

On the Levy-walk Nature of Human Mobility

Injong Rhee, Minsu Shin, Seongik Hong
Dept. of Computer Science
NC State University

Kyunghan Lee, Song Chong
School of EECS
KAIST

Abstract—We report that human walks performed in outdoor settings of tens of kilometers resemble a truncated form of Levy walks 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. This paper shows that many statistical features of human walks follow truncated power-law, showing evidence of scale-freedom and do not conform to the central limit theorem. These traits are similar to those of Levy walks. It is conjectured that the truncation, which makes the mobility deviate from pure Levy walks, comes from geographical constraints including walk boundary, physical obstructions and traffic. None of commonly used mobility models for mobile networks captures these properties. Based on these findings, we construct a simple Levy walk mobility model which is versatile enough in emulating diverse statistical patterns of human walks observed in our traces. The model is also used to recreate similar power-law inter-contact time distributions observed in previous human mobility studies. Our network simulation indicates that the Levy walk features are important in characterizing the performance of mobile network routing performance.

I. INTRODUCTION

Our purpose of studying human mobility patterns is their use in simulating mobile networks of wireless devices carried by people. As wireless devices are often attached to humans, understanding their mobility patterns leads to more realistic network simulation and more accurate understanding of protocol performance. Commonly used mobility models include random way point (RWP) or random walk models such as Brownian motion and Markovian mobility. These models are simple enough for theoretical analysis and experimental simulation. However, there has been little statistical validation of such models for accuracy in describing human mobility.

Biologists [1], [2], [3] have found that the mobility patterns of foraging animals such as spider monkey, albatrosses (seabirds) and jackals can be commonly described in what physicists have long called *Levy Walks*. The term Levy walks was first coined by Schlesinger et al. [4] 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 [5] 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. 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 MSD of particles making Levy walks is proportional to t^γ 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 in which differences in the patterns are visually evident.

In this paper, we study the statistical patterns of human walks observed within a radius of tens of kilometers. 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. Although the number of participants is relatively small in our study, the total duration of tracks taken over the five different sites are over 1000 hours, which adds to the statistical significance of our findings.

From the data analysis of our traces, we find the followings:

- The mobility patterns of the participants in these outdoor settings have features defining Levy walks; their flight distributions and pause time distributions closely match truncated power-law distributions. Their MSD also shows significant influence of these mobility patterns.
- There exist some deviations from pure Levy walks occurring due to various factors specific to human mobility including geographical constraints such as roads, buildings, obstacles and traffic. These deviations are manifested in our traces in the form of flight truncations which may make the flight distribution appear like heavy-tailed or even short-tailed at times.

From these findings, we construct a simple Levy walk (LW) mobility model that can be used for network simulation. LW is shown to be versatile enough to emulate various statistical features and those deviations observed in our traces when run

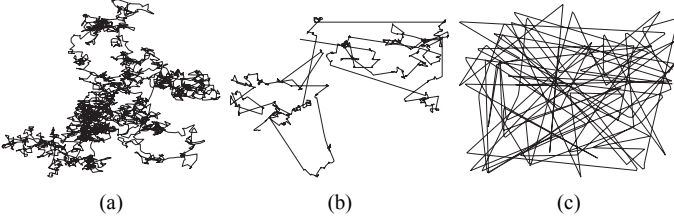


Fig. 1. Sample trajectories of (a) Brownian motion, (b) Levy walk and (c) Random way point

under similar geographical constraints. Using LW, we report the followings:

- LW can reproduce similar power-law inter-contact time (ICT) distributions of human walks observed in earlier studies (e.g., [6]). This result is significant because ICT is an important performance determinant of DTN (delay-tolerant networks) routing performance and our model provides some clues on what caused the power-law tendency.
- The DTN network simulation using the model shows unique routing performance characteristics unobserved in RWP and BM: while RWP shows overly optimistic routing performance because its high occurrences of long flights intensify the chance of meeting destinations, and BM shows overly pessimistic routing performance because its high occurrences of short flights diminish the chance of meeting destinations, LW shows performance somewhere in-between the performance seen from these two models. For instance, the tail distributions of routing delays for LW is between those of RWP (very short) and BM (very heavy), but the significant portion of long flights in LW also causes similar drastic improvement in routing delays as RWP when the amount of message redundancy increases. We also found other unique performance characteristics from MANET simulation, but due to a space constraint, we report only DTN results. The results on MANET simulation can also be found in our full report [7].

