MobIPLity: A trace-based mobility scenario generator for mobile applications*

Nuno Cruz*,+
ncruz@deetc.isel.ipl.pt
*Instituto Superior de Engenharia de Lisboa
Instituto Politécnico de Lisboa
ADEETC

Hugo Miranda+
hmiranda@di.fc.ul.pt
+Faculdade de Ciências
Universidade de Lisboa
LaSIGE

ABSTRACT
The understanding of human mobility patterns is key for the development and evaluation of ubiquitous applications. To circumvent the scarcity and difficulties in capturing mobility data, a number of models has been devised. The accuracy in replicating observed human mobility by these models varies. In general, each model concentrates in replicating some of the metrics that have been observed, while neglecting others. Unfortunately, all tend to neglect diversity, in the roles and goals of the users but also in the devices that are used to access the wireless network.

This paper presents the mobility traces that could be extracted from the access records of 49000 devices on the 190+ access points of the eduroam WiFi network on the campus of the Lisbon Polytechnic Institute between 2005 and 2012. The traces are made publicly available in the expectation that its large scale permits to support evaluations based exclusively on real mobility data, thus removing the uncertainty that emerges from the use of synthetic mobility models. Traces emphasise the differences that can be found between device types, with impact on aspects like the observed trace duration, speed, pause times, ICTs and availability and which can hardly be replicated on synthetic mobility models.

Categories and Subject Descriptors
I.6 [**Simulation and Modeling**]: Model Development; C.2.1 [**Network Architecture and Design**]: Wireless Communication

General Terms
Measurement, Performance, Algorithms

*This work was supported by FCT through funding of LaSIGE Strategic Project, ref. PEst-OE/EEI/UI0408/2014

MOBIQUITOUS 2014, December 02-05, London, Great Britain

DOI 10.4108/icst.mobiquitous.2014.257986

1. INTRODUCTION
Simulation plays an important role in the validation and evaluation of applications and protocols for wireless networks as they permit to circumvent the difficulties in deploying large scale and long term real experiments. Simply stated, network simulators put to test an implementation of the application against abstractions of the environment, traffic and node movement. Therefore, the reliability of the experiments is strongly influenced by the quality with which each of these abstractions reproduces reality.

Research has been pursuing two approaches: pure synthetic and trace-based synthetic mobility models. In pure synthetic mobility models nodes move according to some predefined statistical function in an attempt to realistically represent device/human movement. The commitment of the rules defined for synthetic mobility models in the replication of observed user movement patterns vary but this model has been criticised for its inability to reproduce human movement patterns, when evaluated by metrics like inhomogeneity [4].

Trace-based synthetic mobility models apply statistical distributions to observations of user movement, thus mirroring properties observed in real traces of human mobility (e.g. [5, 9, 15]). Traces are provided either by volunteers, which make their location available, or by a third party performing passive observation. Unfortunately, the number of trace-based samples made publicly available is scarce, present a small time span and/or number of users.

To circumvent the limitations of trace-based mobility models, this paper presents a study that consisted on the observation of the 49000 wireless devices that connected to the Eduroam network of the Polytechnic Institute of Lisbon, Portugal (IPL) between 2005 and 2012. The study permitted to identify approximately 7500 distinct devices running operating systems for small devices (IOS, Android and Windows Mobile) and about 33000 devices running operating systems usually found on laptops [7]. Data collected includes the moments of association and disassociation of mobile devices to access points and the geographic location of the access points.

The distinction between mobile device types allows the generation of mobility scenarios that better purport the characteristics of the group of devices for which an application is developed.The paper shows that different device classes have different behaviours, with impact on the metrics typically modelled by trace-based synthetic mobility models.

The set of traces in this study is made publicly available using a web interface that exports data in bonnmotion [1] format, making it available to a broad range of network simulators used for application framework and protocol evaluation. Expectations are that the scale of the dataset contributes to the definition of *trace-set mobility models*, a new class that, in contrast with previous approaches, creates mobility scenarios using exclusively real data.

The contributions of this paper are two fold. First, it presents MobIPLity, a trace-set mobility model and scenario generator. Secondly, it characterises and compares the traces of MobIPLity using several metrics and trace-based models found in the literature. Results show that the latter fails to emulate multiple device types and that require a careful and difficult parametrization to correctly model reality.

The paper is organised as follows. The next section makes a brief overview of different mobility models, efforts to collect mobility data and metrics that have been used to characterise human mobility. Sec. 3 and 4 respectively present the methodology used to convert the raw dataset on trace-based mobility scenarios and the different metrics associated with it, while on Sec. 5 one can find the discussion and comparison between MobIPLity and other approaches that can be found on the literature. The conclusions appear on Sec. 6.

2. RELATED WORK

Pure synthetic mobility models employ random distributions to simulate device movement. A classical example is the Random Waypoint Mobility Model (RWP) whose simplicity in the generation of mobility scenarios facilitated cross comparison of mobile applications and protocols. However, it has been shown that, in addition to their disparate modelling of human behaviour, synthetic mobility models typically bias node distribution in a non-natural way [4]. Limitations of the RWP have been addressed, for example in [9], were a variation of the traditional RWP to produce patterns presenting the same inhomogeneity as found in human mobility was proposed.

Trace-based synthetic mobility models, on the other hand, attempt to mirror patterns observed in human movement by modelling nodes behaviour according to some probabilistic distribution functions. The mechanisms used for collecting data inspiring trace-based mobility models can be arranged in two categories. Intrusive approaches (for example [18,21]) are those that obtain their data directly from the device carried by the user. These approaches benefit from the precision of the data, captured by dedicated software or hardware. Unfortunately, these studies are constrained by the considerable amount of resources involved, which limit their time scale and number of participants and may bias conclusions concerning the identification of patterns.

Non intrusive approaches use logs collected by external devices (like access points or indoor-localisation devices) to produce traces with the user location at each instant. In spite of the privacy issues raised with the collection of the data, non intrusive approaches are those that present the capability to scale more in both number of users and time span. Unfortunately, surveys on mobility models [3, 13] indicate a scarcity of traces from mid-2008 onward, thus excluding the massification of mobile devices observed with the emergence of the last generation of smart phones and tablets. If available, more recent traces could evidence the emergence of new mobility and contact patterns among users.

