A Human Connectivity Model for Opportunistic Mobile Systems

Roberta Calegari, Mirco Musolesi, Franco Raimondi and Cecilia Mascolo
Department of Computer Science, University College London
Gower Street, London, WC1E 6BT, United Kingdom
{r.calegari|m.musolesi|f.raimondi|c.mascolo}@cs.ucl.ac.uk

Abstract

Opportunistic networking protocols have recently started to emerge in different contexts, ranging from vehicular communications and remote populations connectivity to wildlife monitoring. These protocols are mainly based on the ability to exploit asynchronous communication among hosts who can act as carriers for the messages which are first stored and transported, and then delivered when the destination is reached. At the heart of these protocols is the concept of hosts colocation and connectivity patterns. Often, however, the protocols are evaluated using mobility models which tend not to mirror the connectivity patterns of the domain in which the protocol needs to be applied, failing to give insight into the performance of the protocols in realistic settings.

In this paper we propose a different approach: based on the assumption that opportunistic networking protocols are based on colocation (and connectivity), we present a model for *connectivity* patterns, which can be extracted from real data. To validate our approach, we show how we used the Dartmouth Campus traces as one of the inputs of our framework to generate connectivity traces with a similar behaviour.

1 Introduction

The recent years have seen a growing interest towards opportunistic networking protocols [7]. The applications of these protocols range from pure delay tolerant networking scenarios for the provision of connectivity in presence of intermittent disconnections or network partitions [24], to information dissemination algorithms for specific scenarios, including vehicular ones [4] and wildlife monitoring [16]. At the heart of many of these protocols is the idea that hosts colocation[1] can be exploited to transfer messages from senders to intermediate nodes, such as mobile carriers, and then, from carriers to final receivers, possibly with some delay. Therefore, *connectivity*, more than *mobility*, is one of the pre-eminent aspects to be considered in the design and performance evaluation studies of this class of systems and protocols[2].

Existing mobility models generate random movement traces like the Random Way Point model [11], with no insight into realistic connectivity patterns. However, there is a stringent need of more realistic and sound connectivity patterns for testing mobile systems in the community: for this reason, many research groups have started projects with the aim of collecting traces for different application scenarios, including students patterns in campuses [8, 27], people attending conferences [10] and cities and streets circulation [22]. Repositories have also been created to collect all these measurements (e.g., CRAWDAD Project [12] at Dartmouth College).

No matter how many traces can be collected, this will always look like a small amount, in many case insufficient, with respect to the variation needed for a thorough analysis of the performance of a system. In addition, a *sensitivity analysis simply cannot be performed* using a single set of traces. A number of pioneering works [1, 2, 8, 28] have studied traces in order to gain a better understanding of the real mobility patterns. A key study in this area is the work on connectivity patterns presented by Chaintreau et alii in [6] which illustrates the fundamental insight that contacts duration and inter-contacts time between individuals are distributed according to power-law distributions[3] and that these patterns may be used to develop more efficient opportunistic protocols.

What we have seen until now is that traces are generally used *a posteriori* to validate the model and not for *a priori* study of the connectivity properties. For example in a previous work [21], we developed a model based on social networks able to reproduce patterns observable in real traces, especially in terms of colocation duration and inter-

[1]For the purposes of this work, we speak indistinctly of colocation, contact and connectivity, as they are equivalent in our model.

[2]Clearly, mobility models generating geographical coordinates are necessary to test location-based systems and geo-routing protocols.

[3]Power-law distributions are characterised by the following form:

$$P(x) = x^{-k}$$

with $k \geq 0$.

A power-law distribution is also called scale-free since it remains unchanged to within a multiplicative factor under a re-scaling of the independent variable x [23].

contacts time. However, the real traces are used to validate emergent properties of the synthetic movements and not as input of the model itself. Many research projects focused on the problem of studying the transition between different geographical areas like the models presented in [14,29]. Yoon et alii [31] presented a model extracted from real traces of probability of transitions between locations of the Dartmouth College campus. Connectivity patterns are, again, emergent properties of the model, rather than inputs of the simulation. Moreover, the evaluation of the model is based on the matching of the geographical movements and density of users, rather than on the analysis of the patterns of the hosts connectivity.

In this paper we present a novel model of connectivity. This mathematical model is then used to build a trace generator for the analysis and design of opportunistic systems. Probability distributions describing the patterns of colocation of mobile users are exploited for the first time as *direct inputs* of the generator. The distribution of the average number of people that an individual meets during a certain period of time (e.g., a day) is also an input of the model. All these distributions can be extracted by measurement of connectivity on real traces, or can be defined by researchers to study interesting or limit cases (e.g., scenarios characterised by long disconnections).