We do not claim that human walks are Levy walks; although our human mobility traces contain striking statistical resemblances to Levy walks, there are still some significant deviations between pure Levy walks and our human walks. This point may be less relevant to our research. What is most relevant, perhaps the biggest contribution of our work, is to demonstrate the value of Levy walk models in recreating similar statistical patterns observed in real world traces that are important in characterizing the performance of network routing protocols in mobile environments involving humans. To the best of our knowledge, this is the first work that studies 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 this nature. This paper is an extended abstract of our full report [7].

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: θ 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 stays at 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 $t = 0$.

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 σ^2 and the step time distribution has a finite mean τ . 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 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 α [9]. Such random walks are named as *Levy walks* [4]. The scale-free distribution of flight lengths leads to *super-diffusion* where MSD is proportional to t^γ , $\gamma > 1$.

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 $\psi(\Delta t_p) \sim 1/\Delta t_p^{1+\beta}$.



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
NCSU (20)	35	1.71	10.19	21.69	0.46	1.82	5.84
KAIST (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.

III. MEASUREMENT METHODOLOGY

A. Data collection

Five sites are chosen for collecting human mobility traces. These are two university campuses (NCSU and KAIST), New York City, Disney World (Orlando), and North Carolina state fair. 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 America [10]. 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. Some more details about the participants in our study can be found in [7].

All participants in the five scenarios are different individuals. Although each scenario, except NCSU, may not contain many participants, the number of traces we obtain for each scenario is quite large and the traces are taken for a long period of time which provide enough data for statistical significance. Moreover, we perform our analysis separately for each scenario and we do not aggregate traces from different scenario in our analysis. Nonetheless, as can be seen in the ensuing analysis, many statistical similarities are found among traces from different scenarios. Note that participants are not necessary simultaneously there.

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

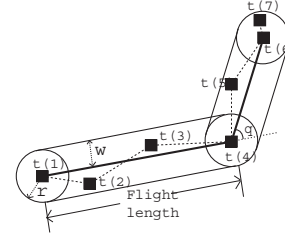


Fig. 3. The rectangular model used to extract flight information from traces.

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. It is not straightforward to extract flight information from a trace because people hardly move in a straight line. Combined with GPS errors, this human “errors” make it difficult to analyze flight data. To reduce noise due to these factors, 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. 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: (a) no pause occurs between consecutive flights and (b) the relative angle (θ as shown in Fig. 3) between any two consecutive flights is less than a_θ 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_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 $a_\theta = 0$ and $a_\theta = 180$, respectively. Fig. 4 presents sample traces produced by the above three flight models (we only show a few scenarios due to the space constraint; all the traces can be found in [7]).

IV. HUMAN MOBILITY

In this section, we analyze the statistical features of human walks from our traces.

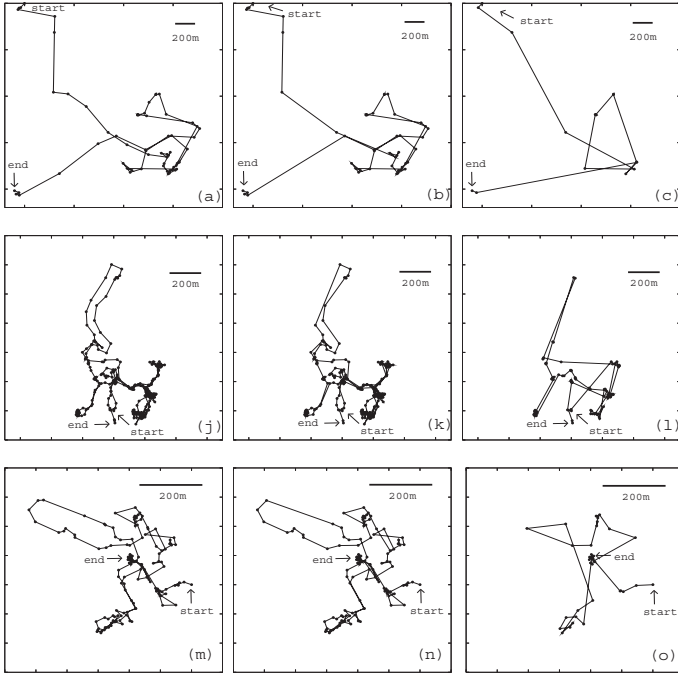


Fig. 4. Traces from NCSU (a)-(c), Disney World (j)-(l) and State fair (m)-(o). The first column represents the rectangular model with $r = w = 5$, the second column the angle model with $a_\theta = 30$, and the third column the pause-based model.