The WiFi network of the Dartmouth College has been serving for collecting a considerable number of traces, for example during the 17 weeks of the 2003/2004 winter semester [10]. The method for collecting the traces is very similar to the one used in this paper and which is further described in Sec. 3. Authors used the logs to model real user tracks and defined a threshold walking speed, below which users were assumed to have stopped before moving to the destination. This knowledge was used to define a trace-based synthetic mobility model [14] inspired on the mobility patterns of 198 VoIP handsets. The model addressed social, spatial and temporal features and considered hotspots, workday/weekend distinction, and mobile and stationary sets. In comparison with MobIPLity, the study of 2003/2004 evaluates a larger number of access points, but a lower number of users and a shorter time frame. Unfortunately, its age prevents it from considering a number of recent advances, like the most recent wave of mobile devices, such as the iPhone and Android OS based smartphones (debuted respectively in 2007[1] and 2008[2]).

Results on a two month study on the eduroam infrastructure of the universities of Minho and Vigo can be found in [16]. The methodology followed is very close to the one used in our study. The access point association to physical spaces allowed to separate network traffic originating in residential from academic areas. Authors found that the APs with more users are not necessarily the ones with more network traffic. In addition, the paper evidences a weekly use pattern for this network, with the vast majority of users connecting only on weekdays. In terms of mobility, authors conclude that 90% of the users connect to more than one AP monthly, with about 35% visiting at least 5 APs. Unfortunately, the small analysis period of this study makes the notion of mobility disperse in time and of little relevance in the characterization of real mobility.

2.1 Mobility Patterns

The mobility of individuals has been characterised along spatial, temporal and social axis [13]. The deployment density is portrayed on the spatial axis by metrics like *jump size* (sometimes referred as *flight*) which characterises the average distance of each movement. Trace-based mobility models have been modelling jump sizes with either log-normal [14] or truncated power law distributions [5, 15]. Jump sizes depended on the physical area where simulation is taking place. For example, the jump size is influenced by the distance between buildings. The *inhomogeneity metric* [20] aims at evaluating the distributions of the individuals on the physical space, highlighting hot-spots. The variation of the Random Way-Point presented in [9] and the Disaster Area [2] are good examples of mobility models enforcing a heterogeneous node distribution. A lower inhomogeneity value is expected

[1] http://en.wikipedia.org/wiki/IPhone

[2] http://en.wikipedia.org/wiki/Android_(operating_system)

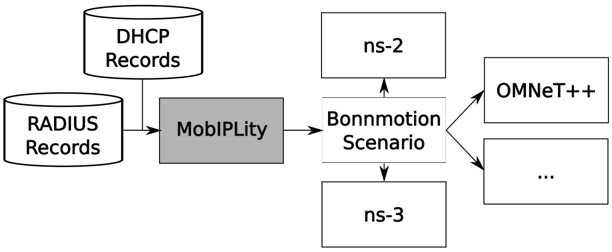


Figure 1: MobIPLity Workflow

from random distributions, while a higher value shows that users are creating groups, formed by nodes placed in popular locations.

Spatial properties are usually tied with temporal ones, associating the time to the distance travelled between two points. Time-varying properties of human mobility characterize patterns such as workday/weekend variations and the time spent on a specific place (known as pause-times).

The social axis characterises the meetings between individuals. In combination with the temporal axis, they contribute to determine how long or how frequently two or more individuals meet. Multiple models with a strong focus on the social relationships established between individuals have been proposed. Metrics considered include attraction (found for example in [5]) but also repulsion. Both properties are explored in [8] by combining the modelling of relationships with individual walks and group trips. The *inter-contact time* (ICT) is a frequently used metric to evaluate the social axes of a mobility model. It is defined by the time between two consecutive contacts between two nodes. ICT is usually modelled using a truncated power law distribution [6,12,15].

3. MOBIPLITY

MobIPLity, the trace-set mobility model proposed in this paper, is responsible for the generation of mobility scenarios. As depicted on Fig. 1, the algorithm combines records produced by the DHCP and RADIUS services, with the former contributing for the identification of the device types and the later identifying the participants and the moment of each association/dissociation event. This section begins with a characterization of the data set and then proceeds with the description of algorithm that is used to generate the mobility scenarios.

3.1 Data-Set Characterisation

The data set used in this study is composed by the log records produced, between January 1, 2005 and December 31, 2012, by all Access Points (APs) of the eduroam Wi-Fi network of the Lisbon Polytechnic Institute (IPL). A total of 48699 devices and 30629 distinct users accessed the network during this time frame, producing about 31 million records.

IPL is the 7th largest teaching institution in Portugal with approximately 1300 teachers and 15000 students, distributed by 10 distinct sites around the Lisbon metropolitan area (see Fig. 2). The IPL's eduroam network is supported by approximately 200 Cisco Systems APs, covering a total of 26 buildings and inter-building areas. Records are originated

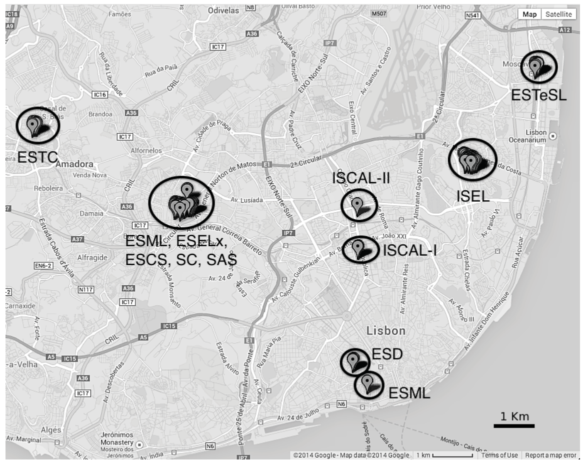


Figure 2: Location of IPL sites

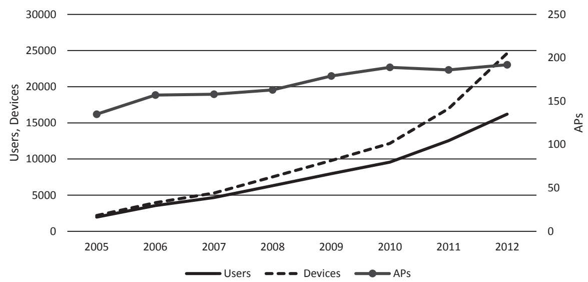


Figure 3: Evolution of devices, users and access points

from all the users accessing the network, thus also including visitors from other national and international institutions.

Our study arranges devices in two classes: *Small Mobile Devices* (SMD) are small and can be used on the move. Examples of mobile devices are smart-phones, PDAs and tablets. The second class, *Laptops*, group the larger devices, usually running over a classical operating system (Linux, Windows or MacOS). DHCP logs permitted to identify 7592 *SMDs* and 33054 *Laptops*.

