Characterizing Opportunistic Communication with Churn for Crowd-counting

Ljubica Pajevic and Gunnar Karlsson
Laboratory for Communication Networks and Linnaeus Center ACCESS
KTH, Electrical Engineering
Stockholm, Sweden
{ljubica, gk}@kth.se

***Abstract*—In opportunistic networking, characterizing contact patterns between mobile users is essential for assessing feasibility and performance of opportunistic applications. There has been significant efforts in deriving this characterization, based on observations and trace analyses; however, most of the findings arise from studying contact opportunities at large spatial and temporal scales. Moreover, the user population is considered to be constant: no users can join or leave the system. Yet, there are many examples of scenarios which do not fully adhere to the previous assumption and cannot be accurately described at large scales. Urban environments, such as smaller city districts, are characterized by highly dynamic user populations. We believe that scenarios with varying population requires further investigation.**

In this paper, we present a novel modeling approach to study operation of opportunistic applications in scenarios where the population size is subjected to frequent changes, that is, it exhibits *churn*. We also propose an application for estimating the size of a mobile crowd, which we then use to validate our model in four scenarios: a city area, subway station, a conference and a scenario with a synthetic mobility model. We show that the model provides good representations of the investigated scenarios.

***Index Terms*—opportunistic networks, crowd-counting, stochastic differential equations**

I. INTRODUCTION

Opportunistic communication in urban areas has been deemed a promising solution for infrastructure-free location-based services and content sharing. The topic of urban opportunistic networking has resulted in a large body of research, with a majority of studies basing their results on scenarios where a fixed number of mobile users roams inside a closed area, following certain mobility patterns. However, the scale of the geographical space where the communication happens plays a significant role. As the region of interest shrinks—consider a small urban district, rather than an entire city—the assumption on closed population will no longer hold: there will eventually be users arriving to the area, staying inside for a while and leaving. The population of users is no longer constant and such a system is known as a system with open population (*open system* in short), or in the networking parlance, a system with *churn*. Neglecting the effect of churn in specific scenarios can be hazardous.

Understanding contact patterns between humans and modeling these patterns realistically is essential for designing opportunistic communication schemes and evaluating their performance. While the research towards this aim has thrived over the past years, there has been little work considering opportunistic systems with churn. In such systems, the user population is in constant change, thus previously obtained results (e.g. the distribution of inter-contact times) may not be reliable. This study takes a step towards increasing our understanding by characterizing user interactions in such dynamic systems.

To this end, we propose an application for estimating the size of a (mobile) crowd, which operates by epidemically spreading small messages between users. Then we define a stochastic model to study the operation of the application. The model is developed by applying the stochastic differential equation (SDE) modeling approach, and we believe that it can be adapted to a variety of opportunistic use-cases.

Our main contributions are the following:

- We study contact patterns in systems with churn and provide an approximation of the contact rate for the system in the transient state.
- We define and analyze a model for a crowd-counting application and validate our model by means of simulation. We investigate four mobility scenarios which include: realistic simulator-based traces, real-life traces and synthetic mobility.
- Based on the presented modeling approach, we provide an estimation method for the system parameters.

This paper is organized as follows. In section II we describe the crowd-counting application. Section III presents the proposed SDE model, and section IV its empirical validation in various mobility scenarios. In section V we describe the application usage. We review related work in section VI and summarize our most important findings in section VII, giving directions for future work.

II. CROWD-COUNTING APPLICATION

A. Motivation and problem statement

Crowd-counting is a technique used to estimate the number of people in a crowd. At ticketed events and events that take place in an enclosed venue, the precise numbers are readily available through turnstiles that count as the visitors enter, or people counters by the entrance. At unticketed events, for example open-space public gatherings such as festivals,

978-1-4799-8461-9/15/$31.00

crowd-counting is more challenging and less precise. Yet, accurate estimation is vital for planning and responding to emergency situations, e.g. for fire evacuation and to provide safety. Crowd-counting can also be a valuable tool for journalists, political groups and event organizers, since biased (politically or utility-motivated) estimations are often reported.