	Rectangular $a_\theta = 0$	Angle	Pause-based $a_\theta = 180$
NCSU	0.65 (0.03)	0.66 (0.02)	0.24 (0.08)
KAIST	1.11 (0.02)	1.07 (0.04)	0.51 (0.10)
NYC	0.45 (0.01)	0.59 (0.04)	0.23 (0.11)
Disney World	1.31 (0.04)	1.19 (0.13)	0.79 (0.08)
State fair	1.55 (0.34)	1.13 (0.18)	0.62 (0.29)

TABLE II
THE ESTIMATED α VALUES (WITH STANDARD DEVIATION) FITTED TO THE LOG-LOG CCDF OF FLIGHT LENGTHS OBTAINED BY VARYING FLIGHT PARAMETERS: r AND w FROM 2.5 METERS TO 10 METERS AND a_θ FROM 15 DEGREES TO 90 DEGREES.

A. Flight length distribution

In this section, we study the distributions of flight lengths from our traces. 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. 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., [3]).

Fig. 5 shows the CCDF (complementary cumulative density function) of flight lengths from each scenario. CCDF is known to show the tail patterns of a distribution better than log-log binned PDF plots. The PDF plots can also be found in our full report [7]. We apply Maximum Likelihood Estimation (MLE) to fit three known distributions, exponential, log-normal, and truncated Pareto distributions [12] to the CCDF. 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 with truncation points over three-orders of magnitudes (i.e., 1000 meters), which is a rule of thumb for power-law distributions. To reduce the sensitivity on particular settings of our flight models, we vary the values of r , w and a_θ from 2.5 meters to 10 meters and from 15 degrees to 90 degrees, respectively. We performed line fitting on the tails of the resulting CCDF over several ranges. 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 -2 (so $\alpha < 2$).

Flight truncations are natural consequences of geographical constraints including boundaries and physical obstructions, and observation artifacts (e.g., we do not consider those flights that leave the area boundary). All the distributions in Fig. 5 suffer from truncations of flights longer than a few kilometers whose effects are shown as sharp drops in the frequency of very long flights. This effect show up evidently with State fair traces shown in Fig.5(e) where even short-tail distributions fit well. 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.

The sharp drops at the tails give rise to a possibility that

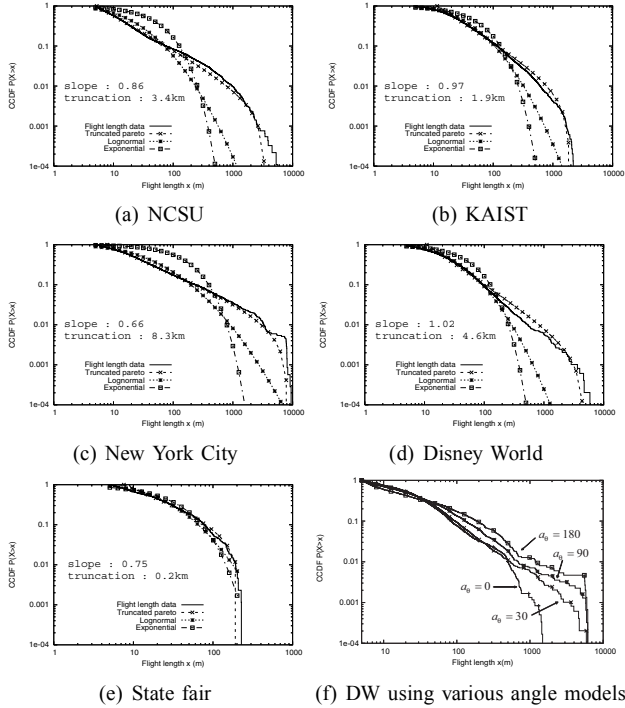


Fig. 5. The CCDF of flight lengths ($\alpha_\theta = 30$). Various known distributions are fitted using maximum likelihood estimation.