Figure 3 shows a continuous growth of the number of users and devices, although at distinct rates, specially since 2010. This is coincidental with an increase in the sales of smartphones observed at the national level and suggests that the number of users accessing the network with more than one device has been increasing. This is confirmed by Fig. 4, which compares the proportion of SMDs and Laptops in each year.

The collection of mobility data is centred on the logs produced by the RADIUS [19] protocol. Log entries reproduce the RADIUS session concept thus considering the association of each user to a single AP. Log session records contain the device MAC address, AP id, user name, session start and stop times. Logs have been edited by:

- merging in a single record consecutive sessions between the same device and AP with an interval of less than 5 seconds. These sessions are attributed to network card

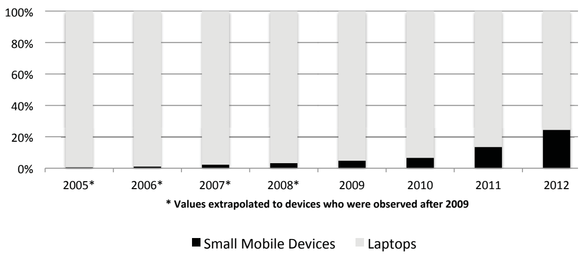


Figure 4: Laptops vs SMDs

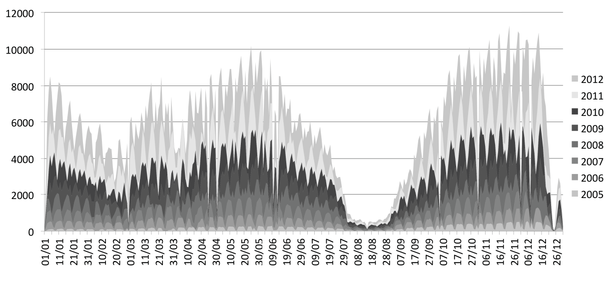


Figure 5: Wifi Devices Connected Per Day

or driver problems;

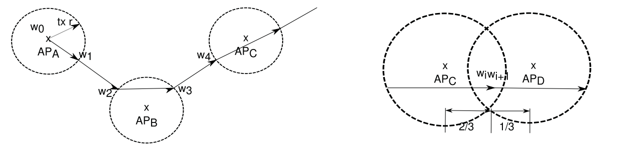
- removing concurrent sessions of the same device to distinct APs. This is an impossibility that can only be explained if the device did not disassociate correctly from one AP before associating to the next and the former artificially defined the session stop time upon a timeout. In this case, the session stop time of the earliest session was corrected to happen immediately before the start time of the latest;
- removing sessions with the stop time equal to the start time. Sessions with this characteristics are created when a user has some problem when connecting to the network, although the network considers the user authenticated (thus creating the RADIUS record).

Fig. 5 depicts the number of devices that connect daily. As expected, the plot exhibits an irregular pattern consistent with the different activity levels that can be found on workdays, weekends and summer and winter breaks in a campus.

3.2 Trace Generation

The MobIPLity trace-set mobility model is created from a set $E \subseteq D \times A \times \{in, out\} \times T$ where D is the set of wireless devices, A the set of access points of the network annotated with their geographical coordinates and T are time stamps. The set is populated with 2 events $(d, a, in, t_1), (d, a, out, t_2)$, for each RADIUS log record (made available by the eduroam network of IPL), with t_1 and t_2 time-stamping respectively d's association/disassociation to AP a. Finally, let $E_d \subseteq E$ be the subset of E containing all the events recorded for device d.

E_d is expected to respect two invariants: *i*) devices are always associated with an access point before being disassociated from it; and *ii*) in any point in time, a device is associated at most to one access point. Note that invariant *i*) is trivially assured by the access points software and invariant *ii*) by the corrections applied to the RADIUS logs that have been outlined in Sec. 3.1.



(a) With disconnections (b) Overlapping coverage

Figure 6: Trace extraction examples

We define $E'_d = e_{d,0}, e_{d,1}, \ldots, e_{d,n}, d \in D, e_{d,i} \in E_d, i > 0$ as the temporally ordered set of events for device d. Note that invariant *i*) ensures that $e_{d,2j}, j \geq 0$ are events of type *in* and, conversely, $e_{d,2j+1}, j \geq 0$ are all events of type *out*.

A trace $W_d = w_0, w_1, \ldots, w_{2n-1}, w_i = F(e_{d,2j+i}), 0 \leq i \leq 2n-1, j, n \geq 0$ for some device d, is defined as a sequence of way-points w_i. The way-points are defined by a geographical coordinate and a time stamp, returned by a function F applied to consecutive events (not necessarily $e_{d,0}$) in E'_d.

The output of function F depends of the position of the way-point on the sequence and of the type (in, out) of the event. The general case is depicted in Fig. 6a. w_0 is set with the time stamp of $e_{d,2j}$ and the coordinates of the access point in the first of the sequence of events selected. Subsequent transformations of pairs of events on pairs of Way-points $w_{2i+1}, w_{2i+2}, i \geq 0$ will return coordinates overlapping a vector $\overrightarrow{AP_A AP_B}$, with AP_A, AP_B being the coordinates of the access points in the corresponding events $e_{d,2j+2i+1}, e_{d,2j+2i+2}$. The precise locations are dictated by the predefined transmission radius of the access points, as w_{2i+1} (resp. w_{2i+2}) will be placed at the intersection of the vector with the transmission radius of AP_A (resp. AP_B). Time stamps of w_1 and w_2 are copied from the corresponding events. Notice that, according to the definition of E'_d above, events $e_{d,2j+2i+1}, e_{d,2j+2i+2}$ are respectively an *out* and an *in* record, thus signalling the moment at which d abandoned the area covered by AP_A and the moment at which d associated with AP_B. The algorithm is successively repeated for each pair of events and way-points.

In the particular case when the coverage area of two consecutive access points visited by the device intersects (Fig. 6b), the algorithm reflects the conservative approach of wireless interface drivers. The two way-points receive the time stamp of the *in* record and are set at 2/3 of the distance between the access points, chosen to reflect expected driver behaviour by selecting the strongest signal only after ensuring that the difference between both signals is enough to ensure better connectivity.