In this work, we present a novel method for crowd-counting and estimating people flows by using a mobile opportunistic application. We envision that the application would be deployed primarily for measurements at outdoor events; nevertheless, the use of the application is not restricted to outdoor scenarios, and can be similarly deployed for indoor measurements where other, common methods are unavailable (or impracticable).

With our method, we can answer the following:

1. What is the number of people currently residing inside the area, and how does this number vary over time?
2. How long do visitors stay in the area?

The second question can be particularly interesting for event organizers to determine turnout—whether the visitors are staying long enough to be considered participants or just roaming in and out of the area. Further practical uses include estimating departure rates at certain exit points, which can be used to re-route users to less congested exits.

In the next section, we describe the crowd-counting application. Note that in this paper we do not address the technical details of the application.

B. Application description

Assume that the area of interest is bounded and well-defined, e.g. given a set of coordinates of a polygon that confines it. Visitors arrive to the area from several entrance points, spend some time inside and eventually leave. Further, visitors are able to determine their approximate position (e.g. from GPS coordinates or by WiFi triangulation methods). If there were a small number of entrance (and exit) points, measuring the people flows could be done by people-counting devices. However, we assume that this is not the case: the area is not strictly confined by physical barriers (think of a large park, or an open-field festival). Rather, we utilize a certain (small) number of nodes—access points—with the crowd-counting application running. The application is also running in the background on a mobile device (e.g. smartphone) carried by each visitor.

The application uses services of a publish/subscribe middleware [1] to detect proximity of other users and access points by exchanging small beacon messages. Whenever a new user arrives to the area and passes by an access point, which is located at some of the entrance points, the user will receive a *registration message* R_{msg} for that area. This user will be denoted a *primary registered* user. The purpose of the message is to inform users that the area is being monitored, that is, crowd-counting is ongoing. Then, the application uploads to a server user details with the location of the node's entrance point and the time of arrival (e.g. via cellular network). While registered users roam inside the area, they will meet unregistered users and recruit them for the measurement task by forwarding R_{msg}. Upon such event, the unregistered become the *secondary (registered)* users (we will interchangeably use the short and the full denotation). In addition, there will be a number of users who do not meet any registered users during their stay, and eventually leave the area without registering themselves. We label these as *unregistered* users. Registered users, primary and secondary, will continue spreading registration messages, and count the nodes they encounter during their stay—both registered and previously unregistered. Finally, when a registered user leaves the area, the application again uploads information about the user: the exit location, departure time and the measured number of encounters. Optionally, the application can periodically send collected measurements while the user is still in the area. As we will describe later, the encounter information will be used to compute contact rate between users.

III. Model description

A. Measuring node interaction

Before delving into modeling, we define contact metrics.

1. *Contact rate* of a single node in an open system is defined as the total number of contacts that the node established during its stay, normalized by its sojourn time, i.e. the number of contacts per unit time. The mean contact rate of the system, c, is obtained from the individual contact rates measured over a longer time interval. In particular, we denote by c_N the mean contact rate measured in a system with the average population of N nodes.
2. Consider a snapshot of the system when the population equals N. The *total encounter rate* Λ_N is given by $\Lambda_N = \frac{N}{2} c_N$. Alternatively, this rate can be represented via the rate at which two arbitrary nodes come into contact η_N, as $\Lambda_N = \frac{N(N-1)}{2} \eta_N \approx \frac{N^2}{2} \eta_N$. Note that we use index N to indicate that rates Λ_N and η_N depend on the current population size.

B. Modeling operation of the crowd-counting application

We want to measure the total number of users, which can be one of the three aforementioned types. Denote by $P(t)$ the number of primary users, by $S(t)$ the number of secondary, and by $U(t)$ the number of unregistered users at time t. As a first step, we will introduce assumptions that allow us to consider the spreading of registration messages as a Markovian process. Let us assume that primary users arrive to the area according to a Poisson process with rate λ_p and that their sojourn time inside is exponentially distributed with mean $T_p = 1/\mu_p$. Assume also that the total arrival rate of all other (non-primary) users is λ_u. The sojourn time of both the S and U type comes from the same distribution with mean value $T_u = 1/\mu_u$[1]. Then, arrivals to the area constitute a Poisson process with the total rate $\lambda_p + \lambda_u$ and all node sojourn times are exponentially distributed.