the flight distributions have long-tails but not power tails since truncated power law distributions can be also fitted with non power-law long-tail distributions such as Weibull [13]. (This truncation problem also appears in earlier studies of animal mobility, e.g., [3].) Our data is inconclusive in disproving this. However, there are some hints that this may not be the case. Fig. 5(f) shows the CCDF of flights as we increase the flight angle in the flight model. We find that as the angle increases, the distribution becomes flatter with a heavier tail. Under the pause-based model (i.e., $\alpha_\theta = 180$), it shows the heaviest tail. While it seems obvious that the frequency of longer flights increases with more angle tolerance in the flight model, this phenomenon also reveals an important feature in human mobility patterns: if we accept that humans tend to pause for a non-zero period of time when they get to a destination, the heavier-tail distribution of flights for the pause-based model implies that it is human intention causing the heavy-tail tendency, not the geographical constraints that force humans to make short flights with no pause. This also implies the scale-free tendency of the flight distribution: as we increase the scale by removing constraints and boundaries or increasing the observation area, we are expected to see longer flights. It does not make sense that human intention to move to a destination is bounded by some invisible boundaries as in Weibull. The power-law tendency of human mobility over a larger scale [14] also provides hints for this scale-freedom and self-similarity. This human intention is not well described by pure non-power-law long-tail distributions.

From the perspective of network simulations, power law distributions are easy to scale because simulation setups

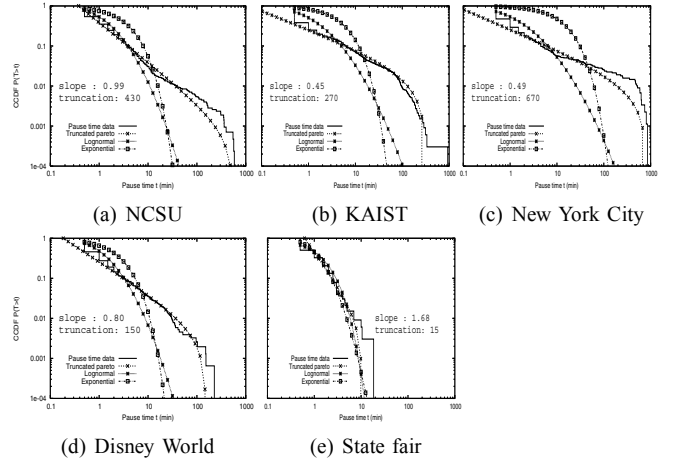


Fig. 6. The pause time CCDF of human walkers in various scenarios along with the MLE of various known distributions.

including geographical constraints may always change; for instance, simulation can run in a small area as well as a large area. It would be impossible to pick a different distribution for different setups. Using power-law distributions while inducing truncations as the natural consequence of adaptations to a given set of geographical constraints offers a much more convenient way of mobility simulation. Our data implies that human intentions and activities for mobility, independent of geographical constraints, are scale-free within our observation scale. Although geographical constraints may vary in different scenarios, this scale-free 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 heavy-tail tendency of human intentions must be inherent in the mobility model to accurately depict human walk patterns.

B. Pause time distribution

Fig. 6 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. The plots 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 than for flight lengths. However, State fair shows a good fit with short-tail distributions as well. 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, high traffic in the setting prevented them from staying at one location for a long time.

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

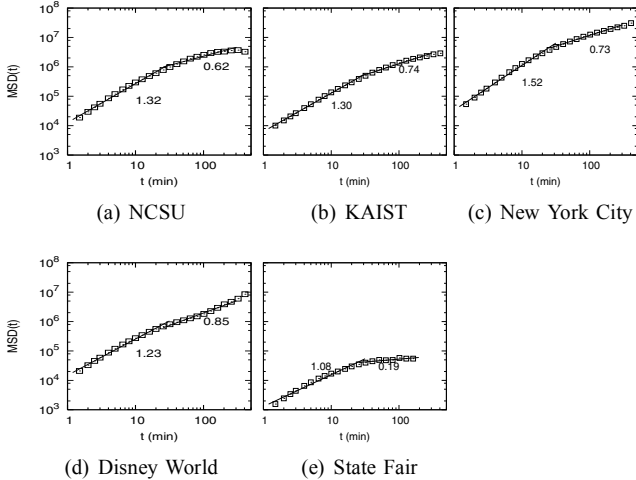


Fig. 7. MSD from various settings.