In addition to these contributions we note that, despite the model having been applied to human contacts, it is general enough to be applied beyond human patterns. Wildlife traces could be incorporated in the model, once their general distributions are studied and understood. We expect that, for some animals, the complete approach presented in this paper would be applicable out of the box, while for some others, e.g., for animal moving in large groups, different connectivity distributions need to be used.

The rest of the paper is organised as follows: Section 2 contains a brief description of the key steps of our approach. Section 3 describes the connectivity model; Section 4 illustrates the application of the model to the Dartmouth traces; Section 5 shows how we automatically generate synthetic connectivity traces which mirror the connectivity patterns of the original traces. Section 6 contains a comparison of our work with the state of the art in this research area, discussing the main contribution of this work and its possible applications. Finally, Section 7 concludes the paper, illustrating possible future work.

2 The Approach at a Glance

In this section we briefly outline the various steps of our approach.

- **Connectivity Model.** We have firstly defined and studied a mathematical model of host connectivity. This aims at representing the properties of the *colocation* of two users as a function of the probability for a user of being at a certain place for a given time. We will refer to this duration as *residence time* in the reminder of this paper. The model is described in Section 3.

- **Derivation of Connectivity Distributions from Real Traces.** In order to determine the inputs of the model we have studied the log session traces of the campus WLAN of Dartmouth College [13], to obtain empirical distributions for residence time, colocation (i.e., contact time) and other variables. The traces were collected by researchers at Dartmouth College from April 2001 to June 2004. The network is composed of 450 access points over an area of about 200 acres. The total number of users logged in these traces is 13889. We have used the measured distributions in the connectivity model to determine the relationship between the distribution of user residence time and user colocation. This analysis is presented in Section 4.

- **Synthetic Trace Generation.** Once the analysis of the real traces is completed we have all the necessary information to start the process of generating synthetic traces. More specifically, we use the parameters of the contacts duration and inter-contact time and the graph of human contacts, which are extracted from the traces, to generate the synthetic traces.

  The process of generation is based on the selection of the desired number of hosts for the synthetic traces and on the construction of a connectivity graph of all the potential contacts of each host. In other words, we map each host to a node of the graph and we link a pair of nodes with an edge if the two hosts have a potential of becoming in contact: the distribution of all potential contacts is another input of the model and is also extracted from the traces. The connectivity graph is then used to unfold a number of connection links between users for each time instant. In other words, we use the connectivity graph as a basis for a *time-varying graph of instant connectivity* for each instant t. In this time-varying graph, each link is either *active* if the two hosts are colocated, or is not present if the two are not.

  A description of our implementation is presented in Section 5.

A possible objection to our approach is that the traces we have used are collected from an infrastructure-based network and not by pure connectivity-oriented technologies such as Bluetooth or 802.11 ad hoc. In response to this, we notice that our analysis is perfectly valid on both types of traces as we only use information related to the simultaneous presence of two hosts in the same location. In our analysis, we assume that colocation implies connectivity. We have used these traces simply because they were many and very complete, but the model is totally independent from which traces are chosen. In addition, we note that all the available human traces show the same kind of distributions [6, 10, 19, 27].

3 The Connectivity Model

In this section we present the details of a methodology to compute the probability distribution of colocation (i.e., connectivity) times between two users, starting from a minimal data set (see below). The definition of our model is based on some simplifying assumptions:

- users' behaviours are *independent*; this means that we assume the behaviour of a user does not depend on other users' behaviours.
- users' behaviours are *uniform*: all users have the same behaviour.

These assumptions are sufficient to capture the real connectivity patterns with great accuracy (see Section 5). A refined model could take into account users' degree of correlation in order to model non-uniform or non-independent users' behaviour.

We denote with X and Y two random variables for the duration of the sessions of two generic users a and b, respectively. The probability that a user a will remain in a given location for a time t (i.e., the residence time) is given, under our assumptions, by a probability density function $p_X(t)$; all users' behaviour is described by the same distribution $p_X(t)$. $p_X(t)$ is interpreted as the probability that the residence time will last t seconds.

In addition to the distribution $p_X(t)$, we assume that a probability density function $p_R(t)$ is available, representing the probability that the temporal distance between the beginning of two sessions of two colocated users is t (see Figure 1: t represents the "delay" of one session with respect to another).

Our aim is to compute a probability density function $p_C(t)$, representing the probability that the colocation (i.e., contact) between any two users a and b lasts t. Without loss of generality, we assume that a's session starts before b's session (the other case is symmetrical). If a and b are colocated (i.e., in contact), then only two cases can occur (see Figure 1):

1) b starts with a delay R and terminates after a (i.e., the two sessions overlap),
2) b starts with a delay R and terminates before a (i.e., b's session is contained in a's session).

