SLAW: Self-Similar Least-Action Human Walk

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

***Abstract*—Many empirical studies of human walks have reported that there exist fundamental statistical features commonly appearing in mobility traces taken in various mobility settings. These include: 1) heavy-tail flight and pause-time distributions; 2) heterogeneously bounded mobility areas of individuals; and 3) truncated power-law intercontact times. This paper reports two additional such features: a) The destinations of people (or we say waypoints) are dispersed in a self-similar manner; and b) people are more likely to choose a destination closer to its current waypoint. These features are known to be influential to the performance of human-assisted mobility networks. The main contribution of this paper is to present a mobility model called Self-similar Least-Action Walk (SLAW) that can produce synthetic mobility traces containing all the five statistical features in various mobility settings including user-created virtual ones for which no empirical information is available. Creating synthetic traces for virtual environments is important for the performance evaluation of mobile networks as network designers test their networks in many diverse network settings. A performance study of mobile routing protocols on top of synthetic traces created by SLAW shows that SLAW brings out the unique performance features of various routing protocols.**

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

I. Introduction

THE PERFORMANCE of mobile networking applications highly depends on the movement patterns of wireless device holders. As wireless devices are often carried by people, understanding human mobility patterns contributes to an accurate performance modeling and prediction of protocols used for these networks. In particular, these patterns can be used for realistic simulation of human-assisted mobile networks. Since simulation is a primary means of performance evaluation in mobile networking, mobility models reproducing realistically inherent and invariant mobility patterns of people are important for accurate performance evaluation.

The goal of this paper is to develop a human mobility model that abstracts out many geographically specific details that might change from one setting to another, and that faithfully reproduces fundamental and invariant statistical properties of human mobility. Because users of a mobility model may test their protocols in diverse mobility environments including user-created virtual settings, often many geographically specific details are not available. Therefore, the model must be simple to manipulate. If users are required to provide about each mobility scenario many specific details that are available only from real traces, such a mobility model is not useful. It is desirable that many diverse location contexts, hotspots, and popular gathering places and their related mobility information such as transition probabilities from one place to another are automatically generated only from a simple set of input parameters, yet the resulting mobility traces must contain sufficient "realism." Note that it is hard to define realistic mobility for such virtual environments. We address this realism by representing faithfully the fundamental statistical properties commonly observed in many real mobility traces of people independent of their mobility settings.

There have been several recent studies [1]–[3] reporting the discovery of fundamental statistical properties of human mobility from real traces of human mobility such as GPS traces of human walks[1] in various locations [1], cell-phone location tracking [2], recordings of wireless device associations with their access points [5], and tracking of bank notes [3]. The following list presents three statistical features observed by the above studies (F1–F3) as well as two new features (F4–F5) that are reported in this paper for the first time.

F1) *Heavy-tail flights and pause-times:* It is shown that the lengths of human flights have a heavy-tail distribution [1]–[3]. A *flight* is a Euclidean distance between two *waypoints* visited in succession by the same person in the same daily trip, and waypoints are the locations where a walker stops for longer than a certain period of time before moving again. Waypoints are intuitively considered as destinations where people stop their travel. Several studies (e.g., [1] and [5]) also show that

[1]We use the terminology *human walks* to represent a general mobility involving humans in a predefined relatively small region such as university campus, airport, shopping malls, and a segment of a city. Mobility in this region is typically carried out by walking, jogging, and running, but can occasionally involve use of transportation such as a car, a bus, and subways.

Manuscript received June 25, 2010; revised January 01, 2011 and June 28, 2011; accepted July 12, 2011; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor C. Westphal. Date of publication November 10, 2011; date of current version April 12, 2012. This work was supported in part by the National Science Foundation under Grants CNS-0910868 and CNS-1016216, the Korea Communications Commission (KCC), Korea, under the R&D program supervised by the Korea Communications Agency (KCA) (KCA-2011-11913-05004)). A preliminary version of this paper was published in the Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Rio de Janeiro, Brazil, April 19–25, 2009. A major portion of this work was done while K. Lee was with KAIST and S. Hong and S. J. Kim were with NCSU.

K. Lee and I. Rhee are with the Department of Computer Science, North Carolina State University, Raleigh, NC 27695 USA (e-mail: klee8@ncsu.edu; rhee@ncsu.edu).

S. Hong is with Samsung Advanced Institute of Technology, Yongin 446-712, Korea (e-mail: seongik.hong@samsung.com).

S. J. Kim is with the DMC Research Center, Samsung Electronics, Suwon 442-742, 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.2011.2172984

the pause-time distributions of human walks follow a truncated power-law distribution. Pause-time is the time duration that a person spends in a waypoint.

F2) *Heterogeneously bounded mobility areas:* Gonzalez *et al.* [2] report that people mostly move only within their own confined areas of mobility and that different people may have widely different mobility areas.

F3) *Truncated power-law intercontact times (ICTs):* The distribution of intercontact times—that is, the time elapsed between two successive meetings of the same persons—can also be modeled by a truncated power-law distribution [4], which consists of a power-law head followed by an exponentially decaying tail after a certain characteristic time [6].

F4) *Self-similar waypoints:* We report in this paper that the waypoints of humans can be modeled by self-similar points. The self-similar dispersion of waypoints intuitively implies that people are always more attracted to more popular places, their visiting destinations tend to be heavily clustered, and such clustering patterns are persistent in various spatial scales.

F5) *Least-Action Trip Planning (LATP):*. We report in this paper that people are more likely to visit destinations nearer to their current waypoint when visiting multiple destinations in succession.

All these properties (*F1–F5*) are intrinsically related to each other. We show in this paper that people planning their daily trips using LATP (*F5*) on top of self-similar waypoints (*F4*) produce heavy-tail flight patterns (*F1*). It is shown in [7] that heavy-tail flights within a confined area (*F2*) result in truncated power-law ICTs (*F3*). Note that self-similar waypoints and power-law flight distributions are related, but not necessarily equivalent. For instance, one may generate power-law flights by selecting a random distance d from a power-law distribution, and then arbitrarily pick a destination randomly among the points located at the radius of distance d from the current waypoint, as it is done in truncated Levy walk (TLW) [1]. When TLW runs in an infinite space, this may provide self-similar waypoints [8], [9]. However, a random selection of destinations within a confined area of mobility does not preserve self-similarity of waypoints. Conversely, one may generate a self-similar point process of waypoints. However, if the sequence of visits over these self-similar destinations is not judiciously selected, the resulting flights are not necessarily power-law.

It is also important that these properties are strong performance determinants of mobile networks such as delay-tolerant networks (DTNs). For instance, ICTs are a very important factor to DTN routing as ICTs decide the delays in meeting a distant node, so short ICTs imply short routing delays. Self-similar waypoints express some social contexts such as gathering places among people that share common interests or those in the same community. These contexts are important as they influence meeting probabilities and periodicity among people. The diffusivity of random walks induced by the heavy-tail flights also has a unique impact on the link lifetimes and intercontact times of mobile networks [1]. Therefore, for an accurate performance evaluation of mobile networks, a mobility model reproducing synthetic mobility traces with a realistic representation of these properties is important.

TABLE I
EXISTING ROUTING MODELS CAN BE CATEGORIZED INTO FOUR GROUPS. TYPICAL MODELS IN EVERY CATEGORY HAVE BEEN LISTED. NONE OF THE EXISTING MODELS HAS ALL THE CHARACTERISTICS OF HUMAN WALKS. "?" MEANS THAT IT IS UNCLEAR FROM THE MODEL DESCRIPTION

Category	Features / Models	$F1$	$F2$	$F3$	$F4$	$F5$
Pure Random	RWP, RD	N	N	N	N	N
	BM(RW)	N	N	Y	N	N
	TLW [1]	Y	N	Y	N	N
Pure Random Variants	MWP [12]	N	N	N	N	N
	GM [24]	N	N	?	N	N
	RPGM [13]	N	N	?	N	N
Geographic	Freeway [25]	N	N	?	N	N
	Manhattan [25]	N	N	?	N	N
	OM [26]	N	N	?	N	N
Social	Dartmouth [5]	N	N	?	N	N
	CMM [11]	N	N	?	N	N
	ORBIT [14]	N	Y	?	N	N

Unfortunately, none of the existing mobility models for mobile networks produces mobility traces that possess all these properties. For instance, no mobility model, except TLW [1], explicitly models heavy-tail flights. Random mobility models such as TLW, random waypoint (RWP), and Brownian motion (BM) do not produce self-similar waypoints, nor do they model heterogeneous mobility ranges of individuals. Although many models (e.g., [5] and [10]–[14]) explicitly model hotspots or grouping effects, they do not produce heavy-tail flights or ICTs. Those producing heavy-tail ICTs (e.g., [6] and [15]) also do not necessarily model heavy-tail flights. Table I summarizes the statistical properties modeled by many existing mobility models.

In this paper, we present a new mobility model, called *Self-similar Least-Action Walk* (SLAW), that produces synthetic mobility traces containing all five properties. This is the first such model. In developing SLAW, we heavily rely on our GPS traces of human walks [16] including 226 daily traces collected from 101 volunteers in five different outdoor sites. In particular, many of these traces are gathered among people sharing common interests such as students in the same university campuses and tourists in a theme park. By faithfully representing the properties present in these traces, SLAW can represent social contexts among walkers manifested by visits to common gathering places and walk patterns therein.

SLAW mainly takes two input parameters, a Hurst parameter value [17] and a weight on distance in performing LATP, along with other trace-specific information such as the size of the mobility areas and the number of mobile nodes. The Hurst parameter value determines the degree of self-similarity for waypoint dispersion. The weight factor for LATP determines the likelihood of visiting nearby destinations when a node has multiple candidate destinations. Users of the model do not need to extract these parameter values from real traces. Instead, we provide the ranges of values for users to choose from for the parameters. The ranges are determined by the statistical properties of their corresponding properties represented by the parameters. For instance, the Hurst value must be in between 1/2 and 1 to ensure the self-similarity of waypoints.

People typically visit the same places every day such as offices and restaurants while making some infrequent irregular trips caused by exceptions such as appointments. By modeling power-law flights, self-similar waypoints, and LATP, SLAW can realistically express regular as well as spontaneous trip patterns of human daily mobility. While other works [18]–[20] model the regularity of daily trip patterns of humans, none of the existing work reflects realistic statistical properties appearing in real human walks.

To measure the impact of these mobility patterns on the performance of mobile network protocols, we study the performance of DTN routing under various mobility models, including SLAW. Our study indicates that SLAW realistically captures the unique performance properties of many existing DTN routing protocols. More specifically, it provides a clear performance differentiation between stateless and stateful protocols where stateful protocols require and utilize past contact information among nodes to predict future contact probability and stateless protocols do not. Examples of stateful protocols are abundant (e.g., [21]–[23]). SLAW induces more frequent and regular contacts among nodes that result in more predictable and shorter routing delays for those protocols. The applications of our work go beyond mobile networks. SLAW can be an important tool for emulating human walk behaviors in diverse application scenarios that can be applied to accurate urban planning, traffic forecasting, and biological and mobile virus spread analysis.