Traces are terminated at an *out* event by creating a way-point with the coordinates of the access point. It is assumed that the device abandoned the eduroam network and therefore that the trace must be terminated when the speed for traversing the distance between two consecutive access points falls below some threshold. Alternatively, traces are terminated when two consecutive connections to the same

Table 1: Trace extraction options

Start/End Date	Not before/after dates of the records
# Devices	Number of devices
Device' Type	Laptops/SMD/Any
Points	Minimum number of APs in a trace
Duration	Trace duration
Axes	2D/3D representation
Location	School for extraction
Warm Up	Duration of the warm up period
Cool Off	Duration of the cool off period
Enhanced trace	Include terminated traces
AP Range	Radius of the AP coverage
Speed	Min. speed for consecutive APs
Time	Max. time between reconnections

AP exceed a time threshold, suggesting that the users have abandoned the location.

3.3 Web Interface

To facilitate the dissemination of the traces, a web interface has been prepared and made publicly available at `http://edata.e.ipl.pt`. Traces are extracted on-demand from an E set stored at a local database and running the algorithm described in Sec. 3.2 . Table 1 lists the configuration parameters that can be configured by the users.

A warm up and a cool off period are defined immediately before and after the period selected by the user for the scenario. These periods ensure that only traces that remain active for the entire duration of the simulation are selected as they must include at least one way-point in each. Alternatively, the "Enhanced trace" option permits to consider traces that start and/or terminate during the period defined for the scenario.

The algorithm outputs traces in the format used by the Bonnmotion mobility scenario generator and analysis tool [1]. The algorithm uniforms the output, shifting the time stamps and creating an initial way-point for each trace with the location predicted for each device at scenario starting time. Finally, all access points are consistently repositioned using a random factor.

3.4 Enforcement of User Privacy

To protect the confidential nature of the data, original records are kept at a secure location and cannot be disclosed. In compliance with the Bonnmotion file format, the algorithm exclusively outputs (time,coordinates) pairs and therefore, no identification that could be associated directly with a user or device is released. In addition, original data is obfuscated by: i) Repositioning way-point coordinates in each new scenario generation while maintaining coherence; and ii) starting all scenarios at time 0, without disclosing the offset between the requested start date and the effective beginning of the scenario.

To somewhat limit any judicious analysis of the data that could be crossed with information made available from other sources, all requests of scenarios will be moderated. Boundaries on the duration of the scenarios may also be applied.

Table 2: Overview of the 2012 trace set

	All		Laptops		SMDs	
	IPL	ISEL	IPL	ISEL	IPL	ISEL
Devices	24141	10080	14947	7066	5403	2056
Traces	2075731	985816	1061686	602620	641394	250209

4. MOBILITY ANALYSIS

This section presents and discusses the characteristics of the mobility patterns found on the MobIPLity trace-set using the inter-contact times (ICT), jump size and pause times metrics, outlined in Sec. 2.1. This analysis addresses exclusively the trace-set of 2012, chosen for being the most recent, and, reflecting the evolving use of the technology, richer in terms of number and variety of device types. The section looks into the 2 alternative configuration options for types of devices, permitting to confirm the existence of distinct mobility patterns for users carrying large ("laptops") and small ("SMD") devices.

To further increase diversity, the analysis considers the ISEL and IPL locations. These are the contrasting extremes concerning node density. ISEL is the engineering school of IPL, located in a single site and provides the largest number of devices from a single location. The IPL option considers records collected from access points at all schools (including ISEL) and presents a very small node density as campus are distributed over Lisbon metropolitan area (Cf. Fig. 2).

4.1 General Dataset Metrics

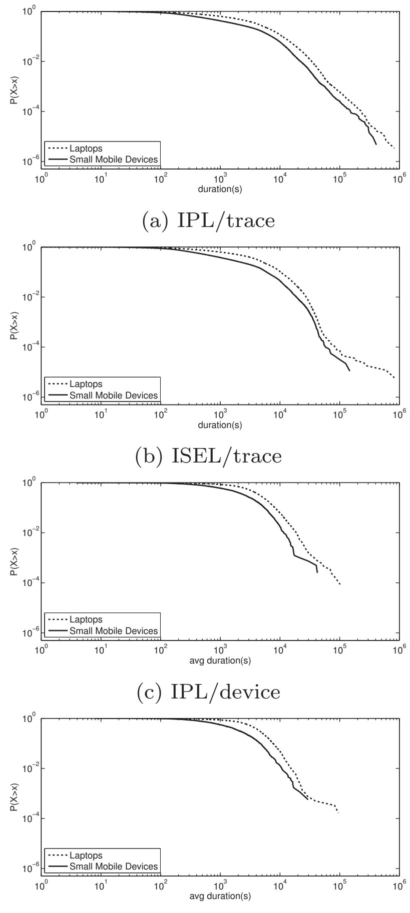
Table 2 show the dimension of the trace sets that are the focus of this section. As it can be seen from the table, ISEL accounts on average approximately 40% of the devices and of the number of traces. Discussion proceeds in two complementary perspectives. A global, "Per trace" perspective considers each trace independently. On the "Per device" perspective the traces are aggregated and the metrics averaged for each device.

4.1.1 Trace duration

Figure 7 shows the complementary cumulative distribution function (CCDF) of the duration of each trace. Long trace durations were expected, specially for Laptops, due to the coverage provided by IPL eduroam to student dorms. The figure also shows that less than 18% of the traces for Laptops exceed 2 hours and that for SMDs this value further decreases to about 10% of the traces in IPL and 7% in ISEL.

Such small proportion of "long traces" is surprising. Note for example that, in contrast with our expectations, there is no 6 hours slot for which 100 SMDs were permanently connected to the network in the whole trace set of 2012. One would expect that the usage pattern reflected the increasing use of mobile devices on the campus and, therefore, that trace durations were consistently higher.