[1]For the purpose of modeling, we allow different sojourn times T_p and T_u, but it is justifiable to assume $T_p = T_u$.

TABLE I
POPULATION CHANGES IN A SMALL TIME INTERVAL Δt

Possible change	Probability
$(\Delta\vec{X}(t))_1 = [1,0,0]^T$	$p_1(t) = \lambda_p\Delta t + o(\Delta t)$
$(\Delta\vec{X}(t))_2 = [-1,0,0]^T$	$p_2(t) = \mu_p P(t)\Delta t + o(\Delta t)$
$(\Delta\vec{X}(t))_3 = [0,1,-1]^T$	$p_3(t) = \beta(t,\vec{X}(t))\Delta t + o(\Delta t)$
$(\Delta\vec{X}(t))_4 = [0,-1,0]^T$	$p_4(t) = \mu_u S(t)\Delta t + o(\Delta t)$
$(\Delta\vec{X}(t))_5 = [0,0,1]^T$	$p_5(t) = \lambda_u\Delta t + o(\Delta t)$
$(\Delta\vec{X}(t))_6 = [0,0,-1]^T$	$p_6(t) = \mu_u U(t)\Delta t + o(\Delta t)$
$(\Delta\vec{X}(t))_7 = [0,0,0]^T$	$p_7(t) = 1 - \sum_{m=1}^{6} p_m(t) + o(\Delta t)$
$(\Delta\vec{X}(t))_8 \neq (\Delta\vec{X}(t))_{i=1,\ldots,7}$	$p_8(t) = o(\Delta t)$

The system state at time t is $\vec{X}(t) = [P(t), S(t), U(t)]^T$; it is not fully observed since we do not know the number $U(t)$. We can however, infer approximate values of $U(t)$ by modeling the system dynamics, governed by stochastic processes which include arrivals, departures and transitions from unregistered to secondary users. However, multivariate stochastic systems often do not easily lend themselves to analysis, due to multiple variables (in our case three) and many interacting factors that drive transitions between states.

A common approach for studying complex system behaviour, which will also be utilized in our study, is by modeling with *stochastic differential equations*. This approach has been used in mathematical epidemiology, where models are often referred to as *compartmental models*; different population categories are called *compartments*. Thus, in our model we will have compartments of: primary (P), secondary (S) and unregistered (U) users. Transitions between the compartments, as well as the external arrivals and departures in our model are illustrated in Fig. 1.

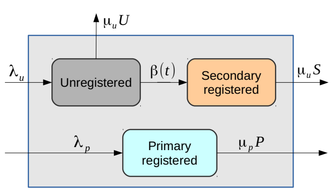


Fig. 1. Compartmental model.

To describe the system behavior, we first need to consider all possible changes from state $\vec{X}(t)$ to state $\vec{X}(t+\Delta t)$ in a small time interval Δt. These changes, denoted by $\Delta\vec{X}(t)$ are given in Table I. During Δt, an arrival can occur either in the P or U compartment with probabilities of these events $\lambda_p\Delta t$ and $\lambda_u\Delta t$, respectively. Users depart with rates $\mu_p P(t)$, $\mu_u S(t)$, and $\mu_u U(t)$. Probability $\beta(t,\vec{X}(t))\Delta t$ defines an encounter event when an unregistered node becomes a secondary registered (transition $U \to S$). We assume that R_{msg} is small enough such sthat it can be transferred in zero time. Lastly, the system can jump to any other state with a small probability $o(\Delta t)$ or remain in the same state $\vec{X}(t)$.