The remainder of this paper is organized as follows. Sections II and III respectively present related work and our findings on self-similar waypoints and LATP. Sections IV and V describe SLAW, the validation of SLAW, and our routing protocol study. Section VI presents our conclusion.

II. Related Work

Table I shows four categories of existing models. RWP, Random Direction (RD), BM, or Random Walk (RW) and TLW [1] are pure random mobility models. In pure random mobility, each waypoint is chosen randomly based on some probability distribution. The Markovian Waypoint (MWP) [12] and Gauss–Markov (GM) model [24] are variants of the above, as they implement some Markovian transition probabilities among waypoints or prohibit unrealistic abrupt velocity changes. In RPGM [13], mobile nodes form several groups, each of which contains one leader. The leader moves according to the RWP, and all other members of a group move along their leader.

Other models consider geographical constraints or social contexts and collective behaviors. The Obstacle Model (OM) with geographical constraints [26] incorporates obstacles for emulating more realistic pathways of humans around obstacles using Voronoi diagrams. The Freeway and Manhattan mobility models [25] emulate pathways by restricting the movements of mobile nodes to follow the pathways.

Modeling hotspots is another way to represent collective human behaviors. The Dartmouth model [5] estimates the locations and movement paths of mobile nodes from real data sets. Based on the estimated information on the users, hotspot regions and the transition probability for moving between hotspots are extracted. This model is developed using the trace data collected by analyzing the access patterns of wireless access points in a university campus. It requires a considerable amount of effort to generate the mobility model because hotspot locations and transition probabilities between hotspots must be given as input (instead of being generated). Thus, it is very hard to change the walkabout areas, the number of nodes, and the locations of hotspots without any corresponding real data sets.

Some hotspot models [10], [11] use scale-free networks [27]. Using the preferential attachment theory [28], a set of *attractors* is established, where attractors are either landmarks or nodes. For instance, Clustered Mobility Model (CMM) [11] is a good example. It first divides the simulation area into a number of subareas and uses them as attractors. Mobile nodes are assigned to a subarea using preferential attachment. The attractiveness of one area is determined by the current number of nodes assigned to that area.

ORBIT [14] randomly creates a specified number of clusters within a given area, and each node is assigned to a subset of clusters. A node moves only among its assigned clusters. The movements between and within clusters are random irrespective of any properties of clusters (e.g., size, distance). ORBIT explicitly models the heterogeneously bounded walkabout areas of mobile nodes. ORBIT does not capture self-similar points, and next destinations are selected uniformly at random. Thus, it lacks the regular patterns present in daily human walks.

Modeling social relationships among people is an interesting way to express mobility patterns. The models proposed in [15], [29], and [30] first construct a social interaction matrix that quantifies the degree of attraction among people. The matrix is then used for computing the transition probability of a person to move from one location to another. These models are based on the intuition that people are attracted to locations where socially close people are gathered around. However, this intuition is not verified.

Hsu *et al.* [31] incorporate to a random walk model the tendency to return to home after some period of time. Using their model, they show the probability of meeting among nodes that can be useful for the performance analysis of mobile networks. However, this model does not represent the realistic statistical patterns of human mobility.

III. Measurement Study of Human Walks

A. GPS Data

We use the same data used in [1]. We describe the data briefly, but for more detailed information, the readers can refer to [1]. Garmin GPS 60CSx handheld receivers are used for data collection, and they are Wide Area Augmentation System (WAAS) capable with a position accuracy of better than 3 m 95% of the time in North America. The GPS receivers automatically record their current positions at every 10 s into a daily track log. The total number of traces from these sites is over 226 daily traces containing more than 200 000 flight samples. The participants in Campus II, New York City (NYC), Disney World (DW), and State Fair (SF) traces are randomly chosen, while those from

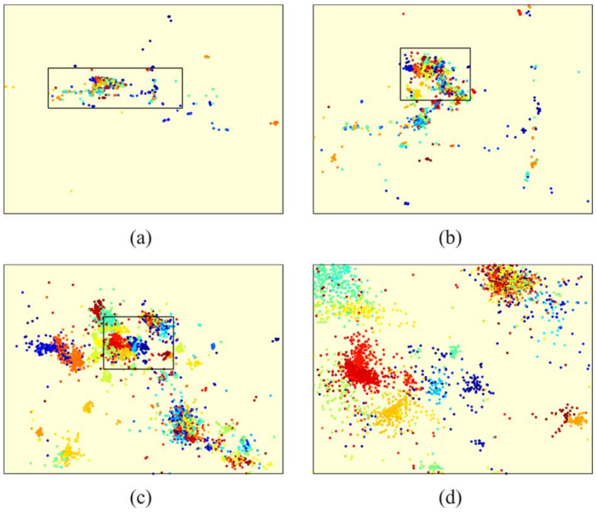


Fig. 1. (a) Campus-II waypoints, and (b)–(d) the bursty dispersion of Campus-II waypoints over different scales. The areas shown are (a) 9781×20902 m^2, (b) 4800×4800 m^2, (c) 1200×1200 m^2, and (d) 300×300 m^2. Different colors are used for different daily traces.

TABLE II
STATISTICS OF COLLECTED MOBILITY TRACES FROM FIVE SITES. THIS TABLE SHOWS THE NUMBER OF PARTICIPANTS, THEIR DAILY TRACES, AND 30-S AVERAGE SAMPLES. IT ALSO SHOWS TIME AND GEOGRAPHICAL INFORMATION

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

Campus I are students from the same department. Table II shows the summary of our daily traces.

The spatial resolution of human mobility traces used in other studies [32] is very low since their traces are collected using association traces to WiFi access points [33], [34] or cellular towers or base stations [2]. The location errors of their data are at most a few hundred meters. Needless to say, the banknote trace data [3] have higher location errors. Compared to those data, our data have much more detailed representations of human walks as they are taken using GPS at every 10-s interval in a few meters scale

B. Self-Similar Waypoints

In [1], we showed that the power-law slopes of the heavy-tail flight distributions from our traces are different from one site to another. This implies that the patterns of human mobility are highly influenced by the geographical contexts such as locations of their destinations. To analyze this behavior in more detail, we register the locations from the GPS traces where participants stop for longer than 30 s within a radius of 5 m, and we call them *waypoints*. Fig. 1 shows the waypoints aggregated from all traces from Campus II over decreasing scales. We can visually inspect that the waypoints are dispersed in a bursty manner forming clusters. People tend to swarm near to a few popular locations, and their popularity measured by the number of waypoints within the swarms of waypoints shows high burstiness.

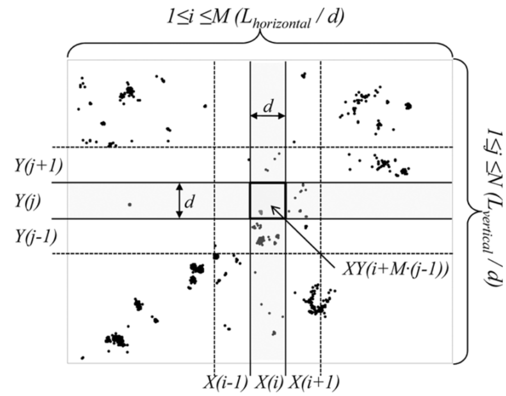


Fig. 2. Measuring aggregated variance of waypoints aggregated from all walk traces. We divide the area by nonoverlapping d by d squares, and count the number of waypoints registered in each square and then normalize the sampled count by the size of the unit square. We compute the normalized variance as we increase d.

Moreover, the burstiness does not disappear as the scale varies, meaning that the waypoint dispersion shows some degree of self-similarity.

Formally, a stochastic process is called *self-similar* or *long-range-dependent* if its auto-correlation function decays slowly [35]. Intuitively, this slow decay indicates a high degree of correlation between distantly separated points of the process. Self-similarity is usually quantified by the *Hurst* parameter, and several methods for measuring this parameter from traces exist in literature [36]. In what follows, we use two such well-known methods to quantify the self-similarity of the waypoint dispersion in our traces, namely the *aggregated variance* and the *R/S* methods.

In the aggregated variance method, we divide the site map by a grid of unit squares (initially of 5×5 m^2), count all the waypoints within each unit square, and then normalize the count by the area of the unit square. Then, we measure the variance in these normalized count samples. Fig. 2 illustrates the method. If there exists long-range dependency in the samples, the aggregated variance should not decay faster than -1 in a log-log scale as we increase the size of the unit square. To see this, we plot aggregated variance in a log-log scale as we increase the unit square size and measure its absolute slope β. The *Hurst* parameter of the samples is defined to be $1 - \beta/2$. The sample data are said to be self-similar if the Hurst parameter is in between 0.5 and 1. Aggregated variance can also be computed over one dimension by mapping waypoints to the x- or y-axis of the map.

Fig. 3 shows the Hurst parameter measured from the aggregated waypoints of Campus II. These values show a self-similarity with a Hurst value larger than 0.7. Fig. 4 shows the Hurst parameter values measured from the aggregated variance test using the aggregated traces of each site over one-dimensional (x and y) and two-dimensional spaces. All values except NYC are over 0.6. We conjecture that the aberration of the NYC traces comes from the small number of waypoints relative to the size of the site.

We also observe that the waypoints registered in each individual trace are self-similar. For each trace, we perform the aggregated variance test on its waypoints. Fig. 5 shows the Hurst parameter values of individual traces from the five sites by the

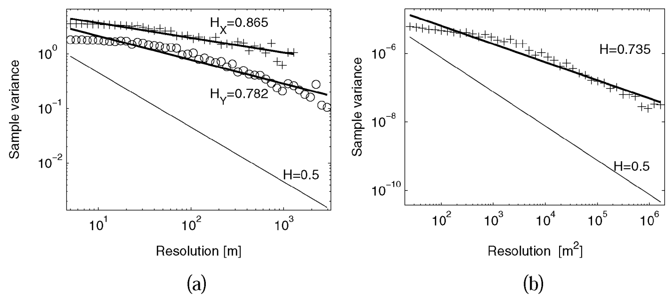


Fig. 3. Hurst parameter estimation of waypoints registered in all Campus-II traces by the aggregated variance method. (a) 1-D stripe. (b) 2-D grid.