The probability of occurrence of case 1) is given by the probability that Y is more than $X - R$, which we write as $p(Y > X - R)$. Analogously, the probability of occurrence of case 2) is given by the probability that Y is less than $X - R$, written as $p(Y \leq X - R)$. Overall, case 1) and 2) contribute to $p_C(t)$ as follows:

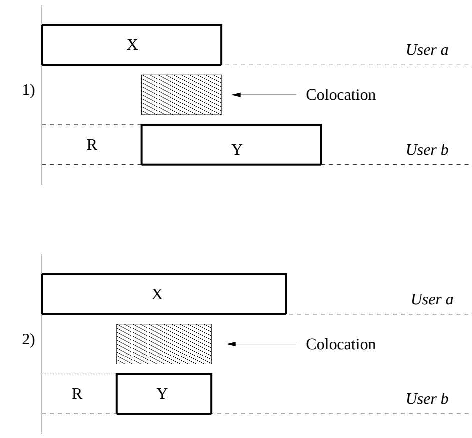

Figure 1: Connectivity cases.

$$p_C(t) = p(Y > X - R)p_{X-R}(t) + p(Y \leq X - R)p_Y(t) \tag{1}$$

where $p_{X-R}(t)$ represents the probability that $X - R$ lasts t. As mentioned above, under our assumptions users are characterised by the same behaviour, therefore, for all t, we have $p_X(t) = p_Y(t)$, and X and R are two independent random variables; thus, we can write

$$p_{X-R}(t) = \int_0^{+\infty} p_X(t+r)p_R(r)dr \tag{2}$$

Intuitively, Equation (2) states that $X - R$ lasts t if X lasts $t+r$ and R lasts r, integrated over all possible delays r from 0 to $+\infty$.

We evaluate now the term $p(Y > X - R) = p(X - R < Y)$. Notice that this is a number and represents the weight of p_{X-R} in Equation (1). For a fixed y, we have

$$p(X - R < y) = \int_0^{y} p_{X-R}(k)dk \tag{3}$$

and therefore

$$p(X - R < Y) = \int_0^{+\infty} \left(\int_0^{y} p_{X-R}(k)dk \right) dy \tag{4}$$

Taking into account Equations (1) and (2), we can rewrite $p_C(t)$ in terms of the known functions $p_X(t)$ and $p_R(t)$, as follows:

$$p_C(t) = \chi \int_0^{+\infty} p_X(t+r)p_R(r)dr + (1-\chi)p_X(t) \tag{5}$$

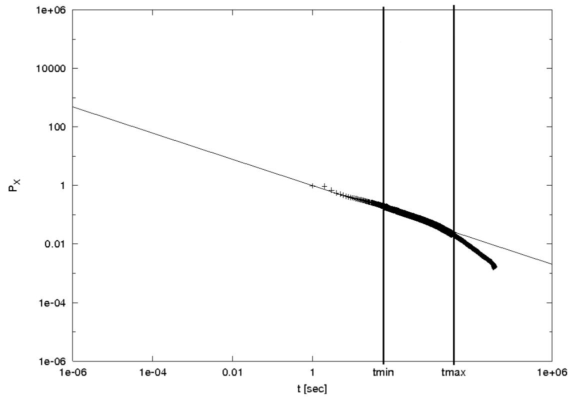

Figure 2: Distribution of residence time in Academic Building 22, log-log scale (all users over 4 years).

where $\chi = p(X - R < Y)$ is defined by Equation (4).

To summarise, we have obtained a formula that allows us to calculate the probability distribution of the colocation (i.e., contact) duration of users in a place as a function of their residence time in that location and the arrival delay. We provide a concrete example of the computation of these functions in the next section, where we instantiate all the parameters using data obtained from real traces for human connectivity.

4 Computing the connectivity distribution from real traces

In the next subsection we analyse traces from [13] to derive $p_X(t)$ and $p_R(t)$. In Section 4.2 we compute $p_C(t)$ using the process presented above; in Section 4.3 we compare the computed $p_C(t)$ with the observed values of colocation from the real traces. We note that our purpose is the generation of traces with probabilities similar to the *average* probabilities over the Datmouth traces *across the whole period of time*. A similar approach can however be applied for the generation of traces of just a specific month period (e.g., April).

4.1 Analysis of real traces

We consider a selection of traces from [13], from 01/04/2001 until 30/06/2004. These traces record connections and disconnections of users at a number of access points in the Dartmouth campus; in particular, the data available include MAC addresses, locations of access, and timestamps. In the traces analysed we found 13889 different users and 178 different locations.

