Advances in Complex Systems

Understanding the Correlation of the Properties of Human Movement Patterns

Rajib Ranjan Maiti,* Arun Mallya
Animesh Mukherjee, Niloy Ganguly

Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
Kharagpur 721302, India
{*rajib.maiti,arun.mallya*}*@gmail.com,*
{*animeshm,niloy*}*@cse.iitkgp.ernet.in*

Received (received date)
Revised (revised date)

Mobility of humans plays an important role in shaping the performance of various mobile and location dependent networks. In this paper, we amalgamate and summarize the parameters reported in various trace analyses, and explore the inter-dependencies among the parameters using a layered approach. We then identify using an experimental set-up (a) several roadblocks in generating an intended synthesis model, (b) the changes in the mobility patterns with variation in inputs (e.g., the underlying map), and (c) the changes in efficiency of the targeted service (here we have considered simple broadcasting) with change in the input parameters. Interestingly, we notice that the efficiency of the service does not necessarily depend on the mobility pattern and one needs to be extremely careful before drawing a direct correlation between the two.

Keywords: Mobility Models; Human Mobility Patterns; Power Law.

1. Introduction

Mobility has become an integral part of a large class of networks such as location based social networks, delay tolerant networks, mobile P2P networks. Mobility defines the underlying dynamics of these networks, and the performance of protocols on these networks is largely dependent on the mobility of the agents in it. To faithfully reproduce the mobility of agents, research has been directed towards (a). building mobility models that can accurately synthesize realistic settings and (b). collection of a large number of real human movement traces which is then used to build better and perfect models. In order to ensure seamless abstraction of model from traces, a framework which would (a). segregate properties of a mobility trace and independently try to demonstrate the importance of each of these in designing a model meant for a specific task, (b). consider the properties that are missing in

*Corresponding author

the available trace data but are required to fulfill the targeted analysis, and finally, (c). point out a target set of properties that should be measured while constructing new real traces in the context of a particular task, would be very helpful and crucial as well. Such a framework can then be used to decide the most important properties which need to be considered while developing a *simple and comprehensible* mobility model to accurately mimic a set of mobility traces.

In order to achieve this, the paper summarizes from the existing methods for mobility analysis and synthesis, and reports a large set of important properties of human mobility patterns. Next, to understand the relationships among these properties, a layered dependency relation (Figure 2) among the various parameters is developed where a layer may correspond to a set of inputs (eg. map) or outputs (eg. flight length of an agent) or task-dependent outputs (eg. coverage). As an instance of task, we consider the simple broadcasting protocol run in delay tolerant networks (DTNs[a]). We investigate the intra- and inter-layer dependencies through a generic experimental set-up similar to [13]. We then investigate the relationship between the change in output of the mobility patterns and the corresponding change in performance of epidemic broadcasting protocol in DTNs.

Thus the study presented in this paper unfolds the intrinsic and cryptic dependencies (that are known to exist but not clear how) among the properties of human movement patterns. It attempts to derive insights from the recently proposed complex models and increase the advantages of using them. Our major contributions in this study are as follows.

(1) A layered dependency relation among a large set of properties of human movement is developed in order to understand the relationships among them.
(2) A generic framework of mobility models is proposed to investigate the intra- and inter-layer dependencies.
(3) Our analysis shows that the distribution of the values of a parameter belonging to the input layer can change an intended distribution of values of another parameter belonging to the input layer itself.
(4) The task-oriented output parameters (i.e., the coverage and the spreading rate) are highly sensitive to some of the input properties (for example, length of the map and area of roaming). However, the output properties (for example, flight length and inter-contact time) are in general not much affected by the input properties. For example, the distribution of area of roaming hardly affects the distribution of inter-contact times, but it has a direct impact on the coverage of broadcasting.

[a]A Delay Tolerant Network (DTN) [5] consists of a set of agents, each agent is capable of sending and receiving messages from each other, moving in a particular geographic location. Unlike traditional communication network, a DTN does not assume any pre-established infrastructure for communication among the agents in the network. A pair of agents in this network can communicate only when they are physically close enough as a result of their movement, and these local interactions lead to eventual end-to-end delivery of messages.

This is particularly significant as existing studies, that analyze the performance of mobility driven applications, have considered the mobility traces (no matter how they get produced, e.g. which model and which data set) as input and the performance of the application as output - this layered dependency approach unfolds the pitfall in doing so.

(5) Our analysis on the real data sets shows that the variation of coverage (a property in task-oriented output layer) may not be explained by the variation of inter-contact times (a property in output layer). However, it can be clearly explained by the variation of radio range (a property in input layer). Thus, our proposed layered dependency structure and the framework can help to correlate, easily and efficiently, the properties across different layers with the performance of a given mobility driven application.

The rest of this paper is organized as follows. Section 2 briefly describes the existing works that have attempted to investigate the relationships among the properties. Section 3 defines the terminologies that are used in this work. Section 4 summarizes the properties identified through realistic trace analysis and categorizes them into layers. Section 5 describes our proposed framework and analyzes the results of this study. The study is concluded in Section 7.

2. Related Works

A large number of mobility models [2, 4, 9, 14] exists in the literature; however, relatively a much less studied area comprises the investigation of the relationships among the properties of human movement patterns. Broadly, the studies in this area can be classified into two types depending on the type of relationships considered. Some studies [3, 10, 18, 19] have investigated the impact of a particular parameter on the performance of a targeted application. For example, the study in [3] has analyzed the impact of power law distributed inter-contact times on the performance of an opportunistic forwarding protocol, and the study in [10] has investigated the impact of visiting a limited number of locations, selected randomly, by each individual on the distribution of inter-contact times. Some studies [8, 15] have shown that the procedure to measure a set of values of a particular parameter can significantly affect the distribution of that parameter. For example, the distribution of inter-contact times can be computed either for every individual pair of agents, or it can be computed by aggregating the inter-contact times from all the pairs in the network, and the resulting distributions need not be always similar [15].

Thus, though existing studies have shown that there exists a significant amount of dependency among the properties, even in the measurement techniques, there has been no study to provide a complete view of inter-dependencies among the properties of human mobility patterns. To the best of our knowledge, this is the first study that considers a large number of properties of human movement, observed recently from available trace data, and investigates the relationships among them through the introduction of a generic mobility framework organized in the form of

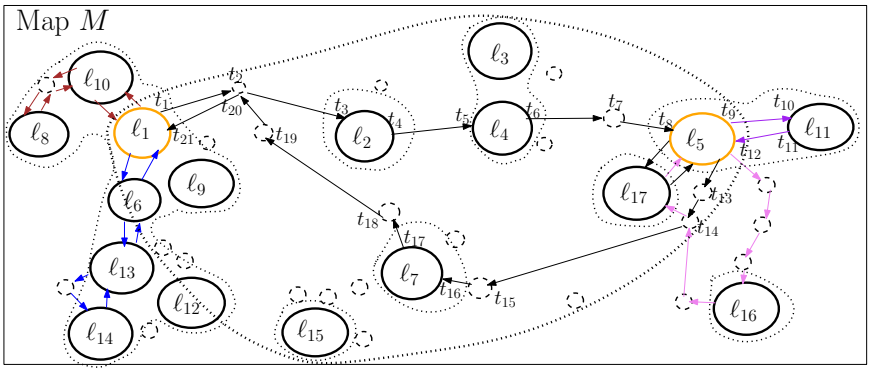

Fig. 1: Trace of four agents (shown by different colored paths) in a time window of length T. A path indicated by directed arrows of one single color indicates a trace of an agent.

a layered dependency structure.