The rate $\beta(t,\vec{X}(t))$ depends on the total number of registered users, $R(t) = P(t) + S(t)$, the number of secondary users $U(t)$, and the pairwise contact rate (defined in III-A). Denote by $N(t)$ the total number of users in the area, $N(t) = P(t) + S(t) + U(t)$. In a dynamic system where population is subjected to frequent changes, the contact rate c depends on the population size: as it has been shown in [2], this dependence is linear for a certain range of population sizes. Therefore, we will consider time-dependent rates $c(t) = c(N(t))$ and $\eta(t) = \eta(N(t))$. We assume homogeneous mixing in the population and for the system in state $\vec{X}(t)$ we have $\beta(t,\vec{X}(t)) = R(t)U(t)\eta(t)$. After substituting $\eta(t) \approx \frac{c(t)}{N(t)}$, the rate at which unregistered become secondary users reads

$$\beta(t,\vec{X}(t)) = R(t)U(t)\frac{c(t)}{N(t)} \tag{1}$$

Now that we have all transition probabilities, we can construct the SDE model. We will use one of the methods described in [3]. The following SDE system describes the crowd-size behavior:

$$\begin{cases} dP(t) = [\lambda_p - \mu_p P(t)]dt + \vec{g}_1 d\vec{W}(t) \\ dS(t) = [\beta(t,\vec{X}(t)) - \mu_u S(t)]dt + \vec{g}_2 d\vec{W}(t) \\ dU(t) = [\lambda_u - \beta(t,\vec{X}(t)) - \mu_u U(t)]dt + \vec{g}_3 d\vec{W}(t) \\ \vec{X}(0) = [P(0), S(0), U(0)]^T \end{cases} \tag{2}$$

where $\vec{W}(t) = [W_i(t)]^T_{i=1,..,6}$ is a vector of six independent Wiener processes, $\vec{g}_{i=1,2,3}$ is the i-th row of the matrix $G(t,\vec{X}(t)) =$

$$\begin{bmatrix} \sqrt{\lambda_p} & \sqrt{\mu_p P(t)} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\beta(t,\vec{X}(t))} & \sqrt{\mu_u S(t)} & 0 & 0 \\ 0 & 0 & -\sqrt{\beta(t,\vec{X}(t))} & 0 & \sqrt{\lambda_u} & \sqrt{\mu_u U(t)} \end{bmatrix} \tag{3}$$

The initial state $\vec{X}(0)$ is (partially) known. The SDE system (2) can be rewritten in the form:

$$\begin{cases} d\vec{X}(t) = \vec{f}(t,\vec{X}(t))dt + G(t,\vec{X}(t))d\vec{W}(t) \\ \vec{X}(0) = [P(0), S(0), U(0)]^T \end{cases} \tag{4}$$

where the vector $\vec{f}(t,\vec{X}(t))$ is given by

$$\vec{f}(t,\vec{X}(t)) = \begin{bmatrix} \lambda_p - \mu_p P(t) \\ \beta(t,\vec{X}(t)) - \mu_u S(t) \\ \lambda_u - \beta(t,\vec{X}(t)) - \mu_u U(t) \end{bmatrix} \tag{5}$$

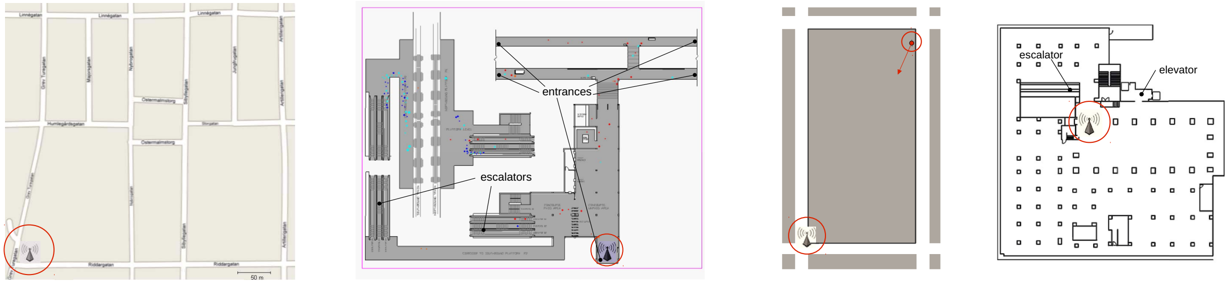
The solution to (4) gives the probability distribution $p(t,\vec{x})$ which satisfies, [3]:

$$\frac{dp(t,\vec{x})}{dt} = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\partial^2}{\partial x_i \partial x_j}\left[p(t,\vec{x})\sum_{k=1}^{6} g_{i,k}(t,\vec{x})g_{j,k}(t,\vec{x})\right] - \sum_{i=1}^{3}\frac{\partial[p(t,\vec{x})f_i(t,\vec{x})]}{\partial x_i} \tag{6}$$

Above, f_i is the $i-$th entry of $\vec{f}$ and $g_{i,j}$ is the i,j entry of the matrix G.

