An Interaction-Based Mobility Model for Dynamic Hot Spot Analysis

Frédéric Morlot, François Baccelli, Salah Eddine Elayoubi

▶ To cite this version:

HAL Id: hal-00627081
https://hal.archives-ouvertes.fr/hal-00627081

Submitted on 27 Sep 2011

An Interaction-Based Mobility Model for Dynamic Hot Spot Analysis

Frédéric Morlot
INRIA - École Normale Supérieure
Paris, France
Email: frederic.morlot@ens.fr

François Baccelli
INRIA - École Normale Supérieure
Paris, France
Email: francois.baccelli@ens.fr

Salah Eddine Elayoubi
Orange Labs
Paris, France
Email: salaheddine.elayoubi@orange-ftgroup.com

Abstract—In this paper, we analyze phenomena related to user clumps and hot spots occuring in mobile networks at the occasion of large urban mass gatherings in large cities. Our analysis is based on observations made on mobility traces of GSM users in several large cities. Classical mobility models, such as the random waypoint, do not allow one to represent the observed dynamics of clumps in a proper manner. This motivates the introduction and the mathematical analysis of a new interaction-based mobility model, which is the main contribution of the present paper. This model is shown to allow one to describe the dynamics of clumps and in particular to predict key phenomena such as the building of hot spots and the scattering between hot spots, which play a key role in the engineering of wireless networks during such events. We show how to obtain the main parameters of this model from simple communication activity measurements and we illustrate this calibration process on real cases.

I. INTRODUCTION

Traces collected by wireless operators in urban environments during multi event mass gatherings reveal that user mobility is extremely volatile and leads to the formation of hot spots, exemplified on the snapshot of Figure 1 which was gathered during one the 2008 Paris *Fête de la Musique*, where unpredicted gatherings are triggered by a large number of spontaneous street concerts taking place throughout the city. Such observations have also been made during other events, like a summer festival in Poland and Euro2008 soccer matches in Spain and in Romania. A key phenomenon within this context concerns the dynamics of these hot spots, namely their formation and their disaggregation which is random both in space and time.

Accurate and yet tractable user mobility models allowing one to represent these phenomena would significantly improve the robustness and the realism of the simulators and the analytical models used by wireless communication operators. They would hence ameliorate the understanding of the impact of mass mobility on QoS within this setting. Predicting hot spots dynamics could also help in the design and the tuning of adaptive radio resource allocation schemes, where it is fundamental to have an idea of some time caracteristics of the system. This allows a better utilization of the corresponding infrastructure and eventually a better QoS and better services to end users. As existing models are not completely adapted to the phenomena we observe on traces, such as the random dynamic hot spot phenomenon alluded to above, we propose a new mobility model taking into account interaction between different users and show by a mathematical analysis based on Markov chain theory that this interaction-mobility model is sufficient to generate random hot spots. For this, we will represent an urban area as a collection of potential hot spot zones. These zones represent potentially attractive points in the city during such mass gatherings.

The basic assumption in our model is that users influence each other. In Section III, we first study the "sheep model", where this mutual influence is particularly strong and where hot spots form at random in time and space, but are stable once formed (i.e. there is no disaggregation of hot spots in this model yet). We start with the 2-zone case, in which we find closed forms for the mean time for the formation of a stable hot spot and for the law of its location. We also give asymptotic estimates for these quantities, and upper bounds on the convergence speed. We then generalize the previous results to an arbitrary number of zones.

Section IV introduces the "sheep and maverick" model, which is more realistic, yet tractable, and which features all key phenomena identified on the traces and alluded to above: the formation of hot spots at random times and places, that we will refer to as *filling*, their disaggregation that we will refer to as *scattering*, and the dynamics between them.

In a last part, we leverage filling and scattering in the "sheep and maverick" model to calibrate our mobility model and show that it can be made consistent with the traces.