The small duration of the traces is attributed to the energy-saving mechanisms that can be found on mobile devices and which automatically disable the wireless interface when not in use or when the screen is turned off. This is an aspect that has been consistently ignored in trace-based mobility models and is even hard to reproduce in network simulators. However, this feature has a non-negligible impact on



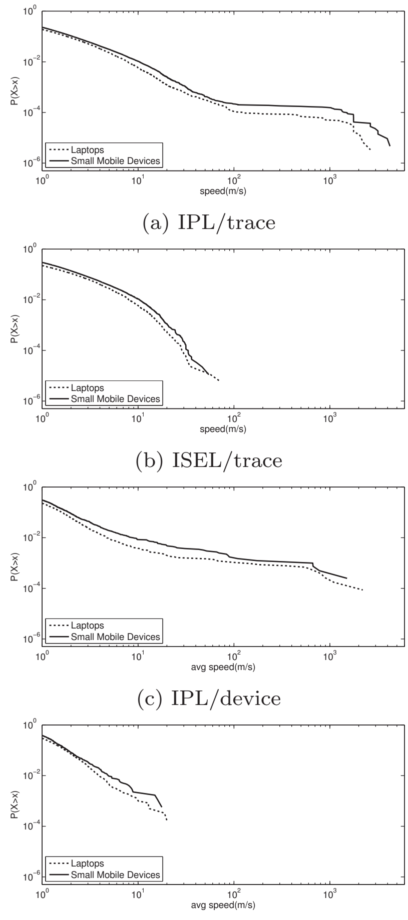
(a) IPL/trace

(b) ISEL/trace

(c) IPL/device

(d) ISEL/device

Figure 7: Trace duration (seconds)



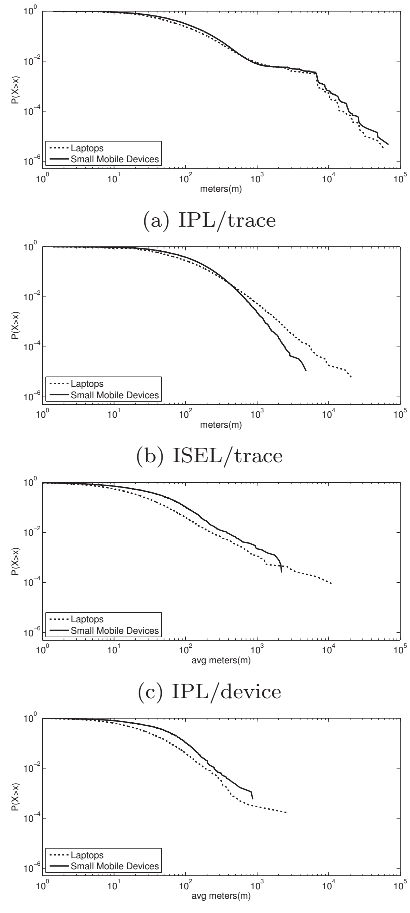
(a) IPL/trace

(b) ISEL/trace

(c) IPL/device

(d) ISEL/device

Figure 8: Trace speed (m/s)



(a) IPL/trace

(b) ISEL/trace

(c) IPL/device

(d) ISEL/device

Figure 9: Trace length (meters)

the design and evaluation of many protocols and applications for ad hoc and delay-tolerant networks which assume "always on" connectivity of the devices. As a simple example, consider the impact of intermittent connectivity on the route discovery phase (that uses flooding) of many reactive routing protocols for MANETS, such as DSR [11] and AODV [17]. A more in depth investigation of the impact of the power saving mechanisms is out of the scope of this paper and left as future work.

4.1.2 Speed

Speed plots (Fig. 8) exhibit some interesting results that are attributed to the ping-pong effect that result from a combination of the fast roaming of the devices between overlapping APs and the trace generation algorithm used. This is a problem that has been experienced in other models (e.g. [14]) and has a negligible impact on the model as these fast speeds occur for very small amounts of time and distances. It should be noted that a portion of the 15% of the traces with an average speed above the average human movement speed on the IPL trace set can be attributed to users moving between sites, and that 18% of the IPL traces and 30% of the ISEL traces have a speed under 1m/s, which simply suggests users walking at low speed. However, when looking at the differences between device types we can find that SMDs always present a higher speed than laptops, which is something we expect considering laptops usage pattern.

4.1.3 Distances Travelled

The distance travelled on each trace is evaluated using two metrics. Trace length (Fig. 9), measures the length of each trace in meters. The geographical disposition of IPL sites and the roles of some of its members results in some traces obtaining surprising values of 100 Km. However, averages are more predictable and only reach 11Km for IPL and 800 meters for ISEL.

As expected, the higher mobility of SMDs is confirmed by longer average traces in conjunction with shorter durations. However, looking at the complete trace-set of IPL we cannot differentiate between different device type distributions. This is expected, as users that carry a laptop also carry a SMD and as such when they travel between IPL locations they carry both devices with them. Results show that about 27% of all devices on IPL and 16% on ISEL are static, contributing for the 70% of the traces without movement. However, the size of the MobIPLity trace set is sufficient to attenuate this large proportion and more than 600000 traces for IPL can be found exhibiting movement. Authors believe that such a large number should be considered sufficient to support a trace-set mobility model.

The distribution of jump sizes (i.e. the distance travelled between way-points) is depicted in Fig. 10 and clearly shows how the multiple sites of the IPL trace set impact the model. The density of ISEL justifies the smaller travelled distances.

4.1.4 Pause times

Figure 11 shows consistently briefer pause times for SMDs on both IPL and ISEL trace sets. This supports common knowledge of SMDs showing a higher mobility, what contrasts with the expected large pause times for laptops, typically operated by steady users. The longer tail on the plot for Laptops can be caused by laptops that are kept at teachers offices, or at students dorms. The logarithmic scale of the graph hides the large difference between the maximum pause time for laptops (almost 8 days) and a maximum of 2 days for SMDs. The difference between these values is consistent on IPL and ISEL.

4.1.5 Disconnection time

Figure 12 presents the CCDF for the time for which devices were disconnected, creating distinct traces. This metric was only obtained for devices that returned to the network after a disconnection. The figure clearly shows the impact of the academic environment where the data was collected.The plot shows steps indicating a considerable number of disconnections of 12 hours, 2 days, 12 days, 2 months and 6 months. These periods represent either weekend/weekday periods, vacations and semesters. We also found that laptops have a higher probability of being disconnected frequently for periods of 90 minutes, which is the duration of classes. In contrast, the figure suggests that SMDs are frequently connected in classrooms during lectures.

4.2 Scenario Metrics

To illustrate the capabilities of MobIPLity, 2 distinct scenarios were generated and compared with the related work using frequently cited metrics. The "3 days" scenario aggregates traces of the 3 days where more SMDs have been found, namely the 22nd of May, 18th of October and 6th of December, 2012. This criteria was used in an attempt to circumvent the difficulties in identifying traces of SMDs with at least 2 hours, discussed in Sec. 4.1.1, and to allow the comparison using a similar set of parameters as used on related work. MobIPLity was configured to select 2h traces with warm up and cool off periods also of 2h. Minimum speed for jumps was set at 0.5m/s and the maximum time for reconnects to be considered as belonging to the same trace was set to 120s. It is assumed that both access points and mobile devices have a 50m transmission range. With these parameters, a second scenario, named "disconnected", was extracted for only the 18th October, and includes interrupted traces, i.e. traces with devices that turn off their radio during the period.