3. Terminologies and Definitions

Before we describe various concepts related to the properties of mobility behavior, let us define some terminologies that are used in this study. Figure 1 shows a sample trajectory of four moving agents in a time window of length T. Let the trajectory indicated by black arrows be the trajectory of an agent u, denoted by $\mathcal{T}(u)$. A circle (either having solid or dashed lines) in the trajectory indicates a traceable location δ_i w.r.t. time t_j.

Waypoint: The agent u can be found at a location, say δ_k, for more than one consecutive time step, which is termed as a waypoint (WP) ℓ, denoted by ℓ_x in $\mathcal{T}(u)$. Thus, the waypoints included in $\mathcal{T}(u)$ are $\ell_1, \ell_2, \ell_4, \ell_5, \ell_{17}, \ell_5, \ell_7$ and ℓ_1.

Waiting time: The period for which u is seen at a WP ℓ_x is termed as waiting time [20] (ω), e.g. $\omega = t_4 - t_3$ at ℓ_2.

Flight length: The geographical distance between two consecutive WPs in $\mathcal{T}(u)$ is termed as flight length (ι), e.g. $\iota = |\ell_1 - \ell_2|$.

Area of gyration and ***radius of gyration***: The smallest circular area that encloses most of the WPs in $\mathcal{T}(u)$ and the radius of that area are termed as area of gyration (π) [6] and radius of gyration (η) respectively.

Return time and ***return-time probability***: The time-gap between two consecutive repetitions is termed as return time (φ) and the probability of repetition after that time-gap to any WP is termed as return-time probability, e.g. $\varphi(\ell_1) = t_{21} - t_1$.

Visit frequency and ***return probability***: The number of times a WP ℓ_x is repeated in $\mathcal{T}(u)$ is termed as visit frequency (λ) and the probability of repetition is termed as the return probability (β) [6], e g., $\lambda(\ell_5) = 2, \beta(\ell_5) = 2/8$.

Map: An area (rectangular area in Figure 1) that accumulates WPs of all the agents is termed as a map M.

Hurst parameter: The Hurst parameter of a given configuration is measured in the following way :- the site map is divided by a grid of unit squares and all the

Table 1: Properties reported in existing studies. UC, SG, UD, PL denote Univ. campus, social gathering, uniform distribution and power law with exponent k resp. Hence, a row, say 1^{st} row, denotes that the study in [20] analyzes the movement traces in a UC and reports that $P(\iota)$ and $P(\omega)$ follow uniform distribution and partially follow power law resp.

Ref.	**Site**	**Parameter**	**Probability Distribution**
[20]	UC	$(\iota; \omega)$	(UD; partially follow PL)
[7]	UC	inter-WP and inter-place transition	$P(\kappa_x) = 1 - e^{-(\frac{\kappa_x}{a})^b}$, a, b are constants and κ_x denotes the popularity of x, x being either a WP or a place
[11]	UC	(ω)	(UD) - for stationary agents
[11]	UC, ν and θ denote speed and direction of movement resp.	$(\nu; \theta; \omega)$	(log-normal; Uniform in range $[0°, 90°]$ with peaks at $0°$ and $90°$, repeated at every $180°$; log-normal) - for moving agents
[11]	UC	start and end time of travel	(Exponential) - for both stationary and moving agents
[12]	UC	$(\iota; \omega; \tau)$	(PL; PL; PL)
[3]	UC, SG	(τ)	(PL, $0.1 \leq k < 1.0$)
[1]	Bank note sitings in US	(ι)	(PL, $k = 1.59 \pm 0.02$)
[6]	Phone call activity	$(\iota; \eta; \lambda; \varphi)$	(PL with Exponential cut off, $k = 1.75 \pm 0.15$; PL with Exponential cut off, $k = 1.65 \pm 0.15$; zipf's law; periodic peak and decrease over time)

waypoints within each unit square are counted and normalized by the area of the unit square. The variance of these normalized count is then measured. The size of the unit square is progressively increased and the variance is recalculated. The gradient at which the variance deceases with increase in the unit size is denoted by β - Hurst parameter of the configuration is denoted as 1 - β. The study in [12] has shown that the values of Hurst parameter (H) in human generated waypoints can vary in the range $0.55 \leq H \leq 0.95$ in realistic scenarios.

Place and ***popularity of a place***: A place is a subset of WPs in a map M that are clustered together depending upon some parameter(s), e.g. inter-WP distance, e.g $[\ell_3, \ell_4]$ and $[\ell_2]$ are two places, and popularity (κ) of a place may be calculated for each such place depending on some other parameter(s), say, number of WPs.

Inter-contact time and ***contact-duration***: Considering multiple agents moving in a map, the time gap between two successive contacts of a pair of agents and the duration for which they are in contact are termed as inter-contact time (τ) and contact-duration (σ), respectively.

A probability distribution $P(p)$ of a parameter p may also be defined where p can denote any one of $\iota, \omega, \eta, \beta, \kappa, \sigma, \tau$.

Table 2: Parameters (symbols and interpretations) used in this paper.

ID	Symbol	Parameter	Distribution
$T_{start}T_{end}$	-	Start and end time	Constant
$TrPlan$	α, VF	Travel Plan	LATP
Map	M, H	Map	-
F_{len}	ι	Flight Length	Power law
P_{pause}	ω	Pause time	Power law
T_{ret}	φ	Return time	Periodic
R_{gyr}	η	Radius of gyration	Power law
ICT	τ	Inter-contact time	Power law
CD	σ	Contact Duration	Power law

4. Summary of the Properties

While summarizing the relevant literature, it was found that various studies on trace analyses have reported about fifteen different properties (briefly stated in Table 1). A property is a parameter-distribution pair where the values of the parameter are measured from real data and a distribution is fitted to describe the characteristics of the observed values. For example, *flight length* is a parameter, whereas "flight length being power law distributed"- is a property. If the information about a set of closely related properties can be expressed by a single representative property, we group them together and take that single property as the group representative. For example, a flight having length ι_i (say, from a WP ℓ_i to another WP ℓ_j) in the direction of θ_j that takes a time duration ι_t can be represented by an ordered tuple $(\iota_i, \theta_j, \iota_t, \omega_j)$, where ω_j is the pause time at ℓ_j [16]. Thus, given the distance and the time for each of the flights taken, the distribution of both speed ν and direction of movement can easily be derived and hence, the distribution of flight length becomes a representative of a group ($P(\iota), P(\theta)$ and $P(\nu)$) of distributions.

Existing studies have identified a *set of independent input parameters*, which are used to generate mobility trace, such as the distribution of start and end time [11] and the travel plan [13] per day of each human, and a map of WPs. Beyond these inputs, there are a host of other parameters which are taken as input. However, these parameters follow certain distributions. We note down the properties and the distributions they follow from literature. The distribution of flight length $P(\iota)$ follows either a uniform distribution [20] or a power law [12] depending on the scenarios considered, $P(\omega)$ follows a power law [12], $P(\varphi)$ shows a decreasing trend with the duration of time and has periodic peaks [6] (φ represents a group consisting of λ [17], β [13], κ), and $P(\eta)$ follows a power law[6] ($P(\eta)$ represents a group consisting of the distribution of trace length, and agent heterogeneity).