II. THE NEED FOR A NEW MODEL

First, let us describe more precisely the phenomenon we want to model.
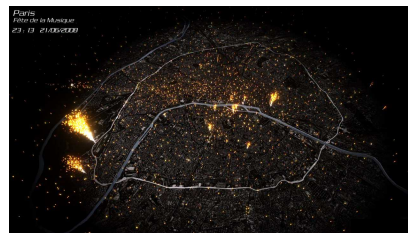

Fig. 1. Communication activity during *Fête de la Musique* in Paris, on June 21st, 2008. It is 11:23 pm. One can observe important hot spots in the center (*Châtelet, Bastille, Saint-Michel*) as well as at the *Parc des Princes* (the city stadium) and *Longchamp* (the city park).

A first remark is the existence of hot spots, that follow some dynamics. On the snaphots of Fig. 2, the crowd moves in large numbers outside a zone: we shall call that a *hot spot scattering*. A new or a few new hot spots then build up at other locations: we shall call that a *hot spot filling*.

(a) at 07:47 pm

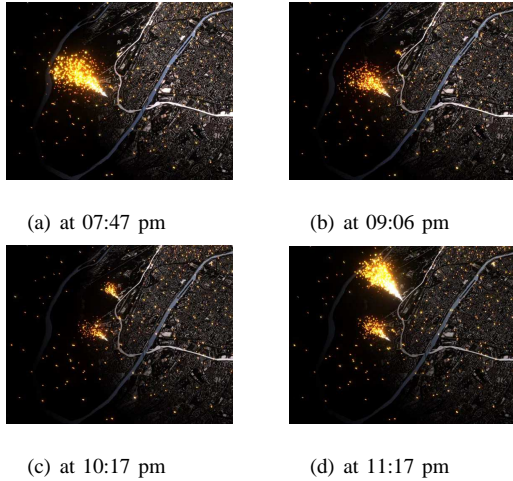(b) at 09:06 pm

(c) at 10:17 pm

(d) at 11:17 pm

Fig. 2. Communication activity in Paris on June 21st, 2008, between 07:47 pm and 11:17 pm. One can observe scattering from the *Parc des Princes* (south) to *Longchamp* (north).

We now survey existing mobility models and discuss how they could be adapted to cope with these phenomena.

The most popular mobility model used in communication networks is probably the Random Waypoint Model (see [8]). This model is easy to simulate and is analytically tractable in many situations depending of the mobility rules of the individuals on the domain (wrap around, edge reflection...). However, it applies to a single mobile user, or to a collection of users moving independently.

Using this model, in order to take clumping into account, we would have to choose a trip selection rule which gives more mass to potentially attractive zones. As a consequence, the law of the location of a single user would clearly favour these zones. However, such a situation prevents any macroscopic time dynamics. Indeed, with a large number of independent users, all following the same law, each snapshot would be more or less the same, very close to the theoretical location law. This follows directly from Sanov's theorem (see for instance [3], Theorem 6.2.10).

Thus, all clumps would happen at predefined locations and the sudden hot spot scattering as well as the sudden new hot spot aggregation observed in our traces would happen with extremely small probability.

More generally, any multi user model built from the superposition of many independent single user motions will exhibit the same "lack of hot spot dynamics", while a key point observed within traces is a strong dependence between users, since there is clearly a joint motion from a hotspot to another leading to *reinforced clumping* in the corresponding zone.

Nomadic Community Mobility Model or Reference Point Group Mobility Model (see [1]) consist in making small groups of users move in a mean given direction, while the

users themselves are allowed to walk here and there around the imposed direction. These models are well adapted for a group of visitors in a museum for instance, where the trajectory is more or less the same for everyone. However, such a system does not allow us to make groups scatter.

Let us now review some ideas about modelling users' interactions. Since we need a strong dependence among users, a natural attempt could be to work with permanental processes ([11]) that are known to exhibit frequent clumping. However, permanental processes are nothing but particular Cox processes (see [11], Remark 3.6.3.). This means that we would have to choose an underlying intensity measure which varies over time to cope with the dynamics of clumping zones, which is as difficult as building directly a convenient model.