IV. EVALUATION

In this section we validate our model via simulations, and we analyze the results and the model's applicability and limitations.



(a) (b) (c) (d)

Fig. 2. Simulation area and the location of the access point: Östermalm (a), Subway (b), Open RWP (c), Conference scenario (d).

A. Mobility scenarios

Having derived the model which is based on rather favorable assumptions (from the modeling perspective), we want to investigate how well it fits more realistic mobility. To cover diverse cases, we consider four open system scenarios: two scenarios with simulator-generated traces, a synthetic mobility model and a real-life trace. Herein we give the most important details about the scenarios.

1. *Legion mobility traces*: For emulating realistic mobility, we use traces generated with *Legion Studio*[2], a multi-agent mobility simulator commonly used for pedestrian traffic planning and public spaces design and dimensioning.

We recreate two scenarios[3]. The first scenario represents an outdoor urban space, modeling a part of the downtown Stockholm, which we will further refer to as the Östermalm scenario. The topology is represented by a grid of streets, Fig. 2(a). There are fourteen streets (and eleven intersections) from where pedestrians enter the area and depart; while inside the area, they are constantly moving. The second scenario captures the movement of passengers in a subway station. The simulation area comprises two levels: the upper, entry level and the lower level with train platforms, which are connected by escalators, Fig. 2(b). Nodes can arrive either by walking in from one of the five entrance points on the upper level, or when a train arrives at the platform. A detailed description of these two scenarios can be found in [2].

2. *Open random waypoint (RWP)* [4]: models nodes traversing a confined area such as a city square. The main difference between this form of random waypoint and the traditional one is the varying population. Nodes arrive to the area according to a Poisson process and immediately start moving towards a randomly chosen point inside. Upon reaching a waypoint, the node decides either to leave the area, with probability P_{exit}, by choosing one of the exits as its next waypoint, or chooses a new waypoint inside. The area in our setup is rectangular, with size 100 m by 200 m and there are four entrances (exits), one at each corner, Fig. 2(c). The exit probability is 0.75. Nodes travel at speed of 1 m/s, without pausing between two consecutive waypoints.

3. *Conference scenario:* To create an open, trace-based mobility model, we use mobility traces collected at the HOPE conference[4] in 2008. The traces contain positioning data of attendees roaming on two floors of the conference hotel. We focus only on the floor where attendees exhibited higher mobility (roaming between a registration desk, exhibition area, commercial stands and similar). By extracting an hour-long sample, we obtain mobility details of 233 attendees. Fig. 2(d) depicts the floor plan and the locations of the elevator and escalators, where attendees could access or exit the floor.

TABLE II
SCENARIO PARAMETERS

	Östermalm	Subway	Open RWP	Hope
λ_p [s^{-1}]	0.040	0.195	0.050	0.012
λ_u [s^{-1}]	0.424	0.956	0.150	0.049
T_p [s]	360	188	266	1016
T_u [s]	307	191	259	954
c_N [s^{-1}]	0.102	1.46	0.070	0.102