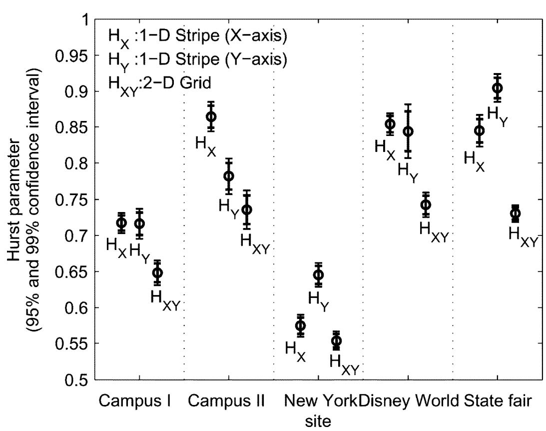


Fig. 4. Hurst parameter values (with 95% and 99% confidence interval) of waypoints extracted from the aggregated traces of each site using the aggregated variance method. Their values indicate the self-similarity of waypoints.

aggregated variance method. Their H-values are slightly less than those from the aggregated waypoints shown in Fig. 4. This also confirms the self-similarity of waypoints as the burstiness gets intensified as individually bursty traces are superimposed together.

In the R/S method, for a one-dimensional data set $X = \{X_i, i \geq 1\}$ with partial sum $Y(n) = \sum_{i=1}^{n} X_i$, the R/S statistic, or the *rescaled adjusted range*, is given by

$$\frac{R}{S}(n) := \frac{1}{S(n)}\left[\max_{0\leq i\leq n}\left(Y(i) - \frac{i}{n}Y(n)\right) - \min_{0\leq i\leq n}\left(Y(i) - \frac{i}{n}Y(n)\right)\right]$$

where $S^2(n)$ is sample variance. For a two-dimensional point process data set, the number of points $X = \{X_{i,j}\}$ (the size of a unit square is chosen to be identical with that of the aggregated variance method), with partial sum $Y(n,m) = \sum_{i=1}^{n}\sum_{j=1}^{m} X_{i,j}$, the R/S statistic can be modified as follows:

$$\frac{R}{S}(n^2) := \frac{1}{S(n^2)}\left[\max_{0\leq i,j\leq n}\left(Y(i,j) - \frac{\sqrt{i^2+j^2}}{\sqrt{2\cdot n^2}}Y(n,n)\right) - \min_{0\leq i,j\leq n}\left(Y(i,j) - \frac{\sqrt{i^2+j^2}}{\sqrt{2\cdot n^2}}Y(n,n)\right)\right]$$

where $S^2(n^2)$ is sample variance of $n \times n$ data.

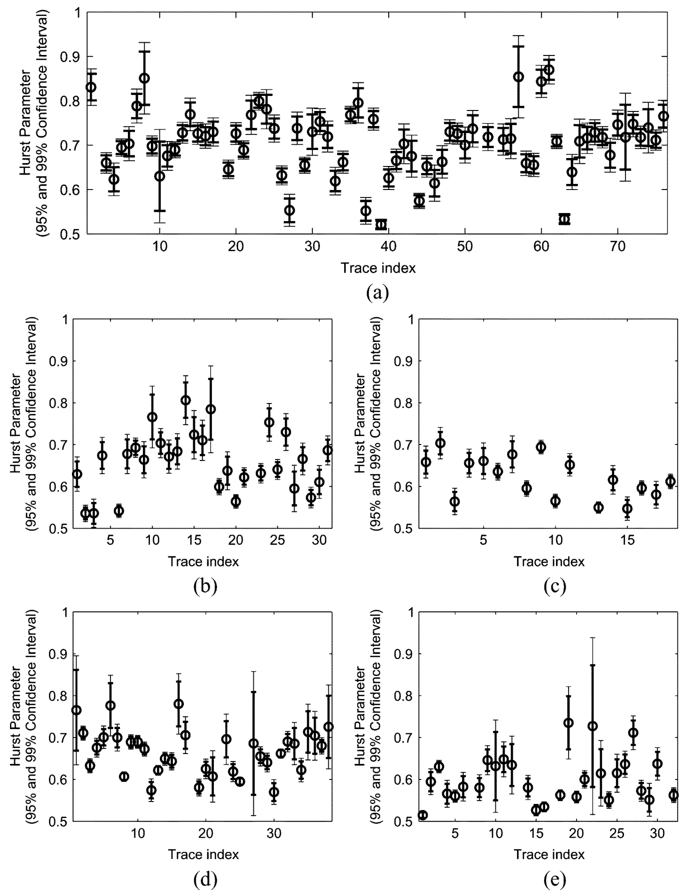


Fig. 5. Hurst parameter values (estimated by the aggregated variance method) of waypoints registered in each individual trace of the five sites. All traces show a tendency of self-similarity. (a) Campus II. (b) Campus I. (c) State Fair. (d) Orlando. (e) New York City.

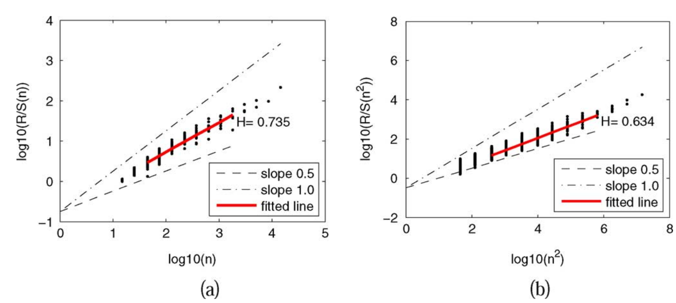


Fig. 6. Hurst parameter estimation of waypoints registered in the Campus-I traces by the R/S method. (a) 1-D (x-axis) stripe. (b) 2-D grid.

Fig. 6 shows the Hurst parameter estimated from the Campus-I traces by the R/S method, and Fig. 7 shows the Hurst parameter values of all traces by the R/S method, which confirms the self-similarity of waypoints. The average Hurst parameter values estimated from the two tests are summarized in Table III.

C. Gap Distributions

Flights are line trips over these waypoints. The order in which a walker visits these waypoints determines his flight patterns. Then what aspects of self-similar waypoints induce a heavy-tail

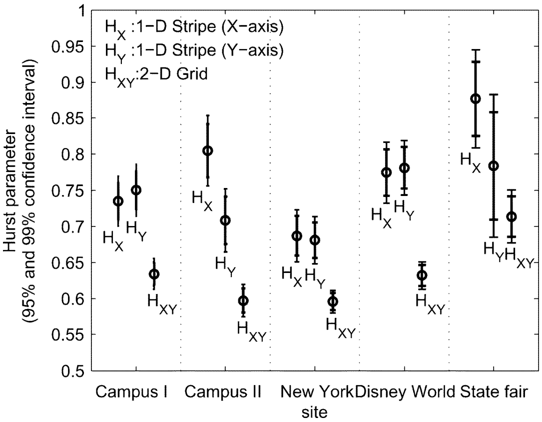


Fig. 7. Hurst parameter values (with 95% and 99% confidence interval) of waypoints extracted from the aggregated traces of each site using the R/S method. Their values indicate the self-similarity of waypoints.

TABLE III
ESTIMATED HURST VALUES BY THE AGGREGATED VARIANCE AND R/S METHODS. BOTH METHODS CONFIRM THAT THEY ARE ALL SELF-SIMILAR

Site	Aggregated variance			R/S		
	H_{XY}	H_X	H_Y	H_{XY}	H_X	H_Y
Campus I	0.648	0.717	0.716	0.634	0.735	0.750
Campus II	0.735	0.865	0.782	0.597	0.805	0.708
NYC	0.554	0.575	0.645	0.596	0.687	0.681
DW	0.742	0.854	0.844	0.632	0.775	0.781
SF	0.730	0.845	0.904	0.714	0.877	0.784

flight distribution? To find this relation, we study the characteristics of the "gaps " formed among these waypoints.

We first examine the relation between one-dimensional self-similar points and their corresponding gap distribution, and later we generalize our analysis to two-dimensional cases through simulation. Consider a process dispersing a set of points over one-dimensional space. Let $Y(x)$ be the number of points over a line interval (x and $x+\Delta$). In other words, $Y(x)$ is a point-count sequence over a small interval Δ. The self-similarity of point counts $Y(x)$ can be manifested in several equivalent ways. First, the aggregated variance $v(m)$ of $Y(x)$, which is a variance of a new series by averaging the original series $Y(x)$ for nonoverlapping blocks consisting of m elements, replacing each block by its mean, has an asymptotic form of $v(m) \sim m^{-\beta}$, $0 < \beta < 1$ as $m \to \infty$. The Hurst parameter can be expressed as $H = 1 - \beta/2$. Second, the power spectrum $S(f)$ of $Y(x)$ has $1/f$ noise around the origin, that is, $S(f) \sim f^{-\theta}$, $0 < \theta < 1$ as $f \to 0$.

The *gap* (interval) between two points can be measured as follows. We first make Δ small enough to hold at most one point and define the distance between any two immediately neighboring points as *gap*. Let $p(x)$ be the probability density function of a random variable x representing a gap among the self-similar point process $Y(x)$.

Theorem 3.1: Self-similar points over one-dimensional space induce power-law gaps, that is, if $v(m) \sim m^{-\beta}$, $0 < \beta < 1$ as $m \to \infty$, then $p(x) \sim x^{-\alpha-1}$, $0 < \alpha < 1$ as $x \to \infty$. Furthermore, $\alpha + \beta = 1$.

Proof: Due to the space limit, we provide the following proof sketch. The proof is adapted from [37] and [38].

From [39], we can have the following relation between $v(m)$ and the autocorrelation function $\rho(x)$ of $Y(x)$:

$$v(m) = v \cdot \left(\frac{1}{m} + \frac{2}{m^2} \sum_{n=1}^{m} (m-n) \cdot \rho(n) \right) \quad (1)$$

where v is the variance of $Y(x)$.

From (1), $\rho(x) \sim x^{-\beta}$ since $v(m) \sim m^{-\beta}$.

A correlation function $c(x)$ of $Y(x)$ is defined as a conditional probability to have a point at x, given that a point occurs at $x = 0$. It is asymptotically the same as the autocorrelation function of $Y(x)$, $\rho(x)$. By definition, $c(x = n\Delta)$ can be represented as follows [40]:

$$c(\Delta) = p(\Delta) \quad (2)$$

$$c(2\Delta) = c(\Delta)p(\Delta) + p(2\Delta) \quad (3)$$

$$\vdots$$

$$c(n\Delta) = \sum_{k=0}^{n} c((n-k)\Delta)p(k\Delta) + \delta_{n,0} \quad (4)$$

where $p(0) = 0, c(0) = 1$, and $\delta_{n,0}$ is the Kronecker delta function, which becomes 1 when $n = 0$, otherwise 0.