Another class of mobility models that could be considered is the class of spatial birth and death processes (see [4]) first studied by Preston in 1975 (see [12]). Such a process can simulate characteristic point patterns if one chooses its stationary distribution in a proper manner. Furthermore, in contrast to the random waypoint model, it provides a real time dynamics. However, this model is known to lead to Gibbs point processes which are usually not analytically tractable.

This state of the art leads us to the conclusion that there is a need for a tractable model allowing one to exhibit the type of hot spot dynamics observed in communication traces.

III. THE SHEEP MODEL

The aim of this section is to model user clumping in some predefined zones. The basic idea is that users tend to move towards places where there are already more people, creating a self-enhanced clumping.

A. General Overview

Let us divide the city into K zones, that are exchanging mobiles one with another. In each zone, a mobile waits during an exponential time with parameter μ before deciding whether to change its zone or not.

In a first step, let us suppose that the mobiles are totally influenced by their environment: they behave as sheep. They decide to move towards one zone or another with a probability proportional to the current repartition of the population (see Fig. 3) of the zone. In other words, if we denote by N the total number of mobiles, by n_1 the number of mobiles in the first zone, by n_2 the number of mobiles in the second zone, etc., a mobile from the k-th zone decides to move to the k'-th zone with probability $n_{k'}/N$. Conversely, he decides to stay in the k-th zone with probability:

$$\frac{n_k}{N} = 1 - \sum_{k' \neq k} \frac{n_{k'}}{N}.$$

Obviously, there are K absorbing states, which correspond to the situation where all mobiles are in the same zone (these are the K stable hot spots mentioned in the introduction).
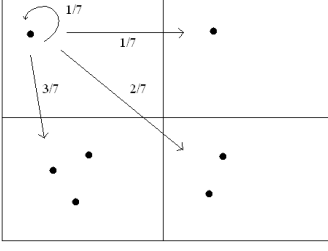
Fig. 3. The sheep model. Here, $N = 7$ and $K = 4$. The mobile in the upper-left corner has 4 choices. The four decision probabilities are proportional to the numbers of mobiles.

If we denote a state by $\mathbf{n} = (n_1, \dots, n_K)$, then an absorbing state is given by:

$$N\mathbf{e}_k = \underbrace{(0, \dots, N, \dots, 0)}_{k-th\ position}, \quad 1 \le k \le K.$$

We first study the 2-zone case for the sake of progressive exposition.

B. The 2-zone Problem

In this section we take $K = 2$. Let us denote by $\mathbf{n}(t) = (m(t), N - m(t))$ the state of the system at time t. Each mobile stays in his zone during a random exponential time with parameter μ, before deciding whether to move or not. All waiting times are supposed to be independent. In other words, the system features a superposition of N exponential clocks (m clocks in the left zone and $N - m$ in the right zone). The rate of events in the left zone is $m\mu$, and that in the right zone $(N - m)\mu$. An event in the left zone leads to a transition with probability $\frac{N-m}{N}$ whereas an event in the right zone eads to a transition with probability $\frac{m}{N}$. The whole process is Markov, with transitions described in Fig. 4.
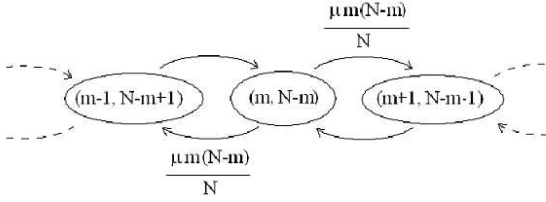

Fig. 4. Diagram of the markovian transitions.

Let us denote by $Q = (q_{m,m'})$ the infinitesimal generator.