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 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 MSD(t) is zero. We compute MSD(t) directly from the GPS traces mapped to the two dimensional space.

Fig. 7 plots the MSD(t) for all scenarios. The shape of 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 ($\gamma < 0.9$). Truncated Levy walks are known to have this pattern of MSD [16].

The change in the diffusion rates can be explained 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 long flights (relative to that scale) frequently appear. 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 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

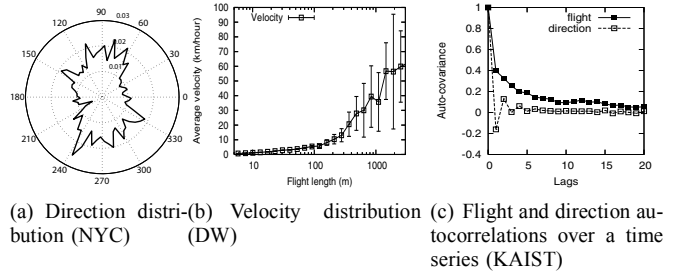


Fig. 8. Various statistics from human walk traces.

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 [17] 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.

D. Directions, velocity and auto-correlation

We also study other statistics that are relevant to generating human mobility models. Fig. 8 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 [7], the New York City traces have more bias in particular directions mostly in 90 and 270 degrees. This pattern is likely related to geographical artifacts since Manhattan tends to induce more perpendicular directional changes. Fig. 8 (a) shows the turning angle distribution from New York City 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. 8(b) 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 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. 8(c) 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. MOBILITY MODEL AND PERFORMANCE EVALUATION

In this section, we devise a Levy walk (LW) mobility model that can emulate the statistical features we observed in the earlier sections. Using the model, we characterize the performance of routing protocols when run in a mobile network carried by humans.

A. Levy-walk (LW) Mobility Model

A step is represented by four variables, flight length (l), direction (θ), flight time (Δt_f), and pause time (Δt_p). Our model selects flight lengths and pause times randomly from their PDFs $p(l)$ and $\psi(\Delta t_p)$ 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,

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx - |ct|^\alpha} dt \quad (4)$$

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 approximately by $\frac{1}{|x|^{1+\alpha}}$. We allow c , α and β to be simulation parameters. We use a uniform angle distribution as shown in most of our data and we do not add any geographical artifacts (other than boundaries) in our synthetic model. The flight speed of our model is set by the following relation between flight times and flight lengths: $\Delta t_f = kl^{1-\rho}$, $0 \leq \rho \leq 1$ 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 $k = 18.72$ 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 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 $(l, \theta, \Delta t_f, \Delta t_p)$ is generated randomly from their corresponding distributions. If l and Δt_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.

B. Model verification

In this section, we verify whether LW can synthetically generate the statistical features we have observed in our traces. Figs. 9 (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 α and β to find synthetic traces that have similar flight length and pause-time distributions of each scenario. We do