The Fourier transform $\mathcal{F}$ of the last equation gives a power spectrum of $Y(x)$. Thus, $S(f) = \mathcal{F}(c(x)) = \frac{1}{1-\mathcal{F}(p(x))}$. Let $\hat{P}(f)$ be an asymptotic function of $\mathcal{F}(p(x))$

$$\hat{P}(f) = 1 - \frac{1}{S(f)} \quad (5)$$

Since $c(x) \sim x^{-\beta}$, $S(f) = \mathcal{F}(c(x)) \sim f^{\beta-1}$ as $f \to 0$. Since $0 < \beta < 1$, the power spectrum of $Y(x)$ has $1/f$ noise.

If we set $\alpha = -\beta+1$, then from (5), $\hat{P}(f) \sim 1-f^{\alpha}$. Note that the inverse Fourier transform of $1-Bf^{\theta}, 0 < \theta < 1$, where B is a constant, is asymptotically $x^{-\theta-1}$. Therefore, $p(x) \sim x^{-\alpha-1}$ as $x \to \infty$. Since $0 < \alpha < 1, p(x)$ is asymptotically power-law, and it proves $\alpha + \beta = 1$. ∎

So far, we provided analytical evidence that gaps over self-similar waypoints over one-dimensional space induce power-law gaps. Since it is hard to define two-dimensional gaps as well as "neighboring" points in 2-D, researchers [41] have shown empirically that two-dimensional gaps over self-similar points have power-law distributions using *Delaunay triangulation*. In these studies, Delaunay triangulation is commonly used to measure two-dimensional gaps [41] as it practically identifies the neighboring points. Formally, Delaunay triangulation for a set P of points in the plane is a triangulation $\mathrm{DT}(P)$ such that no point in P is inside the circumcircle of any triangle in $\mathrm{DT}(P)$.

We perform a Delaunay triangulation over the waypoints extracted from our traces. To illustrate our process of analysis, Fig. 8 shows Delaunay triangles on top of a daily trace of one participant, and in the inset, the complementary cumulative distribution function (CCDF) of the length of triangle sides and flights extracted from the same trace. It is visually striking that the Delaunay triangles and flight patterns and their corresponding distributions are very similar. We explore this point further in the Section III-D.

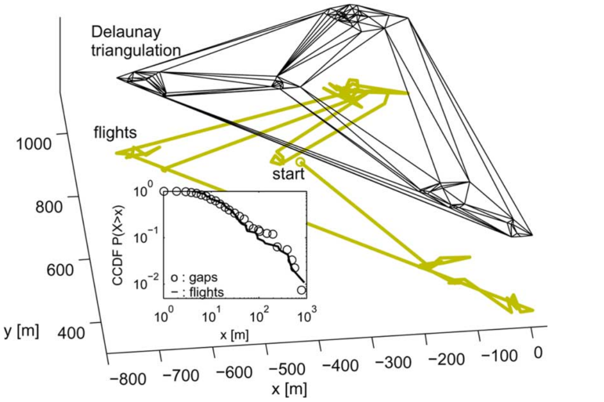
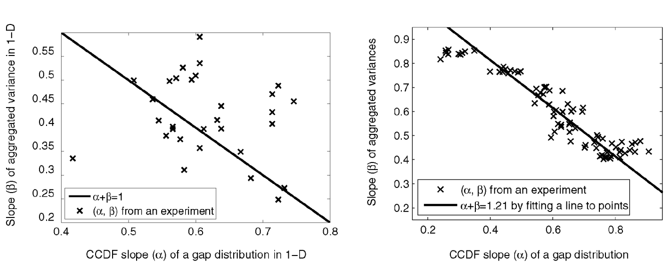


Fig. 8. Delaunay triangulation of waypoints extracted from one daily trace of Campus II. The measured flights involving the waypoints coincidentally resemble the sides of the triangles. In the inset, the CCDFs of the lengths of triangle sides and the flights from the same trace are plotted. The CCDF of flights and gaps are also closely matching.



(a) (b)

Fig. 9. (a) α and β measured with 1-D projected waypoints from traces of Campus II. (b) α and β measured over synthetic 2-D waypoints.

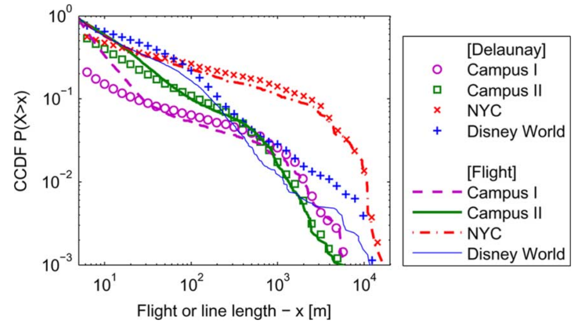


Fig. 10. Delaunay triangulation is performed on all individual daily traces. The line segment lengths in Delaunay triangles are aggregated, and their CCDFs are plotted for different walkabout sites. The CCDF of flights obtained from the corresponding traces are also plotted for comparison.

We measure the power-law slope of the aggregated variances (β) of waypoints extracted from the GPS traces and the power-law slopes of the CCDF of their corresponding gap distributions (α). As we proved, we observe that the values of $\alpha + \beta$ from real traces projected to 1-D follow 1 as shown in Fig. 9(a). For 2-D traces, we find that the values of $\alpha + \beta$ from real traces are close to 1.2. The margin of errors in 2-D may arise from truncations caused by confined measurement areas as the truncations may significantly distort the power-law slope visible at the body of distribution. We also perform Delaunay triangulation on top of synthetically generated self-similar points using a simplified fBm technique [16] over a two-dimensional area and measure α and β. For each Hurst parameter value (H), we generate 10 synthetic waypoint maps. Fig. 9(b) shows that $\alpha + \beta \approx 1.2$ with a similar margin of errors.

D. Relation Between Heavy-Tail Flights and Gaps Over Self-Similar Waypoints

In this section, we examine the hints for how gaps over self-similar waypoints are related to heavy-tail flights. We perform Delaunay triangulation on each individual daily trace in our data and aggregate the lengths of all the resulting triangle sides into a single CCDF. Fig. 10 plots the result along with the CCDF of flights from the traces. The similarity in the shapes of the two CCDFs is clearly visible. People do not consciously consider Delaunay triangles when they plan their trips over multiple destinations. However, the similarity between the gap and flight distributions suggests that there might be a connection between the sequence of visits to multiple destinations (i.e., waypoints) and gaps among self-similar points. Since Delaunay triangulation maximizes the minimum angle of all the angles of the triangles in the triangulation, they tend to avoid skinny triangles. By avoiding skinny triangles, these triangles are formed using nearby points rather than farther points. Thus, we conjecture that the flight traces may have similar tendency: They are more likely to visit nearby destinations before visiting farther destinations. We explore this conjecture further in Section III-E.

E. Least-Action Trip Planning

The similarity in the distributions of the gaps and flights suggests interesting aspects about the order in which a walker visits for a given set of waypoints. Obviously, people are not conscious about gaps when they travel. However, as we can see from Fig. 8, Delaunay triangles are formed among "neighboring" waypoints as Delaunay triangulation produces a planar graph. This may imply that people tend to minimize the traveling distance. Intuitively, when people are to visit multiple destinations located at different distances, people often strive to minimize the total distance of travel. They do this rough minimization by visiting nearby destinations before visiting farther destinations instead of visiting father destinations first and then coming back to nearby destinations.

In fact, this "greedy" way of trip planning is similar to a heuristic to the traveling salesman problem whose objective is to minimize the total distance of travel and aligns well with the least-action principle of Maupertuis [42]. Intuitively, the least-action principle conjectures that all the objects in the universe move toward the direction of minimizing their discomfort. This principle is also used to explain how people make their walking trails in public parks [43].

In our particular situation, this discomfort can be considered the traveling distance. Note that it is entirely possible that people may also bypass nearby unvisited destinations to get to farther destinations and then come back to the nearby ones. This may happen when people have prior engagements with high-priority or time-critical tasks so that they may have to attend to them first. Therefore, there might be the tradeoff between the importance of attending to those special events and the discomfort of traveling longer distances. In this section, we analyze this tradeoff.

We first measure the amount of weight that people put on distance when choosing their next destinations. We estimate this as

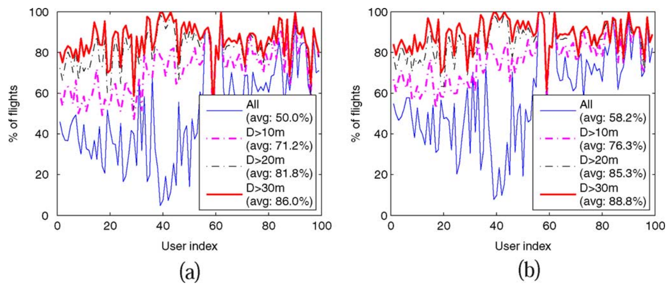


Fig. 11. From the GPS traces of 100 participants, we measure the percentage of flights meeting the least-action criterion. We also plot the case when we exclude those flights whose length is less than 10, 20, and 30 m. (a) $r = 1.5$. (b) $r = 2.0$.

follows. We first measure the *flight-to-nearest-waypoint* (FNW) ratio. For a given flight from x to y, suppose k is the nearest unvisited waypoint from x. The FNW ratio is the ratio of $||x - y||$ over $||x - k||$. We then define the *least-action criterion*: For a given flight, it tests if its FNW ratio is less than some threshold. Fig. 11 plots the percentage of flights meeting the least-action criterion in real traces for all participants when the threshold ratio (r) is less than 2.

The FNW test measures the number of occurrences in each daily trace that a person visits the waypoints located within r times the distance to the nearest unvisited waypoint. It measures how well people meet the criterion when they choose next waypoints. Here, we only consider unvisited waypoints in performing the FNW test, but there is possibility that the endpoint of a flight coincides with a previously visited waypoint. However, among all the waypoints registered in our traces, we confirm that only two flights return to the previously visited waypoints. This can be explained as follows. Although people might come back to the same place repeatedly in a day, they do not necessarily come back to the exact spot in space and stop there to be registered as waypoints. Since the resolution of GPS readings is in meters, these cases are very rare. Therefore, considering only the unvisited waypoints for k does not have much impact on the accuracy of the FNW test.

On average, 58% of flights meet the criterion. However, people are less sensitive to the distance when next destinations are all nearby. Thus, if we exclude the flights whose length is less than a short distance (say 30 m), we get more than 88% of flights meeting the criterion on average. We also tried other distances such as 20 and 10 m and also varied r to a smaller ratio (1.5). All the measurements produce around 80% flights meeting the criterion as shown in Fig. 11. This indicates that most people in our traces may have used distance as an important metric for deciding the next waypoint.