As an example, Figure 2 reports the cumulative distribution of residence times for all users in Academic Building 22 in a log-log scale. For any given duration t, the value on the y axis gives the probability that the session of a user lasts t or more seconds.

As previously observed in a number of works (see for instance [6]), the distribution of the residence time at a given access point follows a power law in a range of values[4], denoted by $[t_{min}, t_{max}]$ in our paper (points between the vertical bars $x = t_{min}$ and $x = t_{max}$ in Figure 2). In the traces analysed we found $t_{min} = 60$ and $t_{max} = 13397$. Figure 2 also reports the interpolated curve to obtain the coefficient for the cumulative distribution $P_X(t)$, from which the actual coefficient k_X of $p_X(t)$ can be computed (see the the straight line in Figure 2).

For the purposes of our automatic trace generation (see Section 5), we extract another parameter: the distribution of the time elapsed between two contacts of the same pair of users. This is called the *inter-contact time* and its distribution is denoted by $p_{IC}(t)$. We proceeded similarly to $P_X(t)$ and, as expected [6], we found a power law distribution with coefficient k_{IC}. Analogously, we evaluated the distribution $p_R(t)$ and we found a power law distribution with coefficient k_R. The interpolated coefficients for the sample location are: $-k_X = -1.448$, $-k_{IC} = -1.745$, and $-k_R = -1.062$. Due to space limitations, we do not include graphs and coefficients for all locations; these are available from the authors upon request.

4.2 Computing the probability distribution for colocation

We now compute $p_C(t)$ using the model presented in Section 3, in terms of its cumulation $P_C(t)$. We start from the distributions $P_X(t)$ and $P_R(t)$ obtained in the previous section and we compute a (theoretical) $P_C(t)$, as defined by Equation (5); finally, we compare this distribution with the experimental results obtained.

We assume that $P_X(t)$ and $P_R(t)$ are defined in the range (t_{min}, t_{max}) by the following power law distributions:

$$P_X(t) = t^{-k_X}$$
$$P_R(t) = t^{-k_R}$$

We substitute these distributions in Equation (2) to obtain

$$P_{X-R}(t) = \int_{t_{min}}^{t_{max}} (t+r)^{-k_X} r^{-k_R} dr \qquad (6)$$

Equation (6) can be resolved analytically using, for instance, the software tool Mathematica, obtaining

$$P_{X-R}(t) \approx \Big(-r^{-k_R}(t+r)^{-(k_X+1)} \left(\frac{r}{t}+1\right)^{k_X+1} \cdot {}_2F_1\left(-k_R, k_X+1, 1-k_R, -\frac{r}{t}\right) \Big)\Big|_{r=60}^{r=tmax} \qquad (7)$$

[4]Notice that if a probability density function is a power law of the form $f(x) = x^{-\alpha}$, then its cumulative distribution is a power law with coefficient $-(\alpha - 1)$.

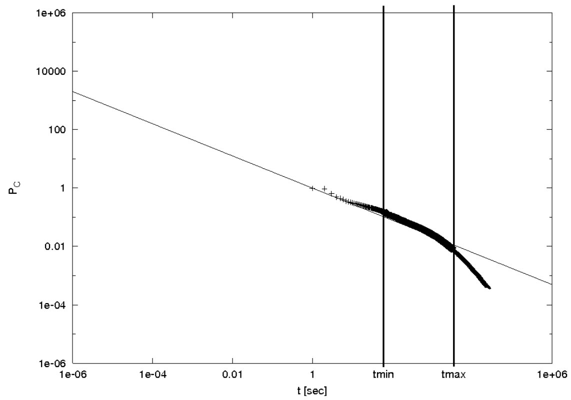

Figure 3: Distribution of colocation time in Academic Building 22, log-log scale (all users over 4 years).

where ${}_2F_1$ is the hyper-geometric function. Using the representation of ${}_2F_1$ in terms of a convergent series (see Appendix A), we can evaluate the behaviour of the previous integral, namely

$$P_{X-R}(t) \approx O(t^{-1})|_{r=60}^{r=tmax}. \quad (8)$$

Mathematica can be used to compute the exact value of χ, as defined in Equation (4). This value, together with the result of Equation (8), imply that Equation (5) can be approximated by

$$p_C(t) \approx \chi O(t^{-1}) + (1-\chi)t^{-k_X} \approx t^{-k_X} \quad (9)$$

for our values of k_X and for values of t greater than 1 second (which is clearly the case in our scenario where $t_{min} = 60$).