Metrics presented below consider the maximum number of devices available during the days extracted for analysis (respectively, 15, 20 and 18 devices for ISEL and 28, 43 and 45 for IPL). The "disconnected" and the "3 day" scenarios share all the configuration parameters with the exception of the option to include traces interrupted during the period. Expectations are that the differences observed between the two scenarios can give some hints on the impact of node disconnection.

4.2.1 ICT

Figure 13 depicts the CCDF for the Inter-contact times of the 8 surveyed scenarios. The irregularity of the plots for connected scenarios is attributed to the small number of devices that remained connected during the 2h period. Results show longer inter-contact times for devices that are geographically bounded to ISEL. This is expected as the density of the network is considerably higher, what increases the probability of the nodes to keep in touch for more time.

4.2.2 Jump size

The CCDF of jump sizes for IPL (Fig. 14) show a distinctive step pattern attributed to the considerable distances between the different schools of the institution and to the need of some users to commute between them. As expected, the considerably smaller area of ISEL constrains the maximum travelled distance. It is interesting to notice the differences between SMDs and Laptops, with jump sizes of the former having a higher probability of being shorter than the latter. Something that can be attributed to the mobility of the devices, which favours more frequent connections to distinct APs.

Figure 14a presents an interesting exception to the relation of the curves presented by laptops and SMDs given that laptops have a lower probabilities of moving throughout all the IPL. However, Fig. 14b contradicts the 1 day results. Since the difference between both traces is restricted to the minimal speed of travel (which in Fig. 14a) must be above 0.5m/s), it is safe to assume that the abnormal behaviour is due to the speed at which the devices travelled such long distances. In general, jump size results tend to support the claim that SMDs have a higher mobility, which produces larger traces passing through multiple APs, while laptops are disconnected and reconnected at a new location.

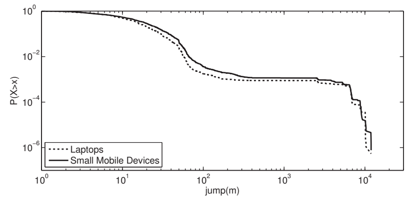
Jump sizes are tightly associated with the physical dispersion of IPL, which creates groups of users that either remain close on one school or travel between several. This issue has been previously identified, for example, in HCMM [5] where scenario generation considers the possibility to set-up a number of groups and create bell shaped normal distributions.

4.2.3 Pause times

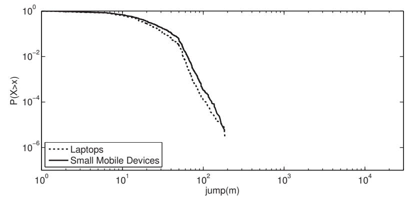
The CCDF of pause times is presented on Fig. 15. It should be noted that pause times are particularly small as the methodology followed for trace definition tend to maintain a node in movement even if at a very small speed. In the results presented on the figure, devices are considered stopped if the distance between way-points is less than 1m. Still, it is interesting to observe that SMDs have different pause time distributions with lower probabilities of having higher values, something that supports the mobility characteristics expected for SMDs.

4.2.4 Inhomogeneity

Table 3 shows results for the inhomogeneity metric. Samples for this metric were obtained at 4 different times on the

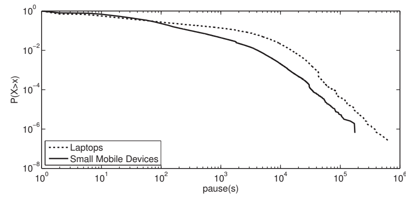


(a) IPL/trace

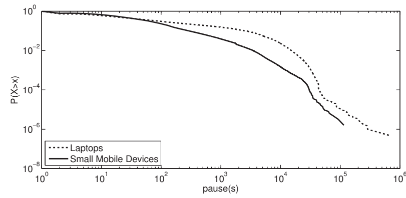


(b) ISEL/trace

Figure 10: Jump size (meters)

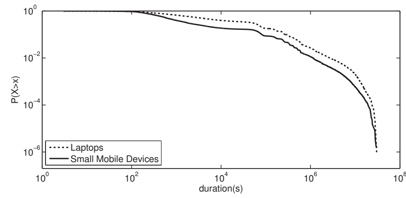


(a) IPL/trace

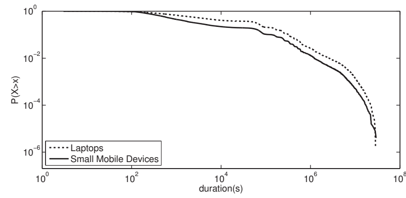


(b) ISEL/trace

Figure 11: Pause times (s)

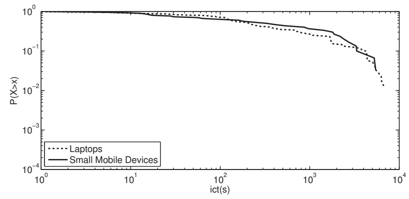


(a) IPL/trace

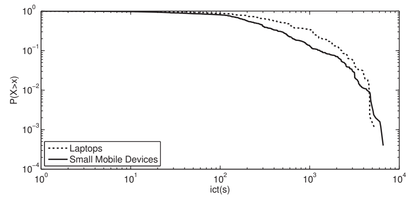


(b) ISEL/trace

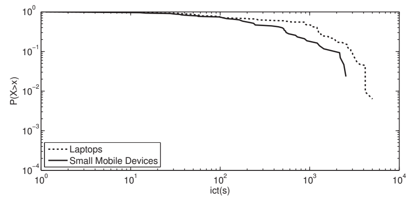
Figure 12: Disconnection time (s)



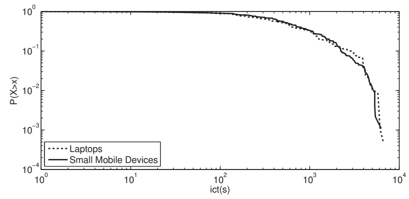
(a) IPL (3 days)



(b) IPL (disconnected)

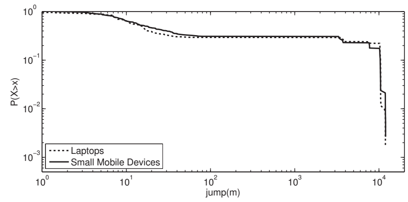


(c) ISEL (3 days)