We construct a new trip planning algorithm called LATP that, given a set of waypoints to visit, decides the order in which a person visits them. Algorithm 1 gives a pseudocode of LATP. The algorithm selects a next unvisited waypoint to visit based on a probability function $P(c, v)$, which uses a weighted function $1/d^a(c, v)$. $d(c, v)$ is the distance from the current waypoint c to an unvisited waypoint v, and a is a constant. If a is larger, then the algorithm is more likely to choose the nearer unvisited waypoint, and if it is zero, then it randomly chooses the next waypoint. LATP finishes when it visits all the unvisited waypoints. Visiting only unvisited waypoints is justified because waypoints are heavily clustered due to their self-similarity. People visiting the same hotspots repeatedly in a day are emulated by having them visit unvisited waypoints in close proximity to each other. This emulates repeated visits to the same hotspots because even if people visit the same hotspot repeatedly in a day, their exact GPS locations can be slightly different despite being in the same cluster.

Algorithm 1: *Least-action trip planning* (LATP) algorithm with a distance weight function d^a

V: set of all vertices (waypoints)
V': set of all visited vertices
$s \in V$: starting vertex
$c \in V$: current vertex
$c \Leftarrow s$
$V' \Leftarrow \{c\}$
while $V' \neq V$ **do**
  Calculate distances to all unvisited vertices,
  $d(c, v) = ||c - v||_2$ for all $v \in V - V'$
  Calculate probability to move to all unvisited vertices,
  $$P(c, v) = \frac{\left\{\frac{1}{d(c,v)}\right\}^a}{\sum_v \left\{\frac{1}{d(c,v)}\right\}^a} \text{ for all } v \in V - V'$$
  Choose a next vertex $v' \in V - V'$ according to the probabilities $P(c, v)$
  $c \Leftarrow v'$
  $V' \Leftarrow V' \cup \{c\}$
end while

Fig. 12 shows the resulting flight distributions obtained from this experiment with various values of a superimposed with the flight distribution obtained from the traces of Disney World and New York City. The other sites show similar trends. Visually, all distributions obtained from LATP fit extremely well to the real flight distributions, especially when a is between 1 and 3. The figure shows the difference of arithmetic sums between the LATP flights and real flights (marked as errors). It shows that when a is equal to 3, the difference is less than 2% in Disney World traces. In all site traces, the error margins are less than 11% with a between 1 and 3. We also measure the first three statistical moments of the LATP flight distributions obtained by performing LATP on top of waypoints extracted from real traces and measure their difference from those of the flight distributions obtained from the real traces. Fig. 13 plots the error percentiles for all the traces in which almost all the moments get the minimums around 1.5 and 3, respectively, with around 10% or less errors.

For Disney World and State Fair, a is close to 3 (so assigning a higher weight to distance), while the other traces have a between 1 and 2. Within a theme park, the objectives of the participants are likely to visit as many attractions as possible within a given time, so the traveling distance plays a bigger role. On the campus scenarios, people may have unexpected urgent events (e.g., appointments) that force them to make trips regardless of their traveling distances. The result indicates that

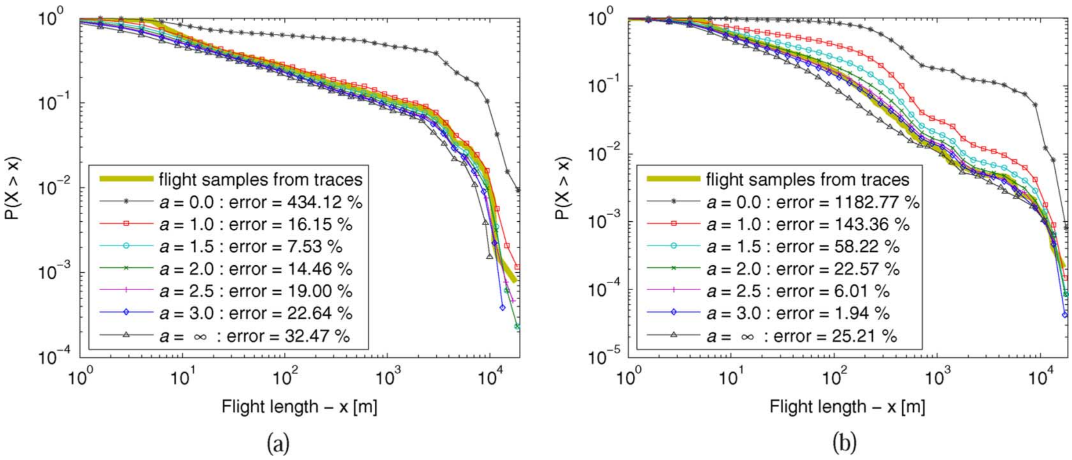


Fig. 12. Results of LATP using waypoints from (a) New York City and (b) Disney World traces. The algorithm is performed on each individual trace for various a-values.

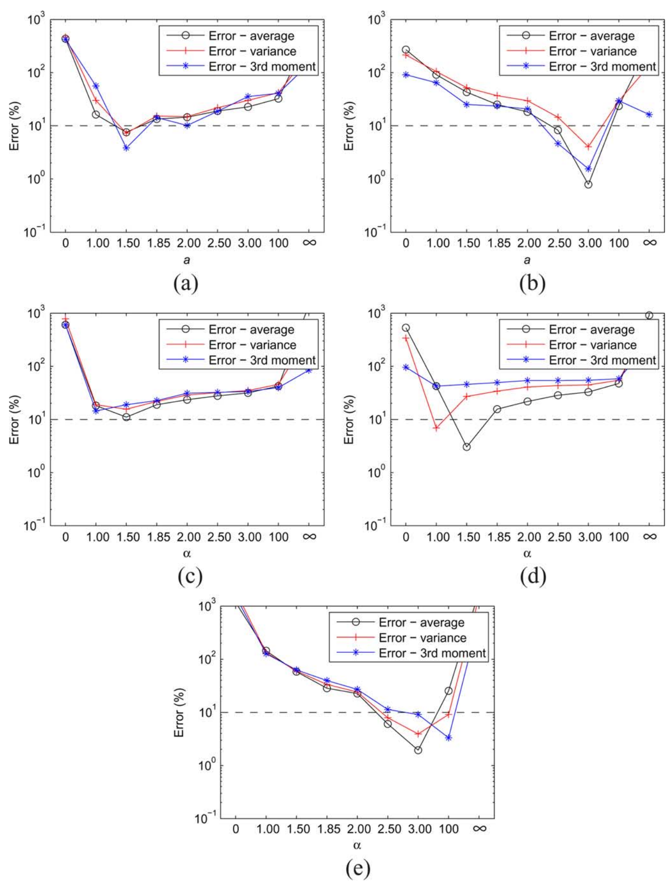


Fig. 13. Errors (%) of LATP in terms of the first three moments. (a) New York City. (b) State Fair. (c) Campus I. (d) Campus II. (e) Disney World.

LATP can recover almost identical flight distributions as the real ones from the traces and thus confirms that people use distance as an important factor in deciding the next destination of their trips. In Section IV, we show by simulation that when combined with self-similar points as found in the traces, LATP can generate traces with heavy-tail flights.

IV. SLAW: Self-Similar Least-Action Walk

A. SLAW Overview

Capturing both power-law flights and self-similar waypoints in walk traces is nontrivial because of mutual dependency among various parameters such as the degree of self-similarity (Hurst parameter) and the characteristics of flight distributions, e.g., the power-law slope of the distribution. SLAW adopts the following approach to the problem. It first generates self-similar waypoints using a technique similar to a fractional Gaussian noise or Brownian motion generation technique (fGn or fBm) (e.g., [44] and [45]) over a 2-D plane. Our analysis in Section III-C recovers algebraic relations between the Hurst parameter of self-similar points and the power-law slope of the corresponding gap distributions. This indicates that by controlling the Hurst parameter value, we can easily control the characteristics of gap distributions.

Gonzales *et al.* [2] report that people tend to make daily mobility within their own bounded areas. To emulate this behavior, SLAW develops an individual walker model restricting the mobility of each walker to a predefined subsection of the total area. It is done by selecting a subset of self-similar waypoint clusters and restricting the movement of each walker to its own designated set of clusters. Since people do not always maintain fixed routines, in order to add randomness, SLAW emulates this spontaneity in daily mobility by allowing walkers to choose one of the other clusters randomly for each day. From these selected clusters of waypoints, each walker chooses a set of waypoints for each daily trip and then applies LATP to select the order of visits over the selected waypoints. The initial point of the daily trips can be arbitrary.

We verify that this walker model combined with LATP and self-similar waypoints generates heavy-tail flights observed from real traces, and furthermore, the collection of individual traces generates power-law ICTs observed in [4].

B. Individual Walker Model

For a given input area S, our self-similar waypoint generation generates a set W of the waypoints. We propose an individual walker model that selects a subset of W and specifies the order in which those selected waypoints are visited. When selecting these waypoints, we need to be careful. Self-similar waypoints have a tendency of creating bursty clusters of various sizes dispersed over S. If waypoints are uniformly selected from W, then it is most likely that all walkers will traverse through most clusters and do not have heterogeneously bounded areas of mobility [2]. To define the heterogeneously bounded areas of mobility, we heuristically define an individual walker model that assigns different walkabout areas to different walkers and restricts each walker to move only around its designated area.

We first build clusters of waypoints by transitively connecting waypoints within a radius of 100 m. The radius represents a typical distance among buildings belonging to the same department in a campus. The clustering radius can be varied to study different scales of clusters (e.g., cities in a nationwide trace). Let $C = \{c_i, i = 1, \ldots, n]\}$ be the cluster set, $|c_i|$ be the number of waypoints, and T be the total number of waypoints in S. We assign a weight $|c_i|/T$ to each cluster i.

According to our GPS traces, each participant in Campus I, Campus II, New York City, Disney World, and State Fair visited 4.55, 3.66, 6.13, 3.34, and 1.67 clusters per day on overage. The overall average number of clusters visited per day by each participant is 4.42. A participant in each trace also visited 121.6, 125.2, 102.8, 36.7, and 30.7 waypoints per day, respectively. The daily traces have duration of 12 h on average. A daily trace of 12 h worth includes approximately 120–150 waypoints on average. SLAW reflects these tendencies as follows. Each walker chooses three to five clusters randomly from C with probability linearly proportional to the weights assigned to clusters. Let C_k be the set of the selected clusters. From C_k, each walker chooses about 120–150 waypoints randomly per day. The speed at which a walker moves from one waypoint to the next waypoint is determined by a speed model discussed in [1]. Two different walkers are allowed to have overlapped waypoints. Let W_k be the set of waypoints that a walker k has selected from S. It also picks a starting waypoint (e.g., home) from W_k from which it always starts its daily trip.