Therefore, by using the distributions $p_X(t)$ and $p_R(t)$ from [13], our theoretical model forecasts a distribution of the colocation with the same behaviour of $p_X(t)$. Notice, however, that this particular result only applies to power law distributions whose coefficient are within a specific range of values, as in the case of the traces analysed here. In the generic case, it might be that the distribution of colocation follows a distribution different from $p_X(t)$. Nevertheless, in this scenario, our result is validated by the analysis of the traces, as described in the next section.

4.3 Colocation results from real traces

We then computed the actual distribution of colocation time for two generic users. The distribution $p_C(t)$ of the duration of colocation obtained form the traces follows a power law. As in Section 4.1, we interpolate the distribution to obtain the coefficient of the power law. An example location is reported in Figure 3. The coefficients for the location described (Academic building 22) is $-k_C = -1.551$ and the average coefficient over all locations is $-k_C = -1.327$

Since the results presented above are averaged over four years, we repeated the process of computing $p_C(t)$ but for a time window of 8 hours only (from 9am to 5pm), averaged over a period of one month in the middle of an academic term (from 19/04/2004 to 19/04/2004). We performed these measures to rule out the possibility that, at a smaller time scale, the behaviour of the distribution of colocation could be different.

Additionally, we interpolated the distribution of inter-contact time for the same time windows, and, again, we obtained a power law. The interpolated values for colocation and inter-contact coefficients for Academic Building 18 are $-k_C = -1.430$ and $-k_{IC} = -1.420$

5 Synthetic Trace Generation

We now show how the connectivity model presented in Section 3 can be used to implement a generic trace generator able to produce traces characterised by given connectivity properties and network of potential contacts. The traces generated by our tool abstract away from spatial movements and only concentrate on connectivity and inter-contact times: in this sense, the tool differs from generators like the *setdest* tool (which implements the Random Way Point mobility model for ns-2 [18]).

The basic idea is to allow the generation of traces of arbitrary time length with connectivity patterns mirroring the ones of the original traces (in our specific example, the Dartmouth traces) and involving a user-defined number of nodes. Below, we give an example of trace generation. In Figures 4, 5, and 6 we show how this mirrors the various patterns of the original traces.

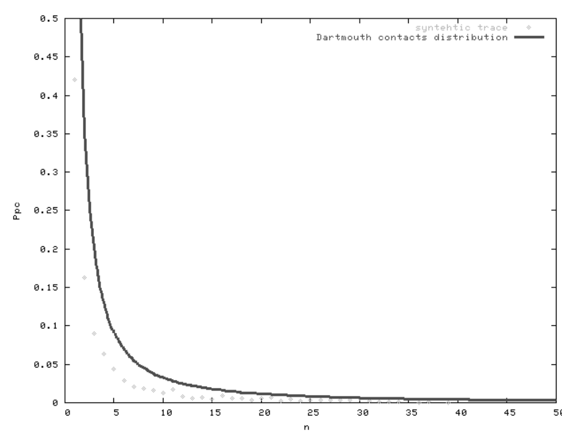

Figure 4: Comparison between synthetic trace (lighter) and real power-law distribution using Dartmouth potential contact coefficient (darker).

5.1 Generation of the Potential Contacts Graph

The first step is the generation of the *Potential Contacts Graph*. Let us assume that we want to generate traces for a simulation time of 8 hours and for 200 users, mirroring the behaviour of Dartmouth traces in the period 19 April to 19 May 2004, a period without holiday days, considering the hours from 9am to 5pm. Each of the 200 hosts is mapped to a vertex of this graph. An edge between two vertexes exists if a potential contact is possible during the total desired simulation time. This means that, in the case of our example, an edge between two vertexes A and B exists if and only if the individuals have a chance of being colocated at least once during the period of the eight hours.

In the real traces for the period considered, we found 1892 users and a maximum number of potential contacts for an individual equal to 42 (i.e., the individual with the maximum number of potential contacts had 42 contacts). The distribution of the number of contacts is *the distribution of the degrees of the vertexes* in the Potential Contact Graph. In order to calculate that in an appropriate manner, we have interpolated the distribution of the degrees for the traces, and obtained a power law in the range up to 42 individuals. We denote this distribution by $p_{pc}(n) = n^{-k_{pc}}$: $p_{pc}(n)$ gives the probability that a node has degree n. The measured value for $-k_{pc}$ is -1.484. In order to be able to generate traces with similar patterns, which may be used in the evaluation of opportunistic protocols, we need to obtain a distribution for the 200 users, instead of 1892. This is done through geometric similarity over the frequency graph of the degrees of vertexes. Intuitively, the maximum number of contacts scales up or down proportionally to the square root of the ratio of the total number of vertexes in the graph. Thus, the maximum number of contacts for a single user when 200 users are present is computed to be 13, and we take the same coefficient for the power law distribution. This enable us to generate a sequence of random degrees following the desired distribution.