The studies have also identified a *set of protocol independent output parameters* namely inter-contact time (τ) and contact-duration (σ). The distribution of ICTs can be computed in two ways, either the ICTs can be collected from individual pair of agents (pairwise ICT) or they can be aggregated from all possible pairs (aggregated ICTs) in the network. Studies [15] have shown that the distributions of

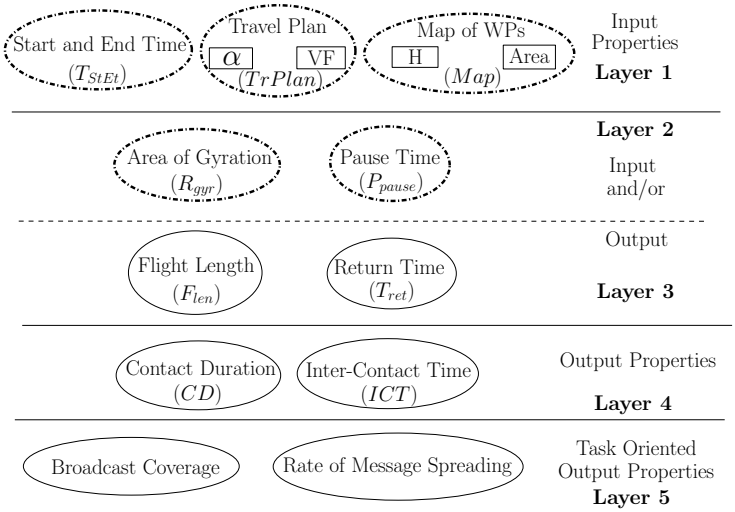

Fig. 2: The layered structure of the properties of human mobility.

pairwise ICTs and aggregated ICTs need not be always similar. Though a consensus on the distribution of both pairwise ICTs and aggregated ICTs is yet to be reached, some studies [3] have reported that the distribution follow a power law, and some other studies [8, 13] have shown that it follows a power law with an exponential cut off beyond a certain threshold of ICTs. It is important to note here that this exponential cutoff in the distribution of ICTs discards the hypothesis of infinite message delivery delay that has been made in earlier studies (as in [3]). For the current experimental purposes we have used aggregated ICTs; however, the framework is suitable to adopt to pairwise ICTs.

Table 2 summarizes the properties where the ID column specifies the property identifiers, *Symbol* column specifies the parameter of the properties, *Distribution* column specifies the distribution of the properties that we consider in this paper. However, note that the layered framework is not limited by such distribution, other distributions can also fit in.

Inter-dependence Analysis: We can thus summarize that depending on the applicability in designing a model, the properties can be categorized into four different sets (Figure 2): a set of input properties consisting of $T_{start}T_{end}$, *TrPlan* and *Map* (*Layer* 1), a set of output properties consisting of ICT and CD (*Layer* 4), and a set of properties that can be input and/or output consisting of F_{len}, P_{pause}, T_{ret} and R_{gyr}. In this paper, P_{pause} and R_{gyr} are considered as inputs (*Layer* 2) and F_{len} and T_{ret} as outputs (*Layer* 3), but this ordering may change. An additional set of properties (*Layer* 5) is added, termed as *Task-Oriented Output Properties*, that specify the metrics that may be used to evaluate the performance of the targeted task. For example, *coverage* and *rate of message spreading* can be two metrics to investigate the performance of a broadcast protocol for DTNs.

The properties exhibited by human movement patterns apparently look to be

inter-related and designing a model capturing even a subset of these properties is a complex and difficult task. Hence, before designing a model, it is necessary to first investigate the relationship among them (properties) and then understand how each of them affects the performance of a protocol. As can be conceived, the space of such investigation is huge, we make an attempt to do so for some of the relationships, through a layered approach (Figure 2). From the input-output dependency logic in designing a model, *Layer* 1 may influence all other layers and possibly *Layer* 2 may determine the nature of the distributions of the properties in the layers below it. A relatively non-trivial dependency in case of designing a human mobility model is that the properties in *Layer* 1 and *Layer* 2 are dependent on each other. We shall investigate various dependencies through extensive simulations and find correlations among the parameters.

5. Implementation of the Layer Dependencies

To investigate the intra- and inter-layer relationships shown in the previous section, we have designed an algorithmic setup detailed below. The algorithm considers $T_{start}T_{end}$, *TrPlan*, *Map*, P_{pause} and R_{gyr} as input properties and generates mobility traces with different characteristics.

5.1. *Algorithm to Generate Mobility Traces*

The algorithm is divided into two sub-procedures: $findWPs$ that finds a set of WPs to travel through, and *preparePlan* that prepares a travel plan to visit the selected WPs. Let us assume an agent u_i whose movement trajectory will be generated by the algorithm and will look like the one indicated by the black arrows in Figure 1. The algorithm takes a map M of WPs as input, which may be generated synthetically, e.g., by using steps in [13], or it can be produced from a real trace. Each WP ℓ_i in M is assigned a weight κ_i depending on its location within M, e.g., popularity which is same as the *place* it belongs to.

findWPs: For each agent u_i, this procedure computes three sets of WPs (R_i, T_i and T_i') and returns two of them (R_i and T_i'). Here R_i refers to a set of regular WPs and T_i' to a set of temporary WPs.

- *Step 1.* (Find home location) Choose a WP ℓ_i in M proportional to ${\kappa_i}^{c_1}$ as home ℓ_i^h of u_i, where κ_i is the weight of ℓ_i and c_1 is a parameter ($c_1 = 0$ indicates random choice, $c_1 = 1$ indicates preferential choice). For example, the WP ℓ_1 indicated by orange circle in Figure 1 is chosen to be the home of u_i.
- *Step 2.* (Find area of gyration) Pick a radius η_i of gyration from a distribution $P(\eta)$ which is an input that can take any form. Assume a circular area of gyration π_i having radius η_i such that ℓ_i^h is located on the boundary of π_i. Let T_i denote the set of all the WPs including ℓ_i^h which are located within π_i. For example, in Figure 1, π_i is the largest circular area indicated by the dotted line that includes the WPs $T_i = \{\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_{17}, \ell_7, \ell_{15}, \ell_6, \ell_9\}$.

- *Step 3.* (Compute R_i set) Put ℓ_i^h in R_i. For each WP in T_i, compute the distance from it to ℓ_i^h. If the difference between this distance and η_i is below a threshold then put this WP in R_i. The number of WPs in R_i is an input parameter. Note that the area may not have suitable WPs to fill the R_i set. In that case, the value of π_i is suitably tuned (increased or decreased; this may actually disturb the intended distribution of η) and the T_i set is recomputed, and this step starts over again. In Figure 1, ℓ_1 and ℓ_5 are in R_i when size of this set is 2.
- *Step 4.* (Compute T_i' set) Randomly select a subset T_i' of WPs of the set $T_i - R_i$. The number of WPs in T_i', say *VF*, is an input parameter. Assuming the size of T_i' as 4, T_i' contains $\ell_2, \ell_4, \ell_{17}$ and ℓ_7 for the agent u_i in Figure 1.

preparePlan: For each agent u_i, given two sets of WPs (R_i and T_i'), this procedure prepares a plan to visit through all the WPs in R_i and T_i' starting from and ending at ℓ_i^h. Let *Trip* denote such a plan that takes a period $\mathbb{T}$ (= 24 hours, every time step correspond to 0.1 hour) of time to complete. Essentially, this procedure creates an ordered set of WPs in $R_i \cup T_i'$ to indicate the WP to be visited next in this present trip.