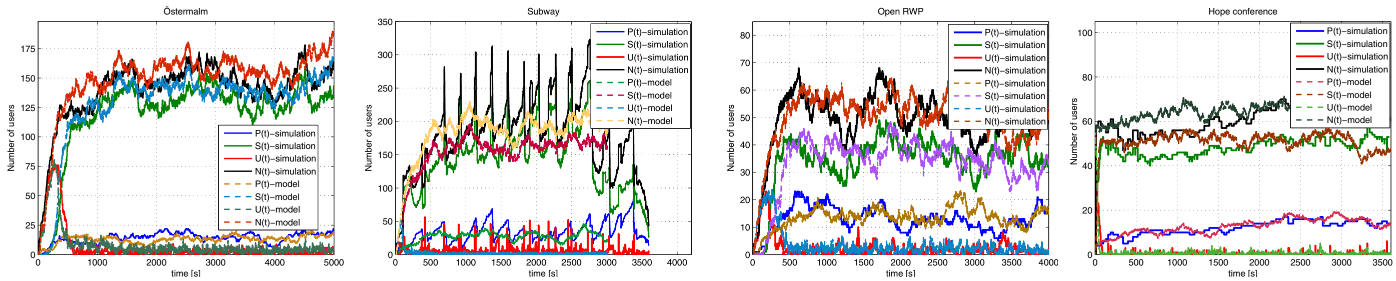
B. Model validation and analysis

To simulate the mobility scenarios, we use The ONE simulator [5]. In each scenario, there is a single access point located as depicted in Fig. 2 (a)–(d). In the simulations, we limit the transmission range for Östermalm and Open RWP to 10 m and for Subway and Hope to 5 m, assuming that communication in outdoor scenarios has longer range as compared to indoor scenarios. Access points have the same transmission range as mobile nodes.

Since the time granularity of the Hope dataset is 30 s, the actual node positions between two consecutive snapshot locations were interpolated as if the nodes were moving at a constant speed. In the Hope trace, some users are already in the system at time $t = 0$, while in the other three scenarios, initially, the system is empty. For computing the scenario parameters (listed in Table II), we discard measurements from the warm-up period, whose length we determine by applying Welch's method. Note that in the model validation step all parameters were computed offline.

First, we look into the average arrival rates for primary users, λ_p. The Östermalm trace was generated such that both ends of each intersecting street feed the area with Poisson arrivals with rate of 0.04 s^{-1}. Likewise, the Open RWP generated arrivals with rate 0.05 s^{-1} at each of the corners. From Little's law, which for the average number of primary users reads $\bar{P} = \lambda_p T_p$, we compute the arrival rates: 0.195 and 0.012 s^{-1} for the Subway and Hope scenario, respectively. Similarly, from the simulation results we find the arrival rates for non-primary users λ_u: 0.424, 0.956, 0.150 and 0.049 s^{-1}. When computing the contact rate c_N, in order to achieve higher

[2]Legion Studio software http://www.legion.com/legion-studio.
[3]Traces are available at http://crawdad.org/kth/walkers.
[4]Traces are available at http://crawdad.org/hope/amd.



(a) (b) (c) (d)

Fig. 3. Simulated counting process and one sample path of the stochastic model: Östermalm (a), Subway (b), Open RWP (c), Conference scenario (d).

confidence, we utilize the encounter information both from primary and secondary users. The contact rate measured by a secondary user is then the number of encounters normalized with its sojourn time in the S state.

It can be interesting to study the stochastic process in Eq. (4) in the transient state. In three scenarios, the system starts from an empty state; thus, before reaching the steady state, the average contact rate will be lower than c_N. Since we can measure only the contact rate c_N, we will introduce a heuristic approximation $c(t) = c_N \frac{N(t)^2}{2N^2}$ for the contact rate in the transient state, while in the steady state, the contact rate is calculated as $c(t) = c_N \frac{N(t)}{N}$. Note that, in a real deployment, the population size estimation would not be possible without any prior knowledge of system parameters, e.g. the expected contact rate in the observed area. The estimation requires an initial data acquisition period, whose length depends on the arrival rates and sojourn times. Once we have obtained the contact rate for a specific setup, this estimation will be readily available for future uses: for example, if we wanted to model the occupancy of the space during different times of the day.

We simulate the stochastic process $\vec{X}(t)$ by plugging the computed values into (4) and plot the results in Figs. 3 and 4. The system evolution over time is plotted in Fig. 3 and we observe that the modeled processes behave similarly to those simulated from the traces in the Östermalm, Open RWP and Hope scenarios. First, observe that the model accurately predicts the system behavior in the transient state, matching the duration of the transient period, and the population size changes for all user types, Fig. 3 (a) and (c). The model also provides a good estimation for the size of stochastic fluctuations; this is an improvement from a deterministic approximation of the process, which would give only steady state population distributions. In the Subway case, the model captures the increasing trends in the number of primary and secondary users, but is unable to account for the burstiness of node arrivals and departures when a train arrives.