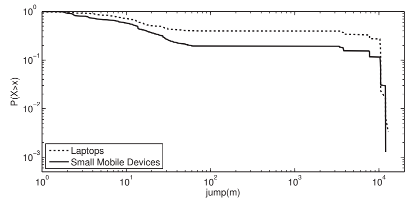


(d) ISEL (disconnected)

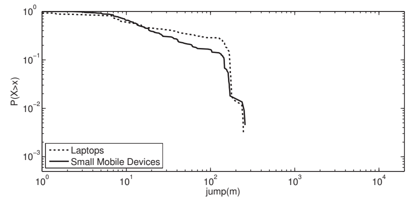
Figure 13: Inter-Contact Times (s)



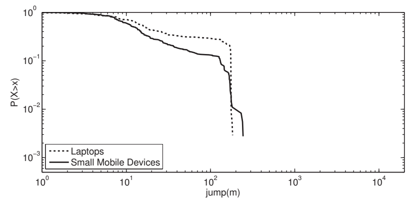
(a) IPL (3 days)



(b) IPL (disconnected)

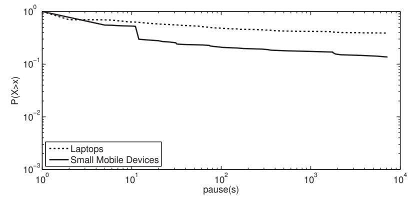


(c) ISEL (3 days)

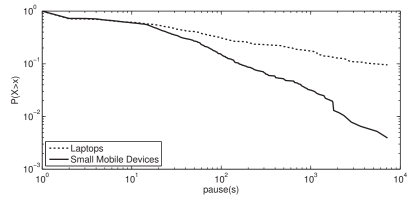


(d) ISEL (disconnected)

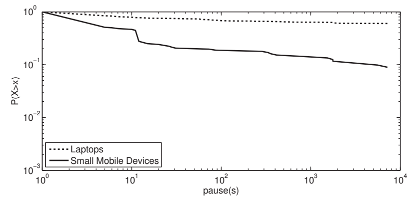
Figure 14: Jump Size (meters)



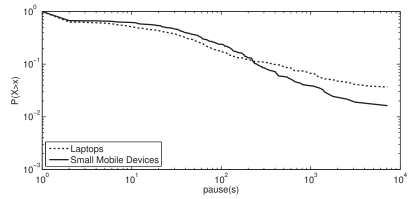
(a) IPL (3 days)



(b) IPL (disconnected)



(c) ISEL (3 days)



(d) ISEL (disconnected)

Figure 15: Pause times (s)

scenarios (at the beginning, 1/3, 2/3 and at the end of the scenario), for the day with the most SMDs present on the network. Laptops show a higher inhomogeneity, indicating a larger concentration and irregular distribution for these devices. Suggests that laptops tend to be grouped, for example in classrooms or libraries. The lower value of inhomogeneity for SMDs confirms the pseudo-random disposition for mobile devices. Again, the data that excludes trajectories that were interrupted, during the analysis period, produces some incongruence with what would be expected, as for example

Table 3: Inhomogeneity values

	All (σ)	Laptops (σ)	SMD (σ)
IPL 1-day	0.79 (0.01)	0.9 (0.01)	0.77 (0)
ISEL 1-day	0.59 (0.01)	0.62 (0.02)	0.68 (0.05)
IPL Disc.	0.76 (0)	0.87 (0.01)	0.75 (0)
ISEL Disc.	0.66 (0.03)	0.71 (0.01)	0.56 (0.03)
IPL RWP	0.4 (0.05)		
ISEL RWP	0.35 (0.06)		
IPL SLAW	0.63 (0.02)		
ISEL SLAW	0.52 (0.05)		

Table 4: Akaike test results

	SLAW	Dartmouth	MobIPLity		
	All	All	All	Laptops	SMDs
Duration	-	-	LN	W	LN
Dur./dev	-	-	G	G	P
Length	-	-	P	P	P
Len./dev	-	-	P	P	P
Speed	-	-	W	W	P
Speed/dev	-	E	P	P	P
Disconnection	-	-	GEV	LN	GEV
Pause	P	LN	P	P	P
Jump	P	-	P	P	P

(P: Pareto, LN: Log-Normal, G: Gamma, E: Exponential, W: Weibull, GEV: Extreme value)

for IPL scenario.

5. DISCUSSION

When choosing a mobility model to evaluate ubiquitous applications, a developer is faced with a number of parameters that allow the evaluation to be adapted to the environment where the application is expected to be deployed. Having a source of real mobility data allows an easier transition from development to evaluation. In spite of the limitations dictated by the specific academic scenario where the data was collected, it is possible to create multiple scenarios if we consider the distinct schools of IPL. For these multiple scenarios, there is a number of metrics that are scenario agnostic (namely trace duration, trace length, pause times, disconnection time and ICT) where differences observed between ISEL and IPL are minimal. On the opposite side we have scenario dependent metrics (trace speed, jump size and inhomogeneity), where metrics evidence the differences between scenarios. Besides the changes on the geographical

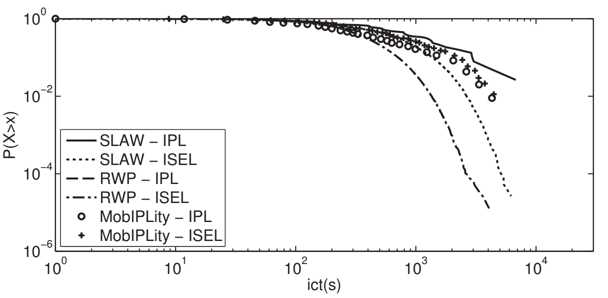


Figure 16: Comparison with ICT for SLAW and RWP

aspect of the scenario, metrics are also influenced by device type which by itself also influences user behaviour. Consider for example the increased mobility of an SMD that allows the user to carry it while connected to the WiFi network and its impact on metrics. We expect that scenario agnostic metrics may also portrait the characteristics of other scenarios, however further research is needed to support this claim.

To facilitate the comparison with trace-based mobility models, we compare MobIPLity with SLAW [15] and RWP using ICT (a scenario agnostic metric) and Inhomogeneity (a scenario dependent metric).