- *Step 1* (Assign waiting times at WPs) The time period $\mathbb{T}$ is divided into two fractions t_r and t_t ($< t_r$). The waiting times at WPs in R_i and in T_i' are sampled from two distributions (may be same), that are considered as inputs. In our case, the distributions in case of R_i and T_i' are considered as normal distribution in range [0.1,t_r] and a power law in range [0.1,t_t] respectively. Thus, in Figure 1, the waiting times at ℓ_1 and ℓ_5 (resp. at $\ell_2, \ell_4, \ell_{17}$ and ℓ_7) are assigned from a normal distribution (resp. from a power law distribution).
- *Step 2* (Select next WP) Let ℓ_p be the present WP in the current trip. The next WP ℓ_x is selected from a set U_i of WPs, where U_i is a subset of $R_i \cup T_i'$ which are not yet visited in this trip. Note that, initially, the number of WPs in the set U_i is same as in $R_i \cup T_i'$, and it is decreased by one every time an unvisited WP is visited in this trip. The WP ℓ_x is chosen with a probability $P(d) \sim d^{-\alpha}$ (similar to that in [13]), where $0.0 \leq \alpha \leq 5.0$ is an input and d is the Euclidean distance between the present and the next WP. In Figure 1, a sample order of visit to the WPs in $R_i \cup T_i'$ is indicated by black arrows.

Such a trip is created for every $\mathbb{T}$ period of time, say *day*, where the WPs in R_i (say, regular WPs) remain fixed, but the WPs in T_i' (say, temporary WPs) are chosen randomly from $(T_i - R_i)$ before starting the trip each day.

5.2. *Simulation Setup*

We have designed a simulator following the steps stated in subsection 5.1 to generate the mobility traces. In this paper, we consider a squared 2D geometric space as a simulation area (i.e., a map), and a WP inside the map be a (x, y) co-ordinate in the geometric space and hence, the length of a flight is the Euclidean distance

between two successive co-ordinates in a trajectory. A large set of maps is generated following the steps in [13] by considering different values of Hurst parameter (H) with different map area.

In this paper, a map having squared area with side length $L = 2400m$ is generated for each value of $H = 0.55, 0.65, 0.75, 0.85, 0.95$, and a map with $L = 1000m, 2400m, 8500m, 15000m$ for a fixed H ($= 0.75$) with 700, 3300, 4000, 10500 WPs respectively, i.e., the map area is varied from a typical social gathering to a big university campus [12]. The transmission range of each individual agent and the range with which places inside a map are created (by transitively connecting the WPs) are kept the same and fixed to 100m.

The number of regular WPs R_i for any agent u_i is kept as 2 in the simulations performed for generating all the results in this paper to denote a *home* and a *work place* for u_i. The radius of gyration (η) for every u_i is sampled from different distributions to generate results in different sections of this paper, the exact distribution is mentioned in the respective sections. It either follows a uniform distribution (three specific ranges are considered in this case $10 \leq \eta \leq 100$, $10 \leq \eta \leq 500$ and $10 \leq \eta \leq 1000$ denoted as $U(100), U(500)$ and $U(1000)$ respectively) or a power law (in this case, the range of η is considered as $10 \leq \eta \leq L/2$, denoted as *PL(Half)*). In a different setup, we have taken the area of gyration to be the entire simulation space, denoted as $U(Full)$.

Recall that a trip is a journey that, starting from home, travel through all other WPs in $R_i \cup T_i'$, and return back to home. Visiting the WPs in $R_i \cup T_i'$ is done using the travel plan *TrPlan* where the probability to visit a WP ℓ_x after the current WP ℓ_c is proportional to $dist(\ell_c,\ell_x)^{-k}$, where $dist(\ell_c,\ell_x)$ is the Euclidean distance between ℓ_c and ℓ_x, and k is a constant (an input parameter). In each simulation run, 100 agents are considered and all the agents start moving at time step 1 and move for 7200 time steps (correspond to 30 days when a single time step corresponds to 0.1 hour). Unless stated otherwise, $P(\eta)$ follows *PL(Half)*, and a map having area 2400^2m^2 with $H = 0.75$ is considered, and the results shown are an average over 100 simulation runs.

5.3. *Input to Input Dependency*

In the first set of experiments, we investigate the dependencies among the properties in *Layer* 1 ($T_{start}T_{end}$, *TrPlan*, and *Map*) and *Layer* 2 (R_{gyr} and P_{pause}), which are considered as inputs (Figure 2). While $T_{start}T_{end}$, *TrPlan*, *Map* are independent of each other, they can potentially affect R_{gyr} and P_{pause}. Figure 3A shows the deviation of the value of radius of gyration η computed from trace vis-a-vis the intended input value. As H decreases, points move away from the $y = x$ line, signifying that it becomes more difficult to get something close to the intended input. Figure 3B shows the deviation in pause time only for temporary WPs (green line) from intended input distribution, as the number of regular WPs is taken to be fixed. Recall *Step 1* in the procedure *preparePlan*() that a given period $\mathbb{T}$ ($= 24$

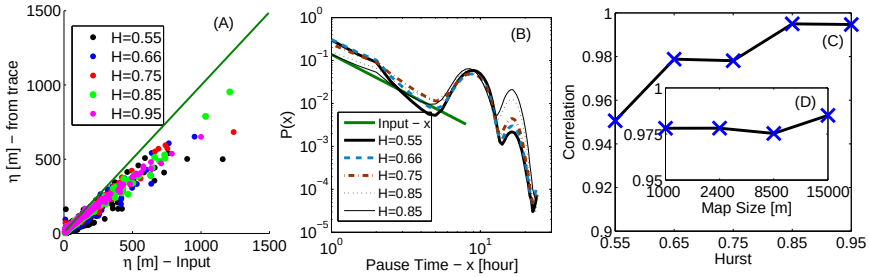

Fig. 3: Dependency among the properties in $Layer1$ (i.e. Input Properties). (A). Deviation of η for various H, (B). deviation of $P(\omega)$ for various H, and (C). correlation among the values that are intended and that are obtained for η with different H (inset of Subplot (C) - with different map sizes).

hours) is divided into two parts t_t and t_r, and t_t is distributed (following a power law) among the WPs in T_i'. Because, the number of WPs in T_i' increases with H for a fixed area of gyration, the probability of having a higher pause time reduces with higher H. Thus, obtaining the intended input distribution from synthetic traces becomes difficult due to input-input dependency.

Figure 3C and the inset measure the correlation between η computed from trace and η drawn from the theoretical distribution $P(\eta)$ for varying H and map sizes respectively. While the correlation increases with H for a fixed size map, it is hardly affected by the change in map size. This indicates that ***the distribution of WPs in the context of movement is more important than the size of the map*** in order to achieve the intended input. Hence, one should take special care to choose a set of input parameters while attempting to design a location specific human mobility model.