Next, we compare the simulations with the analytic results with respect to the estimated number of users. Fig. 4 shows the median values (central marks in the boxes), as well as the 25th and 75th percentiles (edges of the boxes), and the distribution outliers (cross marks) for the numbers of primary, secondary, unregistered users and the total population size. The model yields accurate estimations in all scenarios, slightly overestimating the number of unregistered users (except for the Subway). Due to the high node density, in the simulations there are very few (in fact, close to zero) users who never meet anyone. However, for this particular use-case, an overestimation of the population size is not as critical as an underestimation would be.

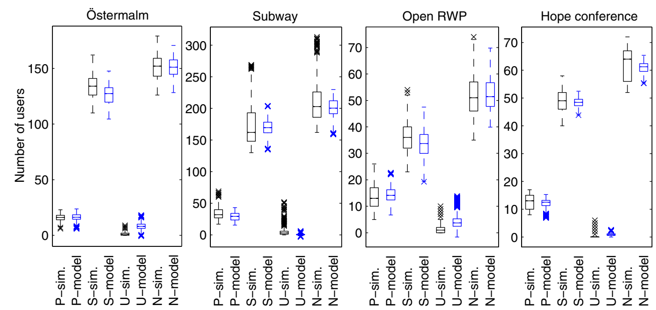


Fig. 4. Primary, secondary, unregistered and the total number of users: simulation results (black) compared with the analytic results (blue).

C. Model limitations

Since the model assumes constant arrival rate λ_u, batch arrivals, such as those manifested in the Subway scenario, cannot be captured. Large batch arrivals (and departures) of size 30 to 50 users are frequent: clearly, they can be identified as peaks in Fig. 3(b). These short-lived events, however, do not seem to have a strong impact on the user population distributions. Thus, we can still obtain reasonable estimations of the average numbers of each user type (measured over several minutes).

To summarize this section, we have shown that our model efficiently counts users of different types in a varying population, and can be applied in various mobility scenarios. The model relies on the Markovian property for user sojourn times and the assumption of the Poisson arrival process; while the investigated scenarios do not meet at least one of the assumptions, we are still able to obtain a very good match with the simulations.

V. Application

Now that we have a better understanding of when and how the model can be applied, we revisit the initial objectives to find the total number of users $N(t)$. To that end, we need to estimate $U(t)$.

Since the application logs the arrival and the departure times, as well as $P(t)$, the arrival rate λ_p and the departure rate μ_p can be readily computed. Likewise, the average time a node spends as a secondary user can be found from the node's registration and departure timestamps. The remaining challenge is to estimate the arrival rate λ_u. To this end, we use approximate Bayesian computation (ABC) implemented in the `abc-sde` software tool [6]. ABC is a "likelihood-free" methodology for Bayesian inference, particularly suitable for models with intractable or computationally demanding likelihood function, as well as partially observed models, like the one we consider. ABC takes as input a (partially observed) trajectory sample and estimates the model parameters by targeting an approximation to their posterior distributions (with respect to the observed sample). Our goal is to estimate a single parameter, λ_u. The method is not restricted to inference of a single unknown parameter, but can be used for a set of parameters; this however becomes more computationally expensive and less reliable for partially observed systems.

For each of the scenarios, we use a 150 s long sample $Y(t_n) = [P(t_n), S(t_n)]^T$, $n = 1, ..., 150$ and run the estimation algorithm. The estimated arrival rates are: 0.329, 1.04, 0.126, and 0.055 s^{-1}. Compare these rates with the corresponding λ_u from Table II: while we have very accurate estimations for the Subway and the Hope scenario, in the Östermalm case underestimating λ_u results in the difference of 16 users (out of around 151 in the system) during the observed time interval. Nevertheless, an approximate estimation can be sufficient to determine the parameter space, and then iteratively improve the estimation by fine tuning. This is done by simulating the counting process $\vec{X}(t)$ with the estimated parameters, and then comparing the simulation results for the number of primary and secondary users with the measured data. In this way, it can be easily inferred whether the actual arrival rate is higher or lower than the estimated value; this inference is subsequently used to iterate simulations until the simulated process is fitted to the measurements.