We have implemented a procedure to generate a graph given its number of vertexes and their degree distribution, within a certain approximation. Intuitively, the procedures non-deterministically tries to build a graph with the desired properties, and iteratively refines the solution up to the desired approximation level. We set the approximation to in the range $[2.5\%, 4\%]$ depending on the connectivity coefficients of the graph (with a higher connectivity we selected a larger margin of error). Figure 4 shows a comparison between the theoretical degree distribution (dark line) and the generated degree distribution using our tool (lighter marks).

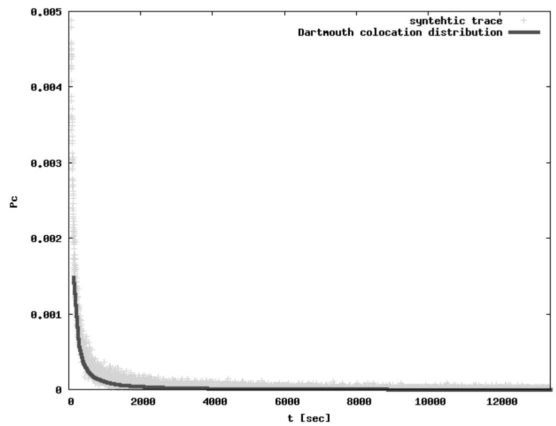

Figure 5: Comparison between synthetic trace (lighter) and real power-law distribution using Dartmouth colocation coefficient (darker).

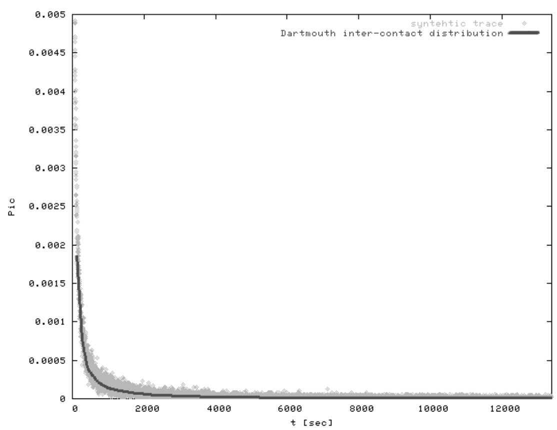

Figure 6: Comparison between synthetic trace (lighter) and real power-law distribution using Dartmouth inter-contact coefficient (darker).

5.2 Generation of the Instant Snapshot Contact Graphs

Once the potential contact graph is generated, the actual connectivity traces can be produced as a sequence of *instant snapshot contact graphs*, one for each instant of time. A connectivity graph for time t represents the network of connected vertexes at time t. The instant snapshot contact graphs are generated as follows: The first connectivity graph is generated from the potential contact graph assuming that each potential edge is non active (i.e., no vertexes are connected and the edge is in an "off" state). Then, the distribution of inter-contact times $p_{IC}(t)$ is used to assign a duration to the "off" time of each edge. When this time has elapsed, the distribution of connectivity time $p_C(t)$ is used to assign a duration to the "on" time of each edge. Thus, for each edge there is a sequence of durations off/on, distributed following $p_{IC}(t)$ and $p_C(t)$. A sequence of instant snapshot contact graphs is obtained by looking at each edge. Finally, the instant snapshot contact graphs are appropriately parsed to generate input traces for Omnet++.This last step provides the desired synthetic traces following the input distributions $p_{pc}(t)$, $p_C(t)$, and $p_{IC}(t)$.

Figure 5 and 6 show a comparison between the cumulative functions for colocation and inter-contact time: the darker lines are the theoretical distributions, while the lighter marks are the distributions obtained from the synthetic traces.

6 Related Work and Discussion

Our approach highlights connectivity over mobility and location in the definition of models. A lot of work has been carried out in the area of novel and realistic mobility models in recent years. There has also been work in terms of analysis of properties of the traces, which have concentrated on determining characteristics of human connectivity and which are seminal to this work.

A comprehensive review of the most popular mobility models used by the research community can be found in [5]. Existing models generate movements in the simulation space towards randomly selected goals [3, 11]. In recent years, many researchers have tried to refine existing models in order to make them more realistic by studying the available mobility traces [12]. Various measurement studies have been conducted both in infrastructure-based and infrastructure-less environments. Extensive measurements about the usage of a WLANs have been conducted, for instance, by Tang and Baker in [28], Balachandran et alii in [1] and by Balazinska and Castro in [2]. A detailed analysis of the usage of the WLAN of the Dartmouth Campus College is presented in [8].

