys (ateles geoffroyi)," *Behav. Ecol. Sociobiol.*, vol. 55, no. 3, pp. 223–230, 2004.

[14] 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, Oct. 1999.

[15] A. M. Edwards, R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, S. V. Buldyrev, M. G. E. da Luz, E. P. Raposo, H. E. Stanley, and G. M. Viswanathan, "Revisiting Levy flight search patterns of wandering albatrosses, bumblebees and deer," *Nature*, vol. 449, pp. 1044–1048, Oct. 2007.

[16] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, pp. 462–465, Jan. 2006.

[17] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, Jun. 2008.

[18] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the Levy-walk nature of human mobility: Do humans walk like monkeys?," 2007.

[19] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the Levy walk nature of human mobility," in *Proc. IEEE INFOCOM*, Phoenix, AZ, Apr. 2008, pp. 924–932.

[20] E. Royer, P. Melliar-Smith, and L. Moser, "An analysis of the optimum node density for ad hoc mobile networks," in *Proc. IEEE ICC*, 2001, pp. 857–861.

[21] A. Balasubramanian, B. N. Levine, and A. Venkataramani, "DTN routing as a resource allocation problem," in *Proc. ACM SIGCOMM*, Kyoto, Japan, Aug. 2007, pp. 373–384.

[22] 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. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006, pp. 1–13.

[23] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic, "Power law and exponential decay of inter contact times between mobile devices," in *Proc. ACM MobiCom*, Montreal, QC, Canada, Sep. 2007, pp. 183–194.

[24] H. Cai and D. Y. Eun, "Crossing over the bounded domain: From exponential to power-law inter-meeting time in MANET," in *Proc. ACM MobiCom*, 2007, pp. 159–170.

[25] G. Sharma and R. R. Mazumdar, "Scaling laws for capacity and delay in wireless ad hoc networks with random mobility," in *Proc. IEEE ICC*, Paris, France, Sep. 2004, vol. 7, pp. 3869–3873.

[26] M. Musolesi and C. Mascolo, "A community based mobility model for ad hoc network research," in *Proc. 2nd ACM/SIGMOBILE REALMAN*, May 2006, pp. 31–38.

[27] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *Mobile Comput. Commun. Rev.*, vol. 11, no. 3, pp. 59–70, Jul. 2007.

[28] V. Borrel, F. Legendre, M. D. de Amorim, and S. Fdida, "SIMPS: Using sociology for personal mobility," *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 831–842, Jun. 2009.

[29] K. Lee, S. Hong, S. Kim, I. Rhee, and S. Chong, "SLAW: A mobility model for human walks," in *Proc. IEEE INFOCOM*, 2009, pp. 855–863.

[30] M. F. Shlesinger, G. M. Zaslavsky, and J. Klafter, "Levy dynamics of enhanced diffusion: Application to turbulence," *Phys. Rev. Lett.*, vol. 58, pp. 1100–1103, Mar. 1987.

[31] M. F. Shlesinger, B. J. West, and J. Klafter, "Strange kinetics," *Nature*, vol. 363, pp. 31–37, May 1993.

[32] *"Garmin GPSMAP 60CSx User's Manual"* Garmin, Olathe, KS [Online]. Available: http://www.garmin.com/products/gpsmap60csx/

[33] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006, pp. 1–13.

[34] S. Asmussen and K. Biswanger, "Simulation of ruin probabilities for subexponential claims," *ASTIN Bull.*, vol. 27, no. 2, pp. 297–318, Nov. 1997.

[35] K. P. Burnham and D. R. Anderson, "Multimodel inference: Understanding AIC and BIC in model selection," *Sociol. Meth. Res.*, vol. 33, no. 2, pp. 261–304, Nov. 2004.

[36] A. Vazquez, O. Sotolongo-costa, and F. Brouers, "Diffusion regimes in Levy flights with trapping," *Phys. A, Stat. Mech. Appl.*, vol. 264, no. 3, pp. 424–431, Mar. 1999.

[37] G. Zumofen and J. Klafter, "Laminar-localized-phase coexistence in dynamical systems," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 51, no. 3, pp. 1818–1821, Mar. 1995.

[38] Y. Maruyama and J. Murakami, "Truncated levy walk of a nanocluster bound weakly to an atomically flat surface: Crossover from superdiffusion to normal diffusion," *Phys. Rev. B, Condens. Matter Mater. Phys.*, vol. 67, no. 8, pp. 085 406.1–085 406.5, Feb. 2003.

[39] 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.f Cell Biol.*, vol. 122, no. 3, pp. 729–737, August 1993.

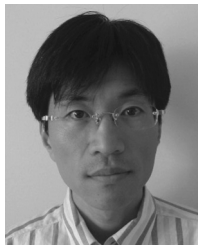
[40] M. McNett and G. M. Voelker, "Access and mobility of wireless PDA users," *Mobile Comput. Commun. Rev.*, vol. 9, no. 2, pp. 40–55, Apr. 2005.
[41] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.
[42] 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. ACM MobiHoc*, Annapolis, MD, Jun. 2003, pp. 245–256.
[43] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*. Norwell, MA: Kluwer, 1996, vol. 353.
[44] "GloMoSim," Univ. California, Los Angeles, Los Angeles, CA [Online]. Available: http://pcl.cs.ucla.edu/projects/glomosim/
[45] V. Srinivasan, M. Motani, and W. T. Ooi, "Analysis and implications of student contact patterns derived from campus schedules," in *Proc. ACM MobiCom*, Los Angeles, CA, Sep. 2006, pp. 86–97.
[46] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, pp. 207–211, May 2005.

Injong Rhee (M'94) received the Ph.D. degree from the University of North Carolina, Chapel Hill, in 1994.

He is a Professor of computer science with North Carolina State University, Raleigh. His areas of research interests include computer networks, congestion control, wireless ad hoc networks, and sensor networks.

Minsu Shin (S'99) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1998, 2000, and 2006, respectively.

Since July 2006, he has been a Visiting Post-Doctoral Research Fellow with the Department of Computer Science, North Carolina State University, Raleigh. His research interests are in congestion control, programmable networks, and wireless networks.



Seongik Hong (S'97) received the Ph.D. degree in computer science from North Carolina State University, Raleigh, in 2010.

He is currently with the Samsung Advanced Institute of Technology (SAIT), Yongin, Korea. From 1998 to 2005, he was a Member of Technical Staff with the Telecommunications and Operations Support Laboratory, Korea Telecom (KT), Seongnam, Korea. His research interests include human mobility modeling and future Internet.

Kyunghan Lee (S'07–A'10) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2002, 2004, and 2009, respectively.

He is currently a Post-Doctoral Researcher with the Department of Computer Science, North Carolina State University, Raleigh. His research interests are in the areas of human mobility, delay-tolerant networks, context-aware service, service-oriented network, and wireless mesh networks.

Seong Joon Kim (S'00–A'08–M'09) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1998, 2000, and 2007, respectively.

He is currently with the DMC Research Center, Samsung Electronics, Suwon, Korea, as a Member of Technical Staff. Prior to joining Samsung Electronics, he was a Post-Doctoral Fellow with the Department of Computer Science, North Carolina State University, Raleigh. His research interests include WLANs and wireless ad hoc networks.

Song Chong (M'95) received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Seoul, Korea, in 1988 and 1990, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin in 1995.

Since March 2000, he has been with the School of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, where he is a Professor and the Director of the Communications and Computing Group.