1) Mean Absorption Time: We are interested in characterizing mobility between different zones. For instance, we would like to give an estimate of τ, the first time to reach an absorbing state, starting from the current state of the network $\mathbf{n}$: $\tau = \inf\{t/\mathbf{n}(t) = (0, N) \text{ or } (N, 0)\}$. The average value of τ is given by $h_\mathbf{n} = \mathbb{E}[\tau|\mathbf{n}(0) = \mathbf{n}]$. We know (see [10], Theorem 3.3.3. about hitting times) that the vector $\mathbf{h}$ is the minimum solution of the equation:

$$\begin{cases} h_{(0,N)} = h_{(N,0)} = 0 \\ \forall m \notin \{0, N\}, \quad \sum_{0 < i < N} q_{mi} h_{(i, N-i)} = -1. \end{cases} \quad (1)$$

We prove in appendix A that:

$$h_\mathbf{n} = \frac{1}{\mu} \left(\sum_{n_1 < i \le N} \frac{n_1}{i} + \sum_{n_2 < i \le N} \frac{n_2}{i} \right). \quad (2)$$

2) Asymptotic Estimate: Typically, N corresponds to the number of users in a city district, thus it is quite large. In order to speed up computations, it can be useful to use an asymptotic estimate of $h_\mathbf{n}$. Let $x = n_1/N$, for $1 \le n_1 \le N - 1$. We want an estimate that can be used for large values of N, and for all values of x. This means that the quality of this estimate should only depend on N. Since N is large, we consider x as a continuous variable taking its values in $]0, 1[$. The computation in appendix A gives:

$$h_\mathbf{n} \sim -\frac{N}{\mu}\left((1 - x)\log(1 - x) + x\log(x)\right) \quad (3)$$

and $r_\mathbf{n} \le 2/\mu$, where $r_\mathbf{n}$ is the absolute error of the approximate. Fig. 5 exemplifies the approximation curve.
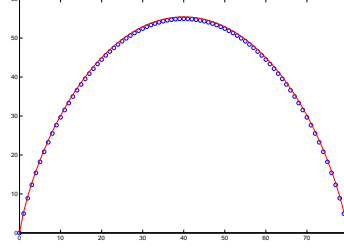

Fig. 5. Asymptotic estimate of $h_\mathbf{n}$ [in min]. $N = 80$ and $\mu = 1$ min^{-1}. The circles represent $h_\mathbf{n}$, and the continuous line represents the estimate.

3) Absorption Probability: So far we have obtained a closed form for the mean absorption time, as well as an asymptotic estimate. Now we would like to determine the probability to reach state $(0, N)$ rather than state $(N, 0)$, starting from $\mathbf{n} = (m, N - m)$. We shall denote this probability by $p_\mathbf{n}$. Using Q, we obtain the transitions for the imbedded Markov chain of the process (see Fig. 6).
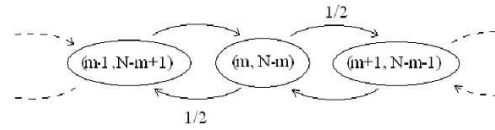

Fig. 6. Imbedded Markov chain of the process.

Conditioning with respect to T_1, the first transition time, and using Markov property, we get $p_\mathbf{n} = \frac{1}{2} p_{n+(-1,1)} + \frac{1}{2} p_{n+(1,-1)}$. Using the fact that $p_{(0,N)} = 1$ and $p_{(N,0)} = 0$, we conclude immediately that:

$$p_\mathbf{n} = n_2/N. \quad (4)$$

C. More Clumping Zones

We now consider the same problem with K zones.