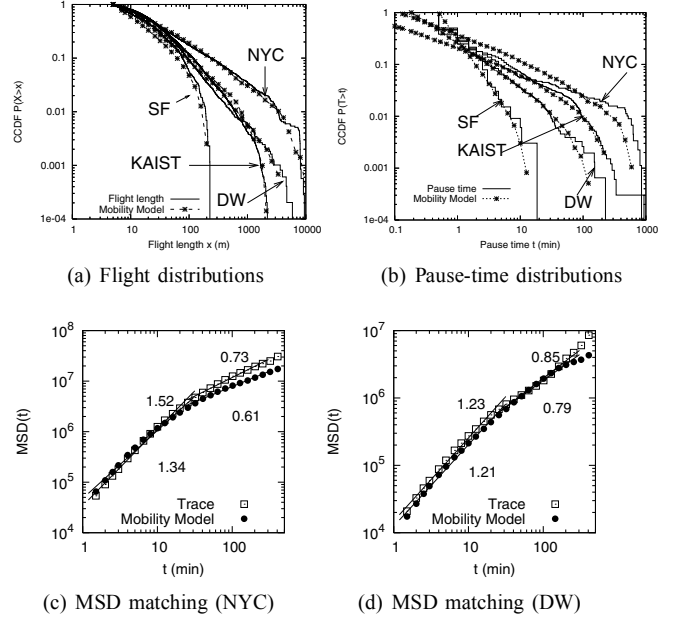


Fig. 9. Our Levy walk model can generate synthetic traces that match the flight and pause time distributions seen from real human walk traces. However, it also shows some discrepancy in the MSD values although its overall shape is similar.

not add any geographical constraints other than the simulation area (i.e., we set τ_l to infinity) and any flight that goes outside the area is abandoned and a new flight is generated. We set the truncation of pause time (τ_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. In addition, note that our truncated LW mobility model has its stationary regime since it has finite pause time and trip duration as shown in [18].

Our model does not exactly mimic human walk patterns. This can be seen from the MSD values measured from the synthetic traces. Figs. 9 (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.

C. Inter-contact time distribution

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 [19]. In this paper, we confirm the same tendency from the ICT distribution generated from our LW model exhibiting a good fit to empirical ICT distributions from real traces.

The earlier measurement studies on ICT (e.g., [6]) report power-law distributions of ICT with human mobility with slopes of 0.3 from the UCSD and Dartmouth traces and 0.4 from the INFOCOM trace. By varying the parameters of α and β of our mobility model, we are able to generate ICT distributions with the similar characteristics as in [6]. Fig. 10(a) shows the result for UCSD and the result for INFOCOM can be found in [7]. In the UCSD simulation, we fix the simulation area to 3.5 km by 3.5 km, τ_l to 3 km and τ_p to 28 hours. These values are chosen based on the data from [20]. We set the scale factors (c) of flight lengths and pause time distributions to 10 and 1, respectively. The transmission range of each node is set to 250 meter radius (which is typical for IEEE 802.11b). 40 nodes are simulated for 300 hours except in Fig. 10(a). In Fig. 10(a) we set simulation hours to 11 weeks as shown in [20].

We also simulate RWP and BM in the same setup as the UCSD environment to compare the results. The BM model uses $\alpha = 2$ and RWP chooses a random destination uniformly within the simulation area. The pause time distributions of these models are set the same as that in the LW model. All the simulation runs are ensured to be in their stationary regimes as all the mobility models have finite pause time and trip duration and we discard the first 100 hours of simulation results to avoid transient effects as shown in [18]. All models use the same velocity model discussed in V-A. Compared to BM's ICT distribution, the ICT distribution of Levy walks fits much better to the measured ICT distribution in UCSD. We are able to fit the power-law slope and also approximate the exponential decay at the tail portion of the measured data. Although there could be other types of mobility patterns that could generate the same ICT distributions as UCSD's, this result allows us to conjecture that the actual mobility that generates these characteristics in these settings is more closely modeled by Levy walks than BM. Furthermore, 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 LW, the diffusivity is in-between and with smaller value of α it becomes more diffusive. The more diffusive the mobility is, the shorter tail its ICT distribution becomes. To confirm this pattern, we run Levy walks with various α while fixing β to one. Fig. 10(b) shows that as α gets smaller, the tail distribution of ICT becomes shorter.

D. DTN routing performance

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* [21] 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 various mobility models including RWP, BM and LW with various α values (small values induce heavier tails in flight distributions). For all the simulations, we assume infinite buffer and that message transfers occur instantaneously. These assumptions are used to

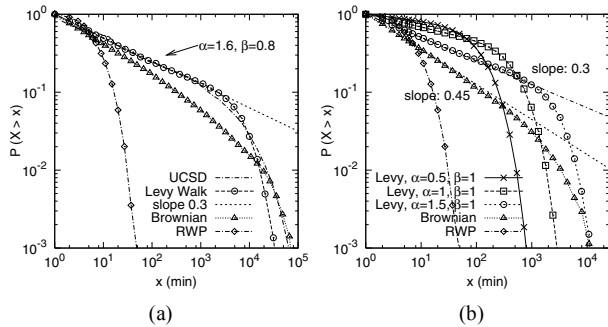


Fig. 10. The ICT distributions of mobility models. (a) Levy walks recreate the ICT distributions seen in the UCSD traces. The measured ICT distribution of UCSD is obtained from [6]. (b) ICT distributions from various mobility models.