Important to note that some of the input-input dependencies shown in Figure 3 can sometimes depend on the implementation of the proposed algorithm. The number of WPs in the set T_i' for an agent u_i may change depending on a particular map, thus affecting the dependencies in Figure 3A and 3B. However, the order in which every agent is assigned its radius of gyration, or the order in which the waiting times are assigned at the WPs (i.e., whether the waiting times at the WPs in R_i is assigned before assigning the waiting times at the WPs in T_i') do not affect any of the observed dependencies.

5.4. *Input to Output Dependency*

Key Results - We have considered five input parameters ($Hurst$ H, visit fraction VF, the distribution of area of gyration $P(\eta)$, map size and number of temporary WPs) as representative measures for analyzing the input to output dependency. Figure 4 shows (A) complementary cumulative distribution function (CCDF) of the flight length, (B) probability distribution of return time, (C) CCDF of contact duration, (D) CCDF of inter-contact time, for different (I) Hurst values, (II) number of temporary WPs (VF), and (III) distribution of area of gyration ($P(\eta)$). The

results show that the CCDF of each of (ι), (σ), and (τ) follows power law (visually evident in a log-log plot) with a truncation. These cut-offs are partially the artifacts of fixed WPs and length of the map and fixed duration of movement each day. Also, as expected, $P(\varphi)$ decreases with increased time-gap and has periodic peaks. Thus, we show that the experimental set-up can faithfully reproduce important aspects of human mobility patterns and can be used to generate traces similar to human movement. In the following, we investigate the impact of these input parameters on the distribution of flight length, return time, contact duration and inter-contact time (Figure 4); the insights can be summarized as follows.

- Effect on flight length (F_{len}): The distributions of F_{len} follow a power law with a cutoff in all the cases, the cutoff threshold varies in different cases
  (1) Probability of having higher flights (e.g. larger than 100m) increases with higher H - Figure 4(I)A. Cutoff threshold in this case is 1000m.
  (2) Large number of temporary WPs (i.e. higher VF) in the trajectory forces F_{len} to be shorter - Figure 4(II)A. For example, probability of having F_{len} as 200m with $VF = 0\%$ (i.e., the waypoints only in R_i are visited, and no waypoint in T_i' is visited) is higher than that with $VF = 100\%$.
  (3) The exponent of the power law distribution of F_{len}s does not get affected by η, only the exponential cut-off shifts to a higher value as η increases (the cutoff shifts from 75m with $U(100)$ to 2250m with $U(Full)$) - Figure 4(III)A.
- Effect on return time (T_{ret}): The distributions of return times have periodic peaks in general that decreases as the return time increases
  (1) T_{ret} is not sensitive to H - Figure 4(I)B.
  (2) Probabilities to return after $12h$ and $24h$ to a place are higher with higher VF - Figure 4(II)B.
  (3) T_{ret} is not affected by the area of gyration - Figure 4(III)B.
- Effect on contact duration (CD): The distributions of CDs follow power law with a cutoff in all the cases, but both the cutoff threshold and the range of CDs for the distributions vary with the input parameters
  (1) Number of contacts with longer duration increases with H - Figure 4(I)C. Cutoff threshold in this case is 80 time steps (i.e. 8 hours).
  (2) Having longer contacts (e.g. more than 7 hours) is highly probable with $VF=0$, it decreases with higher VF - Figure 4(II)C.
  (3) Probability of having longer contact decreases as the area of gyration becomes larger - Figure 4(III)C. CDs with more than 10 time steps (i.e. 1 hour) can be hardly found with $U(Full)$, whereas contacts with 500 time steps can be seen with $U(100)$. Hence, though the distribution of CDs follow power law in both the cases, the distributions differ in the range of values observed for CDs.

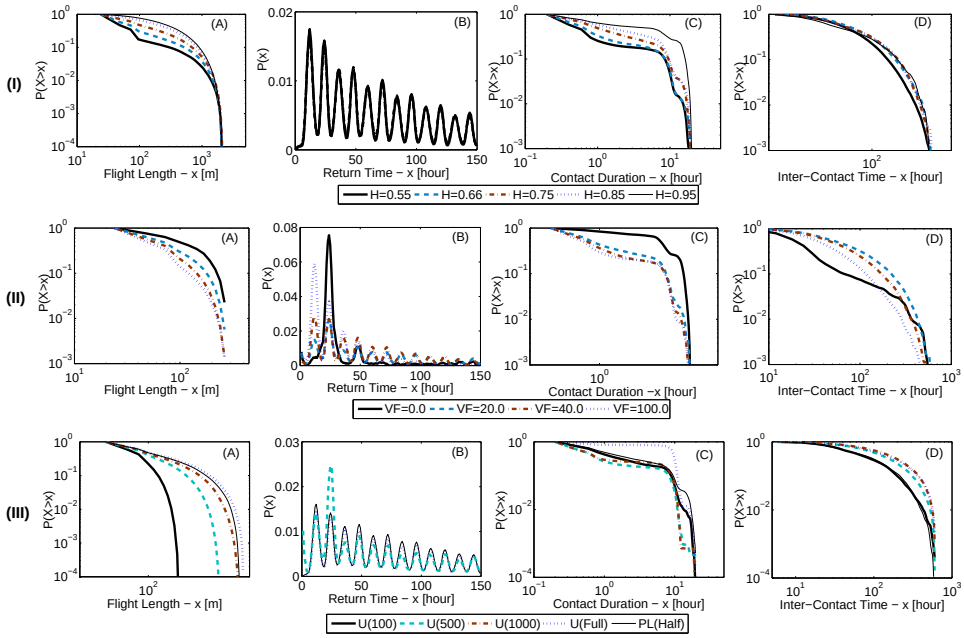

Fig. 4: Variations of (A) complementary cumulative distribution function (CCDF) of flight length ι, (B) distribution $P(\varphi)$ of return time φ, (C) CCDF of contact duration σ (D) CCDF of inter-contact time τ with various (I) Hurst H shown in the 1^{st} row, (II) number of temporary WPs ($|T_i'|$) VF shown in the 2^{nd} row, and (III) distribution of area of gyration shown in the 3^{rd} row.

- Effect on inter-contact time (ICT): The distributions of ICTs follow power law with a cutoff (at 60 time steps) which in general remain insensitive to the input parameters

  (1) ICTs are not affected much by H - Figure 4(I)D.
  (2) Probability to have larger ICTs decreases with higher VF - Figure 4(II)D.
  (3) ICT is insensitive to area of gyration - Figure 4(III)D.

Reasons for these observations can be described by (i) H that makes the WPs to be more clustered or scattered, (ii) η that defines the number of WPs for an agent, and (iii) fixed duration for every trip that makes an agent to routinely visit the WPs.

Impact of some other parameters - We have investigated the impact of other parameters such as the simulation area and the parameter α used in *prepareePlan()* method on the output properties (data not produced). None of these parameters have any significant impact on $P(\iota)$, $P(\varphi)$, $P(\sigma)$ and $P(\tau)$. This is primarily due to the distribution of the radius of gyration which is the same in both of these cases. This possibly indicates that any message spreading dynamics may remain same, but we shall see in next section that this is not true - the simulation area has a high impact on task oriented outputs.