1) Description of the system: Now a state is a K-uplet $\mathbf{n} = (n_1, n_2, \dots, n_K)$, with $\sum n_k = N$. Hence the space of possible states is a $(K - 1)$-dimensional variety. A transition from the k-th zone to the k'-th zone is given by:

$$\mathbf{n} \mapsto \mathbf{n}' = \mathbf{n} + \mathbf{e}_{k'} - \mathbf{e}_k.$$

(a) $\alpha = 0$ (b) $\alpha = 0.02$

(c) $\alpha = 0.04$ (d) $\alpha = 0.06$
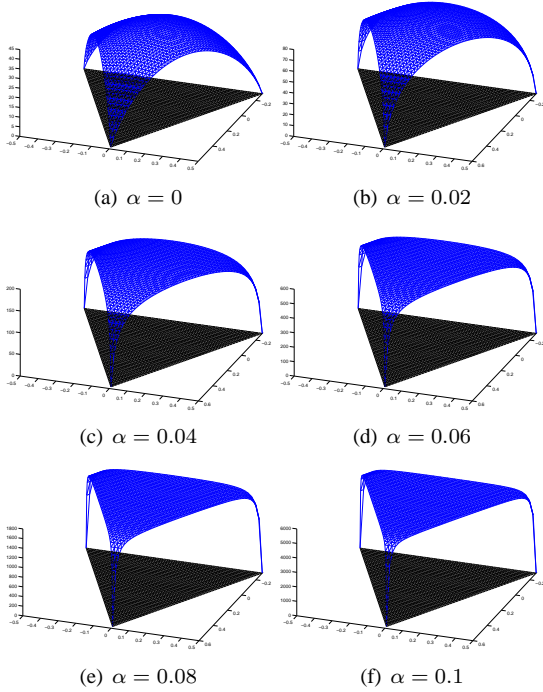
(e) $\alpha = 0.08$ (f) $\alpha = 0.1$
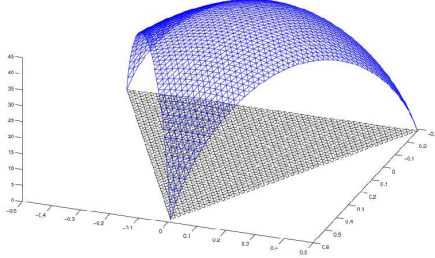
Fig. 8. Mean reaching time [in min] for $N = 50$, $K = 3$ and $\mu = 1$ min^{-1}. The maverickness rate α grows from 0 to 0.1. Note that the scale of the z-axis varies.
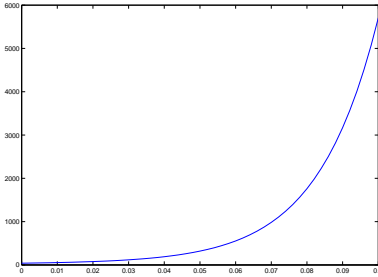

Fig. 9. Maximal mean reaching time [in min] with respect to α. $N = 50$, $K = 3$ and $\mu = 1$ min^{-1}.

B. Phase Transition

To characterize the phase transition observed in Fig. 8, we calculate the invariant measure π of the "sheep and maverick" Markov process, and observe if there is a value of α for which all configurations are equiprobable. π is given by the equation $\pi Q = 0$. Let $\beta = \frac{\alpha}{K(1-\alpha)}$, we have:

$$\pi(\mathbf{n}) = A \prod_{k=1}^{K} \frac{1}{n_k!} \prod_{i=0}^{n_k-1} (N\beta + i), \tag{7}$$

where A is a normalizing constant. This result is proved in appendix C, along with an asymptotic estimate. On Fig. 10, we plot the invariant measure with respect to α for $K = 2$ and $N = 50$. As already indicated by Fig. 8, it appears that when α grows, the influence of mavericks is more and more noticeable. For higher values of α, the probability is more concentrated on uniformly-spread states.

There is a critical value α_c where the curve's shape reverses. What does the transition look like? Is it possible to obtain a completely flat curve (or surface in the general case) ?