To add some randomness in his daily travel, a walker k replaces one of the clusters in C_k as follows. For each daily trip, it first chooses one new cluster c^+ randomly (ignoring weights) not in C_k and selects waypoints W_k^+ randomly from c^+ (about 5%–10% of all waypoints in c^+). Then, it randomly selects a cluster $c^- \in C_k$ and finds all waypoints $W_k^- \in W_k$ associated with that cluster. At the beginning of each day, walker k starts from its starting point and, throughout the day, makes a one-round trip visiting all waypoints in $W_k \setminus W_k^- \cup W_k^+$ using LATP. It uses a truncated power-law pause-time distribution observed in our traces [1] to decide the amount of time to stay at each waypoint. At the end of the day, it comes back to its starting point. The number of chosen clusters and waypoints to visit and the 3 mean of the pause-time distribution are adjusted so that the whole trip will end within a period of 12 h.

Since each walker k always makes daily trips over a fixed set of C_k, its area of mobility is bounded. Also, since walkers pick their sets of mobility randomly, they tend to have different areas of mobility. In addition, walkers are allowed to deviate from these waypoints by picking new waypoints additionally from the other clusters not in C_k. This allows walkers without any overlapping clusters to occasionally meet, thus having some long ICTs. Those with overlapping clusters may have regular periodic contacts, depending on the transmission ranges or the time they arrive to the clusters.

We apply the cluster weights when selecting C_k to build some sense of community among all walkers. This is a heuristic based on the intuition that bigger clusters are more likely to be visited. The probability to choose a cluster for visit in a day is determined by the size of the weight. Because of self-similar waypoints, some clusters are very large, so many walkers are likely to visit them. These clusters are emulating the common popular gathering places for all participants such as a student union, dormitory, shopping center, street malls, or classrooms. In Section V, we verify that SLAW with this individual walker model produces power-law ICTs [4] and the heavy-tail flights observed from our traces.

V. Performance Evaluation

A. Simulation Setup

For validation, we run mobility simulations using various mobility models. We fix the simulation areas to be approximately the same as the measurement sites in [46]. The transmission range of each node is varied from 25 to 150 m. If not explicitly stated, it is set to 50 m. Fifty nodes are simulated for 200 h, and the first 50 h of simulation results are discarded to avoid transient effects. The speed of every user is set to 1 m/s for simplicity. We use a truncated Pareto distribution as the pause-time distribution for which the minimum and maximum values are 30 s and 700 min, respectively.

For simulation of various models, we use the following setup. For the parameters that are common to all models (e.g., the area of simulation), we use the same value for all models. In case a model requires use of real trace information (e.g., Darthmouth requires to use hotspot information and transition probabilities), we use the ones extracted from our traces. Otherwise, we fix the values of the input parameters presented in the original papers. In the original Dartmouth model [5], a 2-D Gaussian distribution is applied to each pause point, and the hotspots are defined as regions that are higher than a given threshold after summing up 2-D Gaussian distributions. These hotspot regions are very similar with the clusters formed by SLAW. In our evaluations, we apply our clustering method to the Dartmouth model. It uses the same waypoints extracted from real traces [46] to build hotspots and also uses the transition probability obtained from the same traces. Note that since this model requires using the above information, the simulation involving the Dartmouth model uses the same waypoint map that we obtained from the real traces in [46]. In the CMM model [11], the level of preferential attachment depends on the parameters such as the number of nodes and the clustering exponent. We set the clustering exponent of the biggest hotspot to 0.5 following the original paper [11]. In ORBIT [14], following the size of the simulation area, we vary the size of one side of hubs from 200 to 500 m while fixing the number of hubs. Each user selects the same number of hubs for daily travels as the number of hotspots chosen by individual walkers in SLAW.

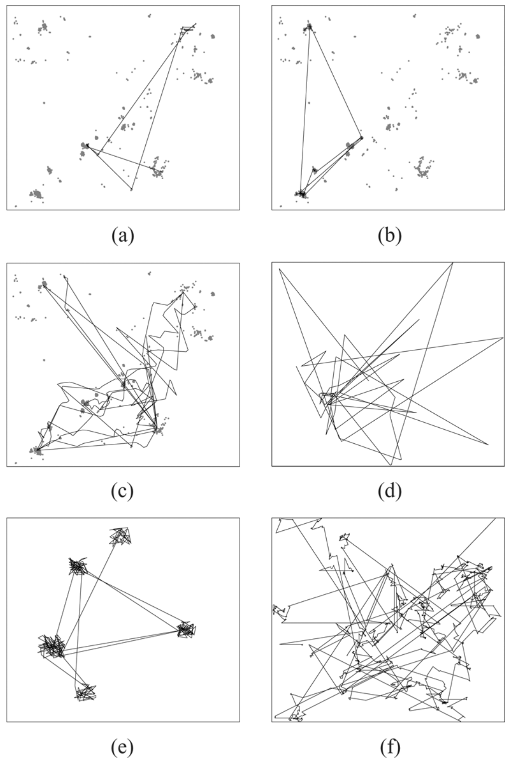


Fig. 14. Sample walk traces of various models. (a) Real GPS traces (Campus I). (b) SLAW. (c) Dartmouth. (c) CMM. (d) ORBIT. (e) TLW.

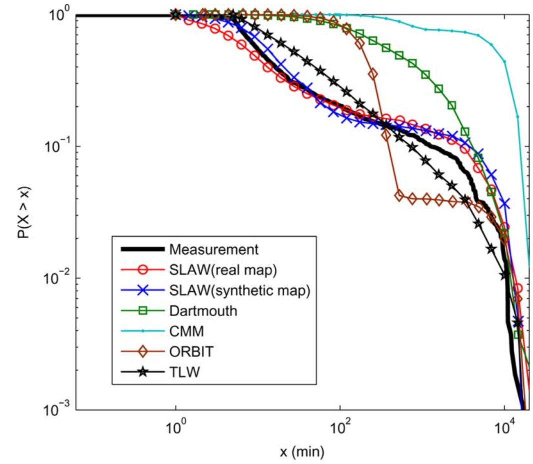


Fig. 15. Flight-length distributions of synthetic traces from various models. "Measurement" is the flight distributions from real traces of New York City.

TABLE IV

RESULT OF THE AKAIKE TEST FOR THE MAXIMUM LIKELIHOOD ESTIMATION OF TRUNCATED PARETO DISTRIBUTIONS (DENOTED Par) AND EXPONENTIAL DISTRIBUTION (DENOTED Exp) OVER FLIGHTS (DENOTED FL) AND ICTs EXTRACTED FROM SYNTHETICALLY GENERATED TRACES FROM VARIOUS MODELS WHOSE PARAMETERS ARE SET BASED ON REAL TRACES OBTAINED FROM FOUR DIFFERENT LOCATIONS (CAMPUS I, CAMPUS II, NYC, AND DISNEY WORLD)

	Campus I		Campus II		NYC		Disney World	
	ICT	FL	ICT	FL	ICT	FL	ICT	FL
Measurement	N/A	Par	N/A	Par	N/A	Par	N/A	Par
$\mathrm{SLAW}_{\mathrm{real}}$	Par	Par	Par	Par	Par	Par	Par	Par
$\mathrm{SLAW}_{\mathrm{syn}}$	Par	Par	Par	Par	Par	Par	Par	Par
Dartmouth	Exp	Exp	Exp	Exp	Par	Par	Exp	Exp
CMM	Exp	Exp	Exp	Exp	Exp	Par	Exp	Exp
ORBIT	Par	Exp	Par	Exp	Par	Exp	Par	Exp
TLW	Par	Par	Par	Par	Par	Par	Par	Par

TABLE V

RESULT OF KULLBACK–LEIBLER (KL) DIVERGENCE: EACH VALUE REPRESENTS THE KL DIVERGENCE VALUE BETWEEN THE FLIGHT-LENGTH DISTRIBUTION FROM THE REAL TRACE AND SYNTHETIC ONES FROM EACH MODEL SHOWN IN FIG. 14

	Campus I	Campus II	NYC	DW
$\mathrm{SLAW}_{\mathrm{real}}$	0.0262	0.0501	0.0125	0.0900
$\mathrm{SLAW}_{\mathrm{syn}}$	0.1182	0.0245	0.0099	0.0284
Dartmouth	0.4095	0.1888	0.1523	0.1477
CMM	0.6300	0.6134	0.3393	0.4980
ORBIT	0.3404	0.2226	0.1742	0.1592
TLW	0.0670	0.0147	0.0249	0.0455

In DTN routing protocol simulations, we generate one message bundle between each of randomly selected 100 source-and-destination pairs. All transmissions are assumed to be reliable and instantaneous when the communicating nodes are within a transmission range. To maximize the effect that mobility models have on routing performance, we assume that all nodes keep the entire history of past contacts with other nodes. All results are averaged over 40 runs.

B. Experimental Validation of SLAW

Fig. 14 shows the sample traces of various mobility models that emulate the mobility in Campus I. It is clearly visible that SLAW generates traces similar to real GPS traces. Dartmouth and SLAW use the same waypoint map extracted from the Campus-I traces. In the Dartmouth model [5], walkers visit every cluster with nonzero transition probability. In ORBIT [14], each user travels a fixed set of clusters (i.e., hubs) daily in a random order with a uniform probability. TLW [1] is a random model, so it does not have common clusters for users. CMM [11] uses one popular cluster using preferential attachment, but it also makes users visit every place in a given area.

We now verify how well SLAW models the statistical features of human mobility. Fig. 15 shows the flight distributions from various models in the experiments from New York City traces. SLAW is also performed on the waypoint maps generated synthetically by our self-similar point generation technique. For the synthetic waypoint map, it uses the average β-value extracted from the real traces. Surprisingly, for both synthetic and real waypoint map inputs, SLAW produces very closely matching flight distributions to and from the GPS traces. The Akaike test [47] tells whether the generated flight distributions fit power-law distributions (e.g., Pareto) or exponential distributions. Table IV shows the result of the Akaike test [47] between Pareto and exponential distributions. In all cases, the flight distributions generated by SLAW are closer to a truncated Pareto distribution than an exponential distribution. We perform the Kullback–Leibler divergence test [48] to measure the closeness of the flight distributions generated from various mobility models to the flight distributions from real traces. The result is reported in Table V, where it shows that SLAW and TLW in general produce the most closely matching flight distributions to real ones. Note that TLW does not have self-similar waypoints, while SLAW does.

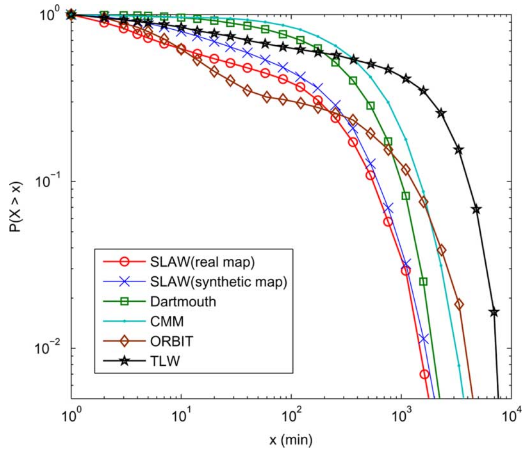


Fig. 16. ICT distributions of synthetic traces from various models running in the Campus-I environment.