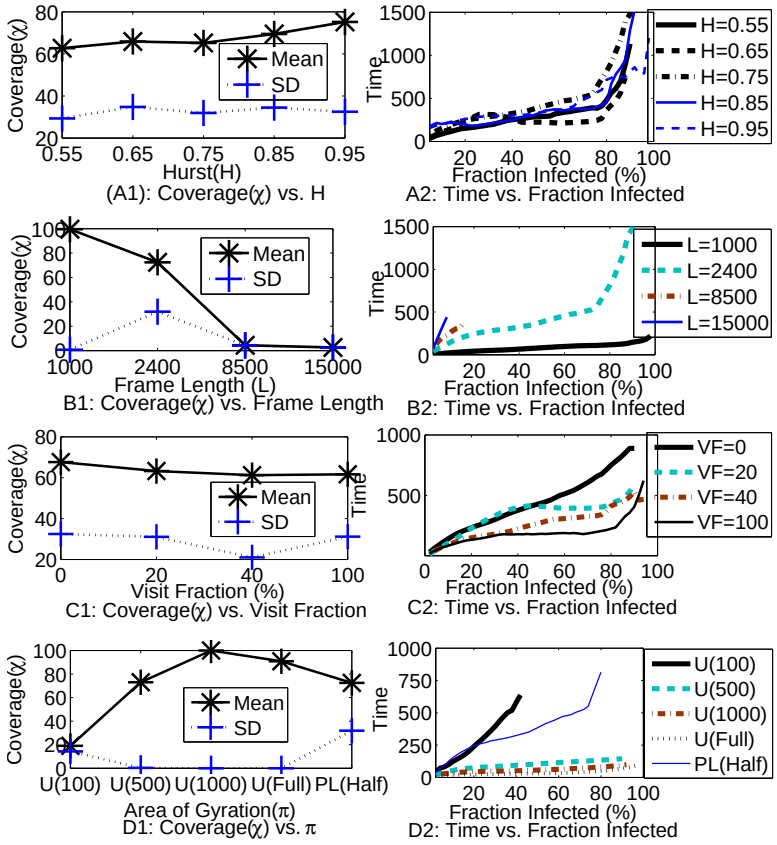

Fig. 5: Variation of the coverage and message spreading rate of epidemic broadcasting with Hurst value in $(A1)$ and $(A2)$ respectively, with map size in $(B1)$ and $(B2)$ respectively, with the number of WPs visited per day (VF) in $(C1)$ and $(C2)$ respectively, and with area of gyration (π) in $(D1)$ and $(D2)$ respectively.

6. Correlating Task-Oriented Output Properties

Mobility patterns are primarily synthesized to test the performance of protocols which run upon such patterns. Hence an additional layer which represents the performance metrics of the protocol need to be an integral part of the mobility framework - we call this layer *task-oriented output* layer. In this section, we investigate the relationships of input and output properties with task-oriented output properties. This analysis helps to identify an important set of input and output properties that can potentially affect a particular task in hand, e.g., coverage and spreading rate during broadcasting.

6.1. *Input to Task-Oriented Output Dependency*

To understand the relation between *Layer 1* and *Layer 5*, we have considered the simple epidemic broadcasting as a test case. The impact of H (Figures $A1$ and $A2$),

map size (L) (Figures $B1$ and $B2$), the number of WPs visited per day (VF) (Figures $C1$ and $C2$), and area of gyration (Figures $D1$ and $D2$) on the performance of epidemic broadcasting are shown in Figure 5 (a randomly selected agent is the source of a message, two metrics are considered *coverage* (χ), a measure of what percentage of the agents can receive the message, and message spreading rate, a measure of the time that is required to spread the message to a certain percentage of agents). The results show that the ***coverage increases with H, but the spreading rate is almost same*** until about 80% of agents are reached. This happens primarily due to the clustered nature of the WPs; intra-cluster spreading is much faster than inter-cluster spreading. The ***coverage steadily decreases with increased area***, mainly because of fixed area of gyration which is very small compared to the size of the map. As a result, the message stays only in the locality of the source.

Importantly, the message ***spreading is much faster when the agents try to visit more WPs per day***. Note that we kept the area of gyration fixed for all the agents in these cases. Even with this restriction, as the agents try to visit every WP in the area following LATP, the message is propagated like a wave from the agents' location and quickly spreads across that place and eventually, in the whole area. Finally, the results show that the ***coverage increase with average area of gyration***. However, there is a limit beyond which performance begins to fall. As we see that $U(Full)$ is performing worse than $U(1000)$ as the agents cover too much hardly coming in contact with other agents. The distribution of area of gyration plays a direct role in spreading rate. In case of *PL(Half)*, since there are a lot of agents with small area of gyration, the broadcast takes time to pick up. It picks up only when it meets an agent with larger area of gyration by chance. So, ***although the coverage is comparable between*** $U(1000)$ ***and*** **PL(Half)*, the spreading rate differs by an order of magnitude***.

6.2. *Analyzing Real Mobility Traces*

Four data sets on human mobility traces [12] (collected using 19, 33, 40, and 90 agents in a Statefair in North Carolina, NC State University (NCSU) campus, Disney World (Oralndo) in Florida, and KAIST campus in South Korea respectively that covers an area of roughly 1000^2m^2, 2400^2m^2, 8500^2m^2, and 15000^2m^2 respectively) are considered to investigate the dependencies between (i) input properties and output properties, and (ii) input properties and task-oriented output properties.

6.2.1. *Dependencies between Input Properties and Output Properties*

Using the four data sets, we analyze the impact of the radio range (an input parameter) on the flight length F_{len} and the inter-contact times ICT (both are output properties). Note that in this case the only input parameter that can be varied is the radio range, and hence we make two choices. We compute the average of the flight lengths ($\overline{F_{len}}$) and measure the average of the radius of gyrations $\overline{R_{gyr}}$ in all

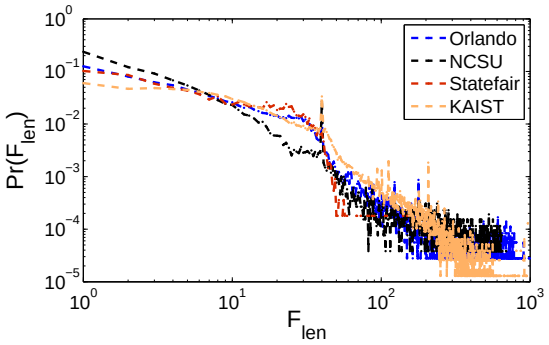

Fig. 6: Variation of the distribution of flight lengths measured from four sites indicated by four different colors in the figure.

the data sets; and the average of the ICTs ($\overline{ICT}$) obtained by varying the radio range from 10m to 100m in each of the data sets. Note that flight lengths and area of gyration are insensitive to the variation of radio range, therefore for a given data set, a single average flight length and area of gyration is a good representative. Note that the distribution of flight length usually follows a power law, as reported in [12]; here we observe the same in the Figure 6 for the four data sets. Results of $\overline{F_{len}}$ and $\overline{ICT}$ on all four data sets are reported in Table 3 along with respective $\overline{R_{gyr}}$.