VI. Related Work

This paper revisits an epidemiological approach to model opportunistic message spreading, and presents new findings in opportunistic contact characterization—specifically when considering user population with churn—and we position our work with respect to these contributions.

Modeling spread of messages in mobile ad hoc networks inspired by the spread of infectious diseases between humans was first applied to epidemic routing in [7]. This approach has been further extended in [8], where the authors proposed a framework for studying a variety of epidemic routing schemes. Their common starting point is the use of ordinary differential equation models in performance evaluation. Following these studies, there has been a wealth of work considering various content spreading schemes through the deterministic approximations of Markovian chains representing content epidemics. While the deterministic models can be useful for estimating important characteristics of epidemics (e.g. conditions under which they occur, the rate at which they grow, the expected number of infections in equilibrium), such models become unreliable when the population is small, and the process exhibits stochastic fluctuations that cannot be neglected. We resort to the field of mathematical biology, and adopt the stochastic differential equation (SDE) modeling approach. There, the SDE approach has already produced a substantial amount of work, encompassing studies on both open and closed populations, different disease transmission patterns, derivation of the stability conditions and so on. To the best of our knowledge, this is the first work to propose the use of SDE models for opportunistic networks analysis.

Characterization of node interactions in open opportunistic systems is a relatively new topic. An empirical study on the impact of pedestrian mobility on connectivity in wireless systems was presented in [2]. Implications of the population churn on the system performance have been investigated in [4], [9]; while these works consider a particular content sharing scheme, our aim is to provide a model which can be adapted to various use-cases.

VII. Conclusion

In this paper we studied the operation of an opportunistic application in systems with varying population. First, we proposed an application for measuring the number of people in a crowd (crowd-counting), which can be used in various situations (both outdoor and indoor) and mobility scenarios. By using the stochastic differential equations modeling approach, we developed a model for the operation of this application, and we analyzed and validated the model in four scenarios. The model proves to be a good match for the examined cases, and since these scenarios have quite diverse features, we believe that the model is suitable for a broad range of applications, under certain assumptions on the user mobility behavior. Thus, a more detailed analysis of the conditions under which the model is applicable will be addressed in our future work. We will also explore the space of applications where the proposed modeling approach can be used.

References

[1] Ó. Helgason, E. A. Yavuz, S. T. Kouyoumdjieva, L. Pajevic, and G. Karlsson, “A mobile peer-to-peer system for opportunistic content-centric networking,” in *Proc. ACM SIGCOMM Mobiheld Workshop*, New Delhi, India, 2010, pp. 21–26.

[2] Ó. Helgason, S. Kouyoumdjieva, and G. Karlsson, “Opportunistic communication and human mobility,” *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1597–1610, July 2014.

[3] E. Allen, L. Allen, A. Arciniega, and P. E. Greenwood, “Construction of equivalent stochastic differential equation models,” *Stochastic Analysis and Applications*, vol. 26, pp. 274–297, 2008.

[4] M. Desta, E. Hyytiä, J. Ott, and J. Kangasharju, “Characterizing content sharing properties for mobile users in open city squares,” in *Proc. IEEE/IFIP WONS*, Banff, Alberta., 2013.

[5] A. Keränen, J. Ott, and T. Kärkkäinen, “The ONE Simulator for DTN Protocol Evaluation,” in *Proc. SIMUTools ’09*, Rome, Italy, 2009.

[6] U. Picchini. (2013) Abc-sde: a Matlab toolbox for approximate Bayesian computation (ABC) in stochastic differential equation models. [Online]. Available: http://sourceforge.net/projects/abc-sde/.

[7] T. Small and Z. J. Haas, “The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way),” in *Proc. ACM MobiHoc ’03*. New York, NY, USA: ACM, 2003, pp. 233–244.

[8] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, “Performance modeling of epidemic routing,” *Elsevier Computer Networks*, vol. 51, no. 10, pp. 2867–2891, 2007.

[9] E. Hyytiä, J. Virtamo, P. Lassila, J. Kangasharju, and J. Ott, “When does content float? characterizing availability of anchored information in opportunistic content sharing,” in *Proc. IEEE INFOCOM*, Shanghai, China, 2011, pp. 3137 –3145.