To compare ICTs, we used the same setup and data from Sec. 4.2.1, considering all devices, and created similar scenarios for SLAW and RWP to emulate ISEL and IPL. The CCDF of ICTs is depicted on Fig. 16. RWP for IPL distributes the nodes so homogeneously that prevents generation of long ICTs, thus limiting the visibility of the data on the figure. Despite setting SLAW to emulate IPL on the number of access points, SLAW has limitations on representing the same ICTs as MobIPLity, which by itself presents similar ICTs for IPL and ISEL. Unfortunately, neither SLAW or RWP distinguish different device types although our results show that the device type plays a significant role on ICT, with SMDs typically presenting shorter ICTs, what is attributed to their higher mobility.

The inhomogeneity metric, depicted in Table 3, was calculated for scenarios synthetically generated by SLAW and RWP that replicate the conditions found in IPL and ISEL (dimension, duration, devices and number of hotspots/access points). Results of our mobility records are similar to the ones obtained by SLAW, and as expected, diverge from the randomness found in RWP where the metric value is low.

To better understand the differences between metrics, mobility models and device types, the Akaike test was used to compare the fitness of ISEL traces for the complete year of 2012 to well-known distributions. ISEL traces were chosen due to their containment in a single campus, thus better reflecting environments found in the literature. Results presented in Table 4 were obtained using Matlab automatic fitting between all supported distributions, sorted by Akaike criteria. It's worth noting the difference between fitting results for different device types, where for the vast majority of SMD metrics Pareto is chosen as the best fit. For Laptops there are more differences between metrics. This confirms our suspicions on the existence of different mobility models for these devices. The table shows the distributions used by SLAW and Dartmouth [14]. The difference in chosen statistical models for pause times are evident. Pause time calculation methodology has impact on the fitting process. In [14], pause times were defined by detecting users walking at a low speed. The paper claims that pause times have a log-normal distribution, something that is not supported by our observations which show a power law distribution.

6. CONCLUSIONS AND FUTURE WORK

The availability of real world mobility traces contributes to the quality of the evaluation and the predictability of new protocols, applications and algorithms. This paper presented a data set composed of the access records to the eduroam network of the Polytechnic Institute of Lisbon between 2005 and 2012 and a methodology and web site to extract mobility scenarios from this set.

The recency of the data set permits to observe the most recent pattern changes on mobility, that result from the increasing popularity of small dimension devices, effectively increasing user mobility. Interestingly, evaluation showed that power saving mechanisms that are standard on these devices reduce the possibility of spontaneous communication between devices and with the environment. This is an aspect that has been neglected by both mobility models and network simulators and left for applications developers to address.

We hope that our effort in providing a public web interface and a set of available metrics will allow developers to rapidly create a scenario in a common, widely adopted format, that could be used in application or framework evaluation using simulation.

Analysis of the data continues. As future work, authors plan to investigate users contact patterns and the changes in the use of the network and devices with time.

7. REFERENCES

[1] N. Aschenbruck, R. Ernst, E. Gerhards-Padilla, and M. Schwamborn. Bonnmotion: A mobility scenario generation and analysis tool. In *Procs. of the 3rd Int'l ICST Conf. on Simulation Tools and Techniques*, 2010.

[2] N. Aschenbruck, E. Gerhards-Padilla, M. Gerharz, M. Frank, and P. Martini. Modelling mobility in disaster area scenarios. In *Procs. of the 10th ACM Symp. on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, 2007.

[3] N. Aschenbruck, A. Munjal, and T. Camp. Trace-based mobility modeling for multi-hop wireless networks. *Computer Communications*, 34(6):704 – 714, 2011.

[4] C. Bettstetter, G. Resta, and P. Santi. The node distribution of the random waypoint mobility model for wireless ad hoc networks. *Mobile Computing, IEEE Transactions on*, 2:257–269, 2003.

[5] C. Boldrini and A. Passarella. Hcmm: Modelling spatial and temporal properties of human mobility driven by users' social relationships. *Computer Communications*, 33:1056 – 1074, 2010.

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

[7] N. Cruz, H. Miranda, and P. Ribeiro. The evolution of user mobility on the eduroam network. In *IEEE Int'l Conf. on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2014.

[8] D. Fischer, K. Herrmann, and K. Rothermel. Gesomo: A general social mobility model for delay tolerant networks. In *Mobile Adhoc and Sensor Systems (MASS), 2010 IEEE 7th Int'l Conf. on*, 2010.

[9] M. Gyarmati, U. Schilcher, G. Brandner, C. Bettstetter, Y. W. Chung, and Y.-H. Kim. Impact of random mobility on the inhomogeneity of spatial distributions. In *IEEE Global Telecommunications Conf. (GLOBECOM)*, 2008.

[10] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. *Computer Networks*, 52:2690 – 2712, 2008.

[11] D. B. Johnson, D. A. Maltz, J. Broch, et al. DSR: the dynamic source routing protocol for multi-hop wireless ad hoc networks. *Ad hoc networking*, 5:139–172, 2001.

[12] T. Karagiannis, J. Y. Le Boudec, and M. Vojnovic. Power law and exponential decay of intercontact times between mobile devices. *Mobile Computing, IEEE Transactions on*, 9:1377–1390, 2010.

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

[14] M. Kim, D. Kotz, and S. Kim. Extracting a mobility model from real user traces. In *INFOCOM 2006. 25th IEEE Int'l Conf. on Computer Communications. Procs.*, 2006.

[15] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. Slaw: A new mobility model for human walks. In *INFOCOM 2009, IEEE*, 2009.

[16] M. Mulhanga, S. Lima, and P. Carvalho. Characterising university wlans within eduroam context. In *Smart Spaces and Next Generation Wired/Wireless Networking*. Springer, 2011.

[17] C. Perkins, E. Belding-Royer, and S. Das. Ad hoc on-demand distance vector (AODV) routing. Request For Comments 3561, IETF, 2003.

[18] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24). Downloaded from http://crawdad.cs.dartmouth.edu, Feb. 2009.

[19] C. Rigney. RADIUS Accounting. RFC 2866 (Informational), June 2000.

[20] U. Schilcher, M. Gyarmati, C. Bettstetter, Y. W. Chung, and Y.-H. Kim. Measuring inhomogeneity in spatial distributions. In *Vehicular Technology Conf., 2008. VTC 2008. IEEE*, 2008.

[21] A. Thiagarajan, L. Ravindranath, H. Balakrishnan, S. Madden, and L. Girod. Accurate, low-energy trajectory mapping for mobile devices. In *Procs. of the 8th USENIX Conf. on Networked Systems Design and Implementation*, 2011.