- Average flight length is found to be highest in KAIST data set, while it is smallest in Statefair data set.
- The average flight length in KAIST data set is higher than that in NCSU and Orlando data sets, even though the average radius of gyration in KAIST data set is smaller than that in NCSU and Orlando data sets. This is partly because of higher Hurst value in KAIST data set, compared to that in NCSU and Orlando data sets. Also in KAIST the waypoints have to cover a large area ($8000^2\ m^2$). Such an observation is also exposed by our framework (shown in Figure 4IA).
- The relation of $\overline{ICT}$ with radio range has two phases - in the first phase when the radio range is too low to cover even a single cluster of waypoints, then increasing radio range decreases $\overline{ICT}$ (see the KAIST-$\overline{ICT}$ row (first three columns)). However, after certain enhancement, all the points in a single cluster fall within the purview of the radio range of an agent, so only points outside the clusters are connected through the antennae. Hence $\overline{ICT}$ value increases ((see the KAIST-$\overline{ICT}$ row (last three columns)). Subsequent increase in radio range again registers slow decline of $\overline{ICT}$ (Statefair-$\overline{ICT}$ (last three columns)).

Note that the distribution of inter-contact times follows a power law (as reported in [12]). Here, we report in Figure 7, the cumulative distribution of ICTs as we observe in the four data sets for various radio ranges.

- In Statefair data set, the probability of having smaller ICTs (less than 50 time steps) is higher when the radio range in tuned to a lower value (less than 40m).

Table 3: Results on real mobility traces. The values H of Hurst parameter (as measured in [12]), the average of flight lengths $\overline{F_{len}}$, the average of radius of gyrations $\overline{R_{gyr}}$, the average of inter-contact times $\overline{ICT}$, and the maximum coverage are reported for the four data sets considered for a set of radio ranges. Note that the distributions of F_{len}, and ICT both follow power law.

	Statefair Area - 1000^2 m^2					**Hurst (H)**	0.75			
						$\overline{R_{gyr}}$	205			
						$\overline{F_{len}}$	14			
Radio Range	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
$\overline{ICT}$	32	56	64	58	58	55	57	58	54	51
Coverage (%)	100	100	100	100	100	100	100	100	100	100
	Orlando Area - 2400^2 m^2					**Hurst (H)**	0.76			
						$\overline{R_{gyr}}$	1236			
						$\overline{F_{len}}$	26			
Radio Range	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
$\overline{ICT}$	40	55	56	59	58	58	55	55	53	54
Coverage (%)	49	72	75	75	86	86	88	89	89	90
	KAIST Area - 8500^2 m^2					**Hurst (H)**	0.82			
						$\overline{R_{gyr}}$	806			
						$\overline{F_{len}}$	30			
Radio Range	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
$\overline{ICT}$	24	21	21	22	23	24	25	27	30	32
Coverage (%)	99	100	100	100	100	100	100	100	100	100
	NCSU Area - 15000^2 m^2					**Hurst (H)**	0.66			
						$\overline{R_{gyr}}$	1060			
						$\overline{F_{len}}$	26			
Radio Range	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
$\overline{ICT}$	18	23	26	31	37	41	48	51	60	60
Coverage (%)	84	98.5	98.5	100	100	100	100	100	100	100

A similar observation is noticed in Orlando data set.

- In NCSU data set, the distribution of ICTs varies significantly compared to that in the other data sets. In this case, the probability of having smaller ICTs is higher using a smaller radio range. Such an affect may be a combined affect

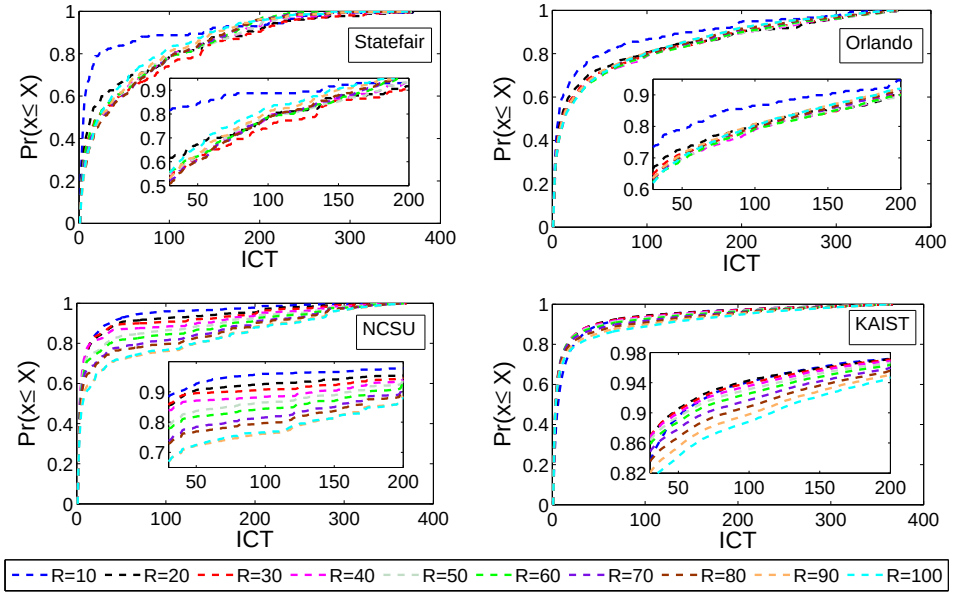

Fig. 7: The distribution of ICTs observed in four sites (Orlando, NCSU, Statefair, and KAIST) are shown in four subplots, using various radio ranges from 10m to 100m indicated by different colored curves in every subplot.

of the larger average radius of gyration and smaller Hurst value in NCSU data set.

6.2.2. *Dependencies between Input Properties and Task-Oriented Output Properties*

Once again the only tunable parameter with respect to these trace data sets is the radio range. We vary the radio range from 10m to 100m to analyze its impact on the coverage and report the results in Table 3.

- In KAIST data set, the coverage reaches 100% using a smaller radio range (=20m), which is mainly due to higher Hurst value ($H = 0.82$). Such a result is also revealed by our framework (see Figure 5$A1$).
- In Statefair data set ($H = 0.75$), the coverage of 100% can be achieved with 10m radio range.
- The coverage does not reach 100% even if radio range is tuned to 100m in the Orlando data set. Note that $H = 0.76$ in Orlando data set which is similar to Statefair data set. A possible reason for this is the fact that the radius of gyration is much higher in Orlando dataset (this observation is in lines with the results shown in Figure 5$E1$).
- Finally, in the NCSU data set, a higher radio range (= 40m) is required to

Table 4: Correlation between properties

Parameters	Correlation
Hurst - Flight Length	0.995
Visit Fraction - Flight Length	-0.906
Hurst - Coverage	0.933
Visit Fraction - Spreading Rate	0.933
Flight Length - Coverage	-0.571
Flight Length - Spreading Rate	-0.188

achieve a higher coverage compared to that (= 20m) in KAIST data set. This is partially due to a lower Hurst value ($H = 0.66$) in NCSU data set and the larger area which needs to be covered.

6.3. *Output to Task-oriented Output Dependency*

Another important aspect is to identify whether the mobility synthesis output (*Layer* 3 & 4) can be used to predict the quality of service (*Layer* 5). In this study we point out some caveats in exploiting such relations. Understanding this dependency through simulation is difficult as they both lie in output layers and therefore we perform correlation analysis for the same.