Existing mobility models based on real traces are based on the concept on the probability of transitions between different geographical points. Tuduce and Gross in [29] present a model based on real data from the campus wireless LAN at ETH in Zurich. Similarly to what we observed, in their case, too, the session duration data follows a power law distribution. This approach can be seen as a refined version of the Weighted Way-Point Mobility Model [9], based on the probability of moving between different areas of a campus using a Markov model. In [17], the authors try to reproduce the movements of pedestrians in downtown Osaka by analysing the characteristics of the crowd in subsequent instants of time and maps of the city using an empirical methodology. In general, the main goal of these works is to try to reproduce the specific scenarios with a high degree of accuracy.

In a previous work [21], we presented a mobility model based on social networks theory that is able to reproduce the typical power-law distribution, similar to the ones observed in real traces. However, in that model, the probabilities of transitions between different areas of the simulation space are based on a pre-defined synthetic social network of the individuals carrying the mobile devices. Additionally, these distributions are not inputs of the model, rather, they are emergent characteristics. In other words, the observed patterns are confirmed by real measurements, but are not derived from them, as it is the case in the proposed model implemented by this paper.

Lelescu et alii in [14] proposed Model T++, based on a study on the correlation between the number of sessions per access point and the time spent at each location. Similarly to ours, their model is extracted and validated using the traces of movements inside the Dartmouth College campus. However, in this study the focus is on the probability of transitions between different locations and not on the characteristics of the connectivity of the hosts over time that are central for the design of opportunistic protocols.

Yoon et alii [31] present a model extracted from real traces based on the study of probability of transitions between different locations. Connectivity is one of the emergent properties of the model, rather than an input of the patterns generator.

In terms of more analytical work, a key study in the field is [6] by Chaintreau et alii, where the authors analyse the distribution of inter-contacts time and the duration of contacts considering different data sets from various measurements exercises. All of these exhibit a similar heavy tail distribution that can be approximated using a power-law function over a large range of values. The work confirms the results from other studies conducted at Dartmouth [8], UCSD [19], the University of Toronto [27] and the National University of Singapore [26]. At the same time, these observed patterns are at odds with the ones that can be extracted from random mobility models that show an exponential decay [25]. In a previous work [10], similar connectivity patterns have also been observed among the participants of INFOCOM'05.

7 Conclusions

In this paper we have presented a connectivity model based on probability distributions for residence time of individuals ($p_X(t)$), the distribution of time intervals between connections ($p_{IC}(t)$), and a distributions of delays in overlapping connections ($p_R(t)$). These distributions can be easily obtained from real traces, and we have presented an example of this in Section 4, where human mobility traces from the Dartmouth College have been analysed. These parameters allow for the concrete computation of the connectivity distribution $p_C(t)$ using our theoretical model. In parallel, we extracted the same distribution from the traces, and we have been able to validate the model against the measured distribution of connectivity.

Building upon the characteristics of these distributions, a tool for the generation of traces has been presented, which uses the properties of real traces and generates synthetic traces with similar connectivity patters. The generator can be used for testing opportunistic protocols such as PRoPHET [15] abd CAR [20], using network simulators like ns-2 [18] or Omnet++ [30].

Acknowledgments We thank Vanessa Didelez, Mike Lonergan, Don Small, Tara Small, for their help with the mathematical and statistical aspects of the model. We would like to acknowledge the support of EPSRC through projects CREAM and UBIVAL and the European Union through projects PLASTIC and RUNES.

References

[1] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan. Characterizing user behavior and network performance in a public wireless LAN. In *Proceedings of SIGMETRICS'02*, pages 195–205, New York, NY, USA, 2002. ACM Press.

[2] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *1st International Conference on Mobile Systems, Applications, and Services (MobiSys'03)*, San Francisco, CA, May 2003.

[3] J.-Y. L. Boudec and M. Vojnovic. Perfect simulation and stationarity of a class of mobility models. In *Proceedings of INFOCOM'05*, pages 2743–2754, March 2005.

[4] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine. MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks. In *Proceedings of INFOCOM'06*, April 2006.

[5] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communication and Mobile Computing Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2(5):483–502, 2002.

[6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proceedings of INFOCOM'06*, Barcelona, Spain, April 2006.

[7] K. Fall. A delay-tolerant network architecture for challenged internets. In *Proceedings of SIGCOMM'03*, August 2003.

[8] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. In *Proceedings of MOBICOM'04*, pages 187–201, New York, NY, USA, 2004. ACM Press.