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. 11 shows the performance of the protocol with one relay and multiple relays. From Fig. 11(a), we can see that 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 nodes do not move out of their original locations very often 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 α , their delays get closer to BM's and as we reduce α , 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. 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 till any copy of the message is delivered to the destination. Fig. 11 (b) shows the 99% quantile delays of the same models normalized by their corresponding one-relay delays as we add more relays. As expected, in BM, the delay does not improve so much as the number of relays increases, since every relay takes long time to meet the destination. However, it is interesting to note 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 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.

VI. 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.,

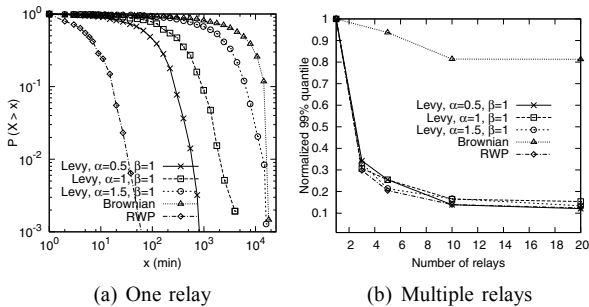


Fig. 11. The DTN delay distributions of various mobility models and normalized 99% quantile delay with multiple relays.

PDA 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 [6] or information of class schedules and class rosters [22], but they do not generate detailed or accurate mobility trajectories suitable for our study. Recently, Brockmann et al. [14] 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.

VII. CONCLUSION

Humans move with context, work, meeting, gathering, favorite places, etc. Thus, it is wrong to claim that human walks are random Levy walks because it does not make sense that humans move in a pure random fashion. What we conclude from our work, though, is not that they are the same, but there exist some common statistical patterns between them, most distinctively, the heavy-tail distribution of flight lengths and pause times and super and sub-diffusive MSD; and these features are likely not caused by artifacts of geographical constraints (obstacles, buildings, roads, boundary, etc.), but rather by human intention ([23] also finds other human activities such as email are also heavy-tailed).

We view that it is an important step forward for the mobility modeling of mobile network simulation. Many network simulation studies use random mobility models to simulate mobility of mobile nodes. Unfortunately, most lack in the features we find from our human traces. Our work points out that there exists some scale-free features inherent of human mobility in outdoor settings and these features are important for the mobility models to provide some realism in the simulation.

While we can easily program the maneuvering around physical obstacles, the scale-free features are invariant and must be explicitly modeled.

REFERENCES

- [1] 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.
- [2] 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.
- [3] G. Ramos-Fernandez, J. L. Morales, O. Miramontes, G. Cocho, H. Larralde, 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the levy-walk nature of human mobility," Tech. Rep., NCSU, 2007, <http://netsrv.csc.ncsu.edu/levy-mobility/>.
- [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. Feldmann and W. Whitt, "Fitting mixtures of exponentials to long-tail distributions to analyze network performance models," in *Proc. of IEEE INFOCOM '97*, Kobe, Japan, April 1997.
- [14] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, pp. 462–465, January 2006.
- [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] A. Vazquez, O. Sotolongo-costa, and F. Brouers, "Diffusion regimes in levy flights with trapping," *Physica A*, vol. 264, pp. 424–431, 1999.
- [18] J.-Y. L. Boudec and M. Vojnovic, "The random trip model: Stability, stationary regime, and perfect simulation the random trip model: Stability, stationary regime, and perfect simulation," *IEEE/ACM Trans. on Networking*, vol. 14, no. 6, 2006.
- [19] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic, "Power law and exponential decay of inter contact times between mobile devices," in *ACM Mobicom*, Montreal, Canada, September 2007.
- [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] 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.
- [22] 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.
- [23] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, pp. 207–211, May 2005.