Table 4 shows some representative correlations, correlations of Hurst - Flight Length (0.995, *Layer* 1 - *Layer* 3) and Hurst - Coverage (0.933, *Layer* 1 - *Layer* 5) are both strongly positive. These are in agreement with Figure 4A and Figure 5A1 respectively. However, the correlation of Flight Length - Coverage (-0.57, *Layer* 3 - *Layer* 5) is negative, when it is computed by varying all the input parameters, H, L, α, and VF. Within a fixed area, an agent can increase its mean flight length (ι) by visiting the farthest point next; however, it might not meet any new agent in the long run as its area of gyration (π) is fixed. Thus, in such a case, it is π that affects the coverage, rather than ι. In our work, we have presently not varied the radius of gyration (η) directly, however, other inputs such as Hurst value which affect η are seen to affect ι and coverage to a large extent. A similar case is observed for changes in (a) flight length and (b) spreading rate with respect to visit fraction (VF). Finally, our analysis indicates that extreme care needs to be taken to establish such indirect relations and they need to be investigated on a case-by-case basis

While analyzing the correlation between the output properties and the task-oriented output properties, we have noticed that even though the coverage increases from 72% to 90% when the radio range (an input parameter) is increased from 20m to 100m, the average inter-contact time remains almost similar in Orlando data set (Table 3). Note that, in the other data sets, the coverage reaches 100% for a very low radio range, and hence no effect of ICT on the coverage can be anyway observed. Interesting to note here is that average inter-contact time (which is an output property belonging to *Layer* 4 in Figure 2) does not help us to understand the change in the coverage (which is a task-oriented output property belonging to

Layer 5) since, at least in the Orlando data set, we do not find any change in ICT whereas the coverage keeps changing with the increasing radio range. This essentially indicates that one has to look at the properties beyond *Layer* 4 in order to derive a proper dependency relationship with the properties in *Layer* 5. Thus, the layered structure proposed in this paper can potentially act as an important tool to unfold the complex interplay among the properties of human movement.

7. Conclusions

One of the major roadblocks in designing services for mobility assisted networks is the poor understanding of how mobility affects/can be leveraged to build better services. The primary contribution of this work is the development of a full-fledged framework to encapsulate the necessary characteristics of mobility models thus allowing for a rigorous analysis of the input-output dependencies. Some of the key insights on the inter-relationships among the properties of mobility patterns revealed by our analysis are as follows.

(1) An intended distribution of radius of gyrations or pause times that are input properties may significantly get changed by the characteristics of Hurst parameter which is another input property, even though these input properties may apparently look independent of each other.
(2) Hurst parameter may highly effect the distribution of contact duration which is an output property, but it may not effect the distribution of inter-contact times which is another output property.
(3) The number of waypoints visited per trip, and the distribution of radius of gyrations, both of which are input properties, can potentially affect all the output properties, i.e., the distributions of flight lengths, return times, contact durations, and inter-contact times.
(4) While Hurst parameter and the number of waypoints visited per trip hardly affect the coverage and the spreading rate of information dissemination which are the task-oriented output properties, the size of the map and the distribution of area of gyrations can have a strong impact on both the task-oriented output properties.
(5) Though both Hurst and flight length, and Hurst and coverage are positively correlated, flight length and coverage are negatively correlated.

One important message of the paper is that change in behavior of mobility may not always equally affect the overall task that we wish to perform. Hence, for each individual class of tasks, we need to identify the influential set of parameters/properties. Our future thrust would be to fill up this *service vs. parameter* matrix with respect to advanced services like routing and load balancing.

While a number of interesting insights are noticeable from the analysis presented in this paper, our study has some certain limitations. First, the dependencies exposed by our study may not be observed in various available real world mobility

traces because the traces are either collected for a short duration or limited in number of agents considered given the resolution of the location to which the agents are exposed. Second, an answer to the problem that we consider in this study may not be found from an existing mobility model because a mobility model, in general, assumes only a few properties (two to three) in order to enhance its usability and simplicity - however, the main issue lies in choosing them in a principled fashion. Importantly, the results of our analysis suggest that a more systematic real world trace collection of human movement is required in order to shed light on the importance of an observed property that may potentially lead to the design of a new mobility model. This would allow the use of realistic mobility models in various application scenarios where toy models (like random walk) are grossly unsuitable.

References

[1] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 2006.

[2] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communications & Mobile Computing: Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2:483–502, 2002.

[3] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6:606–620, June 2007.

[4] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of ACM SIGKDD*, pages 1082–1090, San Diego, California, USA, August 2011.

[5] DTNRG. Delay tolerant network research group. http://www.dtnrg.org.

[6] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008.

[7] R. Jain, D. Lelescu, and M. Balakrishnan. Model t: an empirical model for user registration patterns in a campus wireless lan. In *Proceedings of IEEE/ACM MobiCom*, pages 170–184, Cologne, Germany, August 2005.

[8] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović. Power law and exponential decay of inter contact times between mobile devices. In *Proceedings of IEEE/ACM MobiCom*, pages 183–194, Montreal, QC, Canada, September 2007.

[9] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella. Human mobility models for opportunistic networks. *IEEE Communications Magazine*, 49:157–165, 2011.

[10] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella. An arrival-based framework for human mobility modeling. In *International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM)*, pages 1–9, San Francisco, CA, USA, June 2012.

[11] M. Kim, D. Kotz, and S. Kim. Extracting a mobility model from real user traces. In *Proceedings of IEEE INFOCOM*, pages 1–13, Barcelona, Spain, April 2006.

[12] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. Demystifying levy walk patterns in human walks. Technical report, CSC, NCSU, 2008.

[13] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. Slaw: Self-similar least-action human walk. *IEEE/ACM Transactions on Networking*, 20:515 –529, Apr 2012.

[14] X. Liang, J. Zhao, L. Dong, and K. Xu. Modeling collective human mobility: Understanding exponential law of intra-urban movement. *CoRR*, abs/1212.6331, 2012.

[15] A. Passarella and M. Conti. Analysis of individual pair and aggregate intercontact

times in heterogeneous opportunistic networks. *IEEE Transactions on Mobile Computing*, 12:2483–2495, 2013.

[16] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong. On the levy-walk nature of human mobility. In *IEEE/ACM Transaction on Networking*, volume 19, pages 630–643, 2011.

[17] C. Song, T. Koren, P. Wang, and A.-L. Barabasi. Modelling the scaling properties of human mobility. *Nature Physics*, 6:818–823, 2010.

[18] T. Spyropoulos, A. Jindal, and K. Psounis. An analytical study of fundamental mobility properties for encounter based protocols. *International Journal of Autonomous and Adaptive Communications Systems*, 1:4–40, 2008.

[19] P. Tournoux, J. Leguay, F. Benbadis, J. Whitbeck, V. Conan, and M. Dias de Amorim. Density-aware routing in highly dynamic dtns: The rollernet case. *IEEE Transactions on Mobile Computing*, 10:1755–1768, 2011.

[20] C. Tuduce and T. R. Gross. A mobility model based on wlan traces and its validation. In *Proceedings of IEEE INFOCOM*, pages 664–674, 2005.