[9] W. Hsu, K. Merchant, H. Shu, C. Hsu, and A. Helmy. Weighted Waypoint Mobility Model and its Impact on Ad Hoc Networks. *ACM Mobile Computer Communications Review (MC2R)*, pages 59–63, January 2005.

[10] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 244–251, New York, NY, USA, 2005. ACM Press.

[11] D. Johnson and D. Maltz. Dynamic source routing in ad hoc wireless networks. In T. Imelinsky and H. Korth, editors, *Mobile Computing*, volume 353, pages 153–181. Kluwer Academic Publishers, 1996.

[12] D. Kotz and T. Henderson. CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth. *IEEE Pervasive Computing*, 4(4):12–14, October-December 2005.

[13] D. Kotz, T. Henderson, and I. Abyzov. CRAWDAD trace dartmouth/campus/movement/01_04 (v. 2005-03-08). Downloaded from http://crawdad.cs.dartmouth.edu/, March 2005.

[14] D. Lelescu, U. C. Kozat, R. Jain, and M. Balakrishnan. Model T++:: an empirical joint space-time registration model. In *Proceedings of MobiHoc '06*, pages 61–72, New York, NY, USA, 2006. ACM Press.

[15] A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in intermittently connected networks. *Mobile Computing and Communications Review*, 7(3), July 03.

[16] T. Liu, C. M. Sadler, P. Zhang, and M. Martonosi. Implementing software on resource-constrained mobile sensors: experiences with Impala and ZebraNet. In *Proceedings of MobiSys '04*, pages 256–269, New York, NY, USA, 2004. ACM Press.

[17] K. Maeda, K. Sato, K. Konishi, A. Yamasaki, A. Uchiyama, H. Yamaguchi, K. Yasumotoy, and T. Higashino. Getting urban pedestrian flow from simple observation: Realistic mobility generation in wireless network simulation. In *Proceedings of MSWiM'05*, pages 151–158, September 2005.

[18] S. McCanne and S. Floyd. ns-2 network simulator. http://www.isi.edu/nsnam/ns/.

[19] M. McNett and G. M. Voelker. Access and Mobility of Wireless PDA User. *Mobile Computing Communications Review*, 9(2):40–55, April 2005.

[20] M. Musolesi, S. Hailes, and C. Mascolo. Adaptive routing for intermittently connected mobile ad hoc networks. In *Proceedings of the 6th International Symposium on a World of Wireless, Mobile, and Multimedia Networks (WoWMoM'05). Taormina, Italy.* IEEE press, June 2005.

[21] M. Musolesi and C. Mascolo. A Community based Mobility Model for Ad Hoc Network Research. In *Proceedings of the 2nd ACM/SIGMOBILE International Workshop on Multi-hop Ad Hoc Networks: from theory to reality (REALMAN'06)*. ACM Press, May 2006.

[22] V. Naumov, R. Baumann, and T. Gross. An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces. In *Proceedings of MobiHoc'06*, pages 108–119, New York, NY, USA, 2006. ACM Press.

[23] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46:323, 2005.

[24] A. Seth, D. Kroeker, M. Zaharia, S. Guo, and S. Keshav. Low-cost communication for rural internet kiosks using mechanical backhaul. In *Proceedings of MOBICOM'06*, September 2006.

[25] G. Sharma and R. R. Mazumdar. Scaling laws for capacity and delay in wireless ad hoc networks with random mobility. In *IEEE International Conference on Communications (ICC'04)*, pages 3869– 3873, June 2004.

[26] V. Srinivasan, M. Motani, and W. T. Ooi. Analysis and implications of student contact patterns derived from campus schedules. In *Proceedings of MOBICOM'06*, pages 86–97, New York, NY, USA, 2006. ACM Press.

[27] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara. User mobility for opportunistic ad-hoc networking. In *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'04)*, pages 41–50, Washington, DC, USA, 2004. IEEE Computer Society.

[28] D. Tang and M. Baker. Analysis of a local-area wireless network. In *Proceedings of MOBICOM'00*, pages 1–10, New York, NY, USA, 2000. ACM Press.

[29] C. Tuduce and T. Gross. A Mobility Model Based on WLAN Traces and its Validation. In *Proceedings of INFOCOM'05*, pages 19–24, March 2005.

[30] A. Vargas. The OMNeT++ discrete event simulation system. In *Proceedings of the European Simulation Multiconference (ESM'2001)*, Prague, June 2001.

[31] J. Yoon, B. D. Noble, M. Liu, and M. Kim. Building realistic mobility models from coarse-grained traces. In *Proceedings of MobiSys'06*, pages 177–190, New York, NY, USA, 2006. ACM Press.