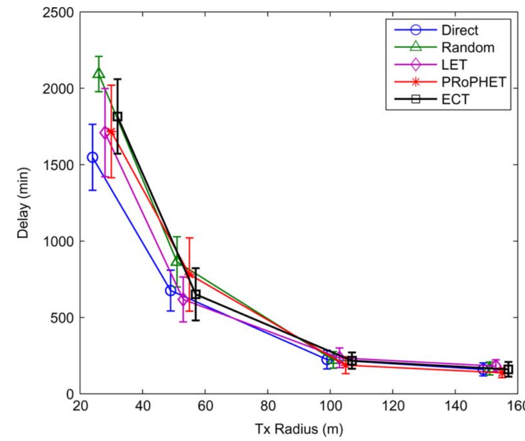


Fig. 17. Average routing delays of various protocols under CMM.

In the same test environments as the above, we run 50 nodes simultaneously to obtain ICT distributions. It is reported in [4] that human mobility induces a truncated power-law ICT distribution. Fig. 16 shows the resulting ICT distributions for various models. Since we do not have any ICT traces corresponding to our GPS traces, we cannot verify the realism of these ICT distributions. However, we can verify whether the ICT distributions follow a truncated power-law pattern. Table IV shows the result of the Akaike test on the ICT distributions. It shows that the ICTs of SLAW, TLW, and ORBIT fit better to power-law distributions, while the ICTs of the other models fit better to exponential distributions. The ICTs of ORBIT shows significantly higher occurrences of very long ICTs. This is because, in ORBIT, the mobile nodes (e.g., humans) with nonoverlapping orbits do not meet at all, while the others may also meet rarely because of randomness in picking the waypoints within their own orbits. The ICTs of CMM and Dartmouth have exponential distributions in most cases and also tend to have much more occurrences of long ICTs than SLAW (note that the scales are logarithmic). This is because they choose the next clusters (or hotspots) to visit randomly without much periodicity so that the chances of two nodes meeting again after the first meeting are much lower.

C. DTN Routing Performance

We test the following five DTN routing protocols: Random forwarding [21], Direct transmission [21], PRoPHET [22], Last Encounter Time (LET) [49], and Expected Contact Time (ECT). Note that all routing protocols are tested with the mobility traces generated from different mobility models. The GPS traces from [1] are individual traces that are not recorded at the same time. Thus, it is infeasible to test routing protocols over the traces.

In LET, a forwarding node of a message picks, as a next relay, the node with the most recent history of meeting the destination of the message among its current neighboring nodes. Each node updates its neighbor set at every minute. ECT is a new metric we developed. It computes the expected time that a node meets the destination by subtracting the last encounter time from the *expected intercontact time*, which is computed by averaging the past intercontact times with the destination.

We can categorize these protocols as *stateless* and *stateful* protocols. Random forwarding and direct transmission are stateless as they do not use any past meeting history information. The other protocols are stateful as they all use past contact information to predict the future probability of meeting the destination.

We find that the routing performance on TLW, RWP, and CMM has almost the same patterns, although their average delays are different. For simplicity, we only show the result of CMM out of them in Fig. 17. In these models, both stateless and stateful protocols perform almost the same. This pattern happens because the mobility of nodes in these models is highly random, so the prediction of stateful protocols is not effective. The lack of performance differentiation among various types of protocols limits the usefulness of these models for mobile network simulation.

Dartmouth [Fig. 18(a)] shows results similar to CMM as the performance of various protocols, except LET is not very distinguishable. The pattern can be explained as follows. The probability that Dartmouth nodes jump to any other hotspots is determined by the transition probability; so any node can jump to any other hotspots as long as the transition probability is nonzero. This causes some randomness in visiting hotspots and, likewise, ICT patterns similar to those in random mobility models such as CMM that are manifested in its exponentially distributed ICTs. However, because of higher transition probability to visit bigger hotspots, Darmouth exhibits much shorter delays than CMM and TLW. The low performance of LET in Darthmouth is because Dartmouth represents a hotspot as a single waypoint (i.e., a cluster reduces to one point), which has a side-effect of increasing pause-times at a hotspot (since all the pause-times for a hotspot are aggregated for all the points inside a hotspot). Thus, when a forwarding node meets a new node with a shorter LET than its LET, it is likely that the new node has just arrived to that hotspot. Thus, the new node is more likely to stay in that hotspot much longer than the forwarding node, thus causing a longer delay to meet the destination next time. These features are an artifact of rather unusual and unrealistic setups of hotspots.

ORBIT [Fig. 18(b)] shows a clear performance differentiation among different protocols. In ORBIT, nodes in nonoverlapping

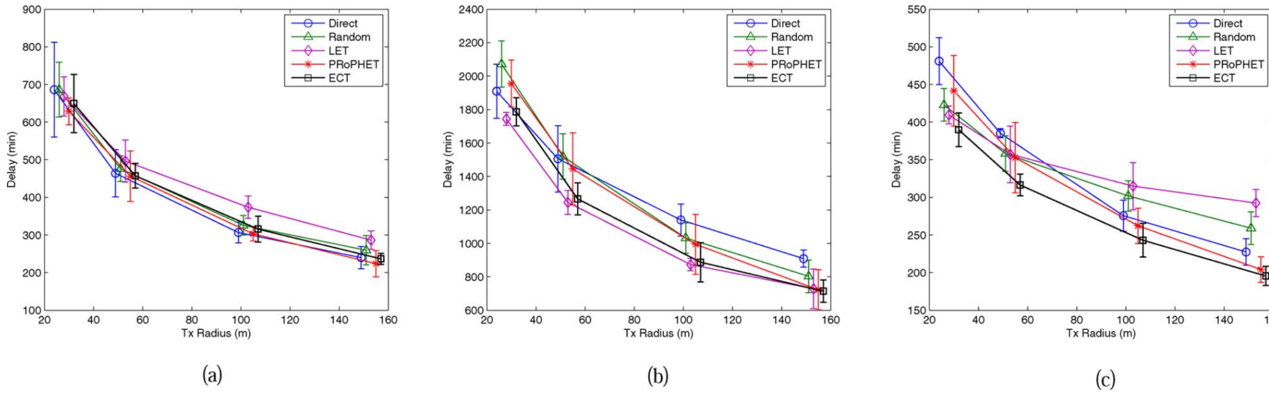


Fig. 18. Average routing delays of various protocols under (a) Dartmouth, (b) ORBIT, and (c) SLAW.

"orbits" (i.e., they do not share a common hotspot) do not meet at all. The only way to deliver messages between two nonoverlapping orbits is through the other nodes with overlapping orbits with the destination. This means that direct and random forwarding can perform really poorly. On the other hand, in ORBIT, stateful protocols can perform much better. The performance difference between ECT and LET are relatively small compared to the difference between random forwarding and LET. This is because in ORBIT, the nodes with long LETs are likely to meet the destination fairly rarely due to only a small overlap in their orbits. Since each node in ORBIT moves like RWP among hotspots in its orbit, a small overlap results in a very long ICT and is thus likely to have long LETs. Thus, the nodes with shorter LETs are likely to have more overlapping orbits with the orbit of the destination. Thus, choosing these nodes as relays leads to short routing delays. A similar argument is applicable for ECT because those with long LETs are likely to have long ECTs in ORBIT.

The mobility patterns of SLAW are quite different from those of ORBIT. The most salient feature is that SLAW has much shorter routing delays. SLAW has much more regularity in their trip patterns than any other models because it uses LATP for the selected set of waypoints and each node visits almost the same set of waypoints every day. In this type of scenario, relaying to a node with short LETs can be detrimental because those nodes that just met the destination are likely to have long ECTs. That means that choosing as relays those nodes with short ECTs would always result in shorter routing delays because expected ICTs are very accurate because of the regularity in trip. The performance of PRoPHET is not as good as ECT because PRoPHET updates its probability only after meeting a destination. Thus, its behavior is slightly similar to that of LET. In SLAW, the power-law ICT distributions play a significant role. Due to the inspection paradox property of the renewal process and the power-law ICTs, when a node meets another node, it is more likely to meet a node with a long ICT. If it is with short ICTs, then ECT would perform as well as LET. However, with long (predictable) ICTs, those nodes with short LETs are not likely to meet the destination for a long time. Thus, ECT can perform better than LET. The fact that ECT performs best indicates that the regularity of trip patterns is well represented in SLAW without loss of inherent statistical features such as power-law flight and ICT distributions.

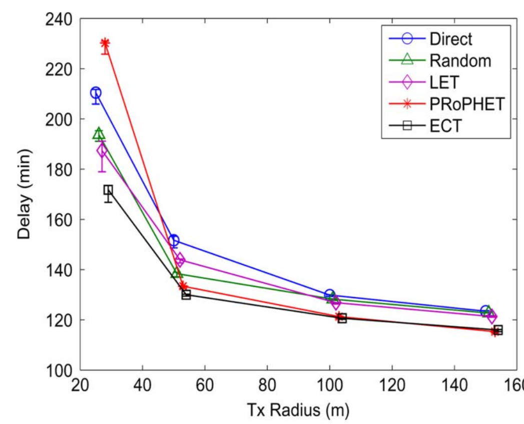


Fig. 19. Average routing delays of various protocols on a real GPS trace [50].

To confirm how much SLAW is realistic in simulating the performance of routing protocols, a measurement data set that records mobility patterns of participants in the same region during the same time period is required. Unfortunately, not only the GPS data set we used throughout the paper, but also most of GPS traces publicly available, have only few number of participants in a region during the same time period. Therefore, for the verification of SLAW, we use our recently published GPS traces [50], which had recorded the movement patterns of 97 students in a university campus (Campus II in Section III) through the same week. This is a unique GPS trace enabling the routing performance evaluation on realistic human mobility patterns. Fig. 19 shows the performance of the routing protocols with the same setting used for the evaluation of mobility models. ECT again performs better than LET or PRoPHET as it is expected and explained in the evaluation through SLAW. This well demonstrates the realism of SLAW.

VI. Conclusion

In this paper, we present a new mobility model, called SLAW,[2] that captures the statistical features found in real human mobility traces. We report many pieces of both analytical and empirical evidence that the movement of people

[2]SLAW and TLW are implemented in network simulators (e.g., MATLAB, NS-2, and GlomoSim). The simulation codes are publicly available through our Web page, http://research.csc.ncsu.edu/netsrv/?q=content/human-mobility-models-download-tlw-slaw.

can be expressed very well using spatial gaps among fractal waypoints and present confirming data for the use of the least-action principle in human trip planning. Based on this, we develop a simple heuristic algorithm called LATP that generates heavy-tail flights on top of fractal waypoints. Combining with heterogeneously bounded walkabout areas, we can successfully reproduce many statistical features important to the study of mobile network performance, especially truncated power-law ICTs. Our routing performance study indicates that SLAW effectively expresses mobility patterns arising from people with some common interests or within a single community like students in the same university campus or people in theme parks where people tend to share common gathering places. We find that LATP over heterogeneously bounded areas realistically expresses some periodicity in the daily mobility of humans. This feature makes many stateful routing protocols such as utility-based routing very effective.

A. Applications of SLAW

Our work is the first in explaining the causes of heavy-tail flight distributions in human mobility. It shows that humans move with contexts (e.g., home, work, meeting, gathering, and favorite places) and explains how these contexts can influence the way that humans make trips. Thus, the insight that our work provides is not limited to computer networking. Realistic human mobility models have applications in many diverse disciplines outside computer networks including civil engineering for city and escape planning, disease control for studying virus outbreak spread, telecommunication for planning cell-phone towers and understanding handoff patterns, and sociology for studying the interaction and social network patterns of humans.

B. Future Work

We believe that SLAW is a few steps closer toward representing realistic synthetic mobility patterns than the existing work. However, it is clear that SLAW may not necessarily capture all the important statistical features of human mobility. There could be many other features that SLAW may not capture. One of them is temporal features of mobility. People meet because they are in the same place and also at the same "time," and also people move to a certain location at a certain time because they have to be there at that time. A hotspot may become popular only at a certain time, e.g., a restaurant area. Real mobility patterns must be a result of representing these spatial and temporal correlations. In the construction of mobility traces, SLAW chooses a set of destinations to visit randomly from a given set of hotspots, and the visits to a certain waypoint are determined by a function of distance from the current waypoint. While this decision process may capture some realism in spatial constraints, this does not represent temporal tendencies of human mobility. We leave this problem as an open problem.

REFERENCES

[1] 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.

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

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

[4] 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, DOI: 10.1109/INFOCOM.2006.172.

[5] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006, DOI: 10/1109/INFOCOM.2006.173.

[6] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović, "Power law and exponential decay of inter contact times between mobile devices," in *Proc. ACM MobiCom*, 2007, pp. 183–194.

[7] S. Hong, I. Rhee, S. J. Kim, K. Lee, and S. Chong, "Routing performance analysis of human-driven delay tolerant networks using the truncated Levy walk model," in *Proc. 1st ACM SIGMOBILE Workshop Mobility Models*, 2008, pp. 25–32.

[8] B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York: Freeman, 1983.

[9] N. Chakravarti, "Beyond Brownian motion: A Levy flight in magic boots," *Resonance*, vol. 9, no. 1, pp. 50–60, Jan. 2004.

[10] V. Borrel, M. D. de Amorim, and S. Fdida, "A preferential attachment gathering mobility model," *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 900–902, Oct. 2005.

[11] S. Lim, C. Yu, and C. R. Das, "Clustered mobility model for scale-free wireless networks," in *Proc. IEEE LCN*, Tampa, FL, Nov. 2006, pp. 231–238.

[12] E. Hyytia, P. Lassila, and J. Virtamo, "A Markovian waypoint mobility model with application to hotspot modeling," in *Proc. IEEE ICC*, Istanbul, Turkey, Jun. 2006, pp. 979–986.

[13] X. Hong, M. Gerla, G. Pei, and C. Chiang, "A group mobility model for ad hoc wireless networks," in *Proc. ACM MSWiM*, Seattle, WA, Aug. 1999, pp. 53–60.

[14] J. Ghosh, S. J. Philip, and C. Qiao, "Sociological orbit aware location approximation and routing (solar) in MANET," *Ad Hoc Netw.*, vol. 5, pp. 189–209, 2007.

[15] 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.

[16] I. Rhee, K. Lee, S. Hong, S. J. Kim, and S. Chong, "Demystifying the Levy-walk nature of human walks," CS Dept., NCSU, Raleigh, NC, Tech. Rep., 2008 [Online]. Available: http://netsrv.csc.ncsu.edu/export/Demystifying_Levy_Walk_Patterns.pdf

[17] H. Hurst, R. Black, and Y. Simaika, *Long-Term Storage: An Experimental Study*. London, U.K.: Constable, 1965.

[18] M. Kim and D. Kotz, "Periodic properties of user mobility and access-point popularity," *Pers. Ubiquitous Comput.*, vol. 11, no. 6, pp. 465–479, Aug. 2007.

[19] F. Ekman, A. Keränen, J. Karvo, and J. Ott, "Working day movement model," in *Proc. 1st ACM SIGMOBILE Workshop Mobility Models*, 2008, pp. 33–40.

[20] W. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *Proc. IEEE INFOCOM*, Anchorage, AL, May 2007, pp. 758–766.

[21] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The single-copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 63–76, Feb. 2008.

[22] A. Lindgren, A. Doria, and O. Schelen, "Probabilistic routing in intermittently connected networks," *Mobile Comput. Commun. Rev.*, vol. 7, no. 3, pp. 19–20, Jul. 2003.

[23] 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.

[24] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for multidimensional PCS networks," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 718–732, Oct. 2003.

[25] F. Bai, N. Sadagopan, and A. Helmy, "Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for ad hoc networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, Apr. 2003, vol. 2, pp. 825–835.

[26] A. Jardosh, E. Belding-Royer, K. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," in *Proc. ACM MobiCom*, San Diego, CA, Sep. 2003, pp. 217–229.

[27] A.-L. Barabasi and E. Bonabeau, "Scale-free networks," *Sci. Amer.*, vol. 288, pp. 50–59, 2003.

[28] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.

[29] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *Mobile Comput. Commun. Rev.*, vol. 11, pp. 59–70, Jul. 2007.

[30] C. Boldrini, M. Conti, and A. Passarella, "Users mobility models for opportunistic networks: The role of physical locations," in *Proc. IEEE WRECOM*, 2007, pp. 1–6.

[31] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling spatial and temporal dependencies of user mobility in wireless mobile networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1564–1577, Oct. 2009.

[32] "CRAWDAD: A community resource for archiving wireless data at Dartmouth," Dartmouth College, Hanover, NH [Online]. Available: http://crawdad.cs.dartmouth.edu/

[33] M. McNett and G. M. Voelker, "Access and mobility of wireless PDA users," *Mobile Comput. Commun. Rev.*, vol. 9, no. 2, pp. 40–55, 2005.

[34] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proc. ACM MobiCom*, 2004, pp. 187–201.

[35] J. Beran, R. Sherman, M. Taqqu, and W. Willinger, "Long-range dependence in variable-bit rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 2, pp. 1566–1579, Feb. 1995.

[36] M. S. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for long-range dependence: An empirical study," *Fractals*, vol. 3, pp. 785–798, 1995.

[37] J. Banerjee, M. K. Verma, S. Manna, and S. Ghosh, "Self-organised criticality and 1/f noise in single-channel current of voltage-dependent anion channel," *Europhys. Lett.*, vol. 73, no. 3, pp. 457–463, Feb. 2006.

[38] J. Gordon, "Pareto process as a model of self-similar packet traffic," in *Proc. IEEE GLOBECOM*, Singapore, Nov. 1995, vol. 3, pp. 2232–2236.

[39] D. R. Cox, "Long-range dependence: A review," in *Statistics: An Appraisal*. Ames, IA: Iowa State Univ. Press, 1984, pp. 55–74.

[40] H. G. Schuster, *Deterministic Chaos*. New York: VCH, 1989.

[41] J. Gaite, "Zipf's law for fractal voids and a new void-finder," *Eur. Phys. J. B*, vol. 47, no. 11, pp. 93–98, 2005.

[42] P. L. M. de Maupertuis, "Accord des differentes lois de la nature qui avaient jusqu'ici paru incompatibles," *Memoires de l'Academie des Sciences*, p. 417, 1744.

[43] D. Helbing, J. Keltsch, and P. Molnár, "Active walker model for the formation of human and animal trail systems," *Nature*, vol. 388, pp. 47–50, 1997.

[44] I. Norros, P. Mannersalo, and J. L. Wang, "Simulation of fractional Brownian motion with conditionalized random midpoint displacement," *Adv. Perform. Anal.*, vol. 2, pp. 77–101, 1999.

[45] D. Koutsoyiannis, "The Hurst phenomenon and fractional Gaussian noise made easy," *Hydrol. Sci. J.*, vol. 47, no. 4, pp. 573–595, Aug. 2002.

[46] "NCSU-KAIST GPS data set," 2008 [Online]. Available: http://research.csc.ncsu.edu/netsrv/?q=content/mobility-models-trace-data

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

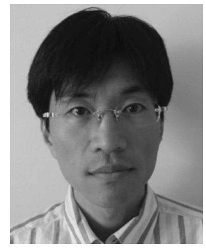
[48] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.

[49] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli, "Age matters: Efficient route discovery in mobile ad hoc networks using encounter ages," in *Proc. ACM MobiHoc*, Annapolis, MD, Jun. 2003, pp. 257–266.

[50] S. Hong, K. Lee, and I. Rhee, "STEP: A spatio-temporal mobility model for human walks," in *Proc. IEEE MASS*, 2010, pp. 630–635.

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

He is currently a Senior Research Scholar 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 services and applications, mobile system, and protocol design.



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

From 1998 to 2005, he was a Member of Technical Staff with the Telecommunications and Operations Support Laboratory, Korea Telecom (KT), Seongnam, Korea. He is currently a Principal Researcher with the Samsung Advanced Institute of Technology (SAIT), Samsung Electronics, Yongin, Korea. His research interests include human mobility modeling and routing algorithms in various wireless 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), Daejon, Korea, in 1998, 2000, and 2007, respectively.

He is currently a Member of Technical Staff with the DMC Research Center, Samsung Electronics, Suwon, Korea. 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.

Injong Rhee (S'89–M'94) received the Ph.D. degree in computer science from the University of North Carolina at Chapel Hill.

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

Prof. Rhee is an Editor of the IEEE TRANSACTIONS ON MOBILE COMPUTING.

Song Chong (M'93) 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 Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, where he is a Professor and was the head of the department's Communications and Computing Group. Prior to joining KAIST, he was a Member of Technical Staff with the Performance Analysis Department, AT&T Bell Laboratories, Holmdel, NJ. He has published more than 100 papers in international journals and conferences. His current research interests include wireless networks, future Internet, and human mobility characterization and its application to mobile networking.

Prof. Chong is currently the Chair of the Wireless Working Group of the Future Internet Forum of Korea and Vice President of the Information and Communication Society of Korea. He is an Editor of *Computer Communications* and the *Journal of Communications and Networks*. He has served on the Technical Program Committee of a number of leading international conferences including IEEE INFOCOM and ACM CoNEXT. He serves on the Steering Committee of WiOpt and was the General Chair of WiOpt 2009.