Expression (7) shows that it is possible. It is sufficient to

(a) $\alpha = 0.01$ (b) $\alpha = 0.03$ (c) $\alpha = 0.038$

(d) $\alpha = 0.04$ (e) $\alpha = 0.05$ (f) $\alpha = 0.2$
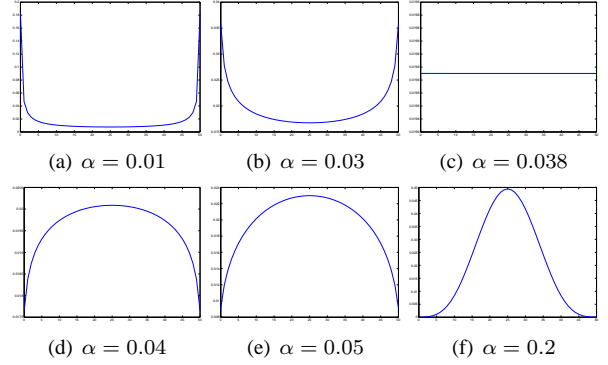
Fig. 10. Invariant measure for $N = 50$, $K = 2$ and $\mu = 1$ min^{-1}. When α exceeds 0.038, the shape of the curve reverses.

impose $N\beta - 1 = 0$ to obtain a uniform distribution. This corresponds to the critical value:

$$\alpha_c = \frac{K}{N + K}.$$

- for $\alpha < \alpha_c$, the influence of the sheep remains noticeable. This is the most interesting case, because the system still exhibits clumping.
- for $\alpha = \alpha_c$, all states are equivalent[3].
- for $\alpha > \alpha_c$, the sheep loose very fast their influence.

As traces reveal the existence of clumping, we shall assume until the end of the paper that $\underline{\alpha < \alpha_c}$.

C. Macro-states and Hot Spot Dynamics

The system is fully characterized by four parameters, which are N, K, μ and α. The three first parameters can be directly measured, but α cannot. And yet, it is of major importance. In a sense, it represents the propensity of users to clump. How can we make it best fit reality ? We would like to find a typical characteristic of the system, easy to measure and strongly related to α. This is what this subsection aims at.

Let us study further the case $\alpha < \alpha_c$. If we simulate such a system, we observe clumping as expected, but we also observe scattering. We would like to evaluate how often a scattering happens. However, extremal states of the form $N\mathbf{e}_k$ are extremely rare, so that it would be difficult to base our analysis on them. We will rather consider "macro-states", which are the union of several states close to one of the extremal states.

For example, let us denote by A_k the macro-state in which the k-th zone contains a significant proportion of the population: $A_k = \{\mathbf{n} : n_k > N_{\text{sup}}\}$, where N_{sup} is an upper bound for a zone to be considered as dense. Conversely, we define $B_k = \{\mathbf{n} : n_k < N_{\text{inf}}\}$, where N_{inf} is a lower bound. Then a scattering is simply a transition from A_k to B_k. More precisely, for k fixed, we define the corresponding scattering time S_k as the duration between the first instant when we reach A_k and the first instant when we reach B_k after that. Another

---

[3]As $N \to \infty$, this assumption can be interpreted as follows: since $\alpha_c \sim K/N$, it means that, asymptotically, there are as many mavericks as zones.

option would have consisted in using the last time we leave A_k before reaching B_k in place. But this is not a stopping time, which would make such a choice more difficult to handle using Markov chain theory. Hence the former definition. By symmetry, we can suppose that $k = 1$ for example, and omit subscript k in the following. At the first instant when we reach A, we are necessarily in $M = \{\mathbf{n}/n_1 = N_{\sup} + 1\}$. From the strong Markov property, the prediction of the future of the chain is condionally independent of the past given the state of the system at this stopping time.

It is shown in appendix D that $\mathbb{E}[S|\mathbf{n}^{(0)} = \mathbf{n}]$ is given the following analytic expression:

$$\frac{N}{\mu(1-\alpha)} \sum_{l=N_{\inf}}^{N_{\sup}+1} \sum_{j=l}^{N} \frac{\prod_{l \le i < j} \delta(i)}{j(N(K-1)\beta + N - j)}, \qquad (8)$$

with

$$\delta(i) = \frac{(N-i)(N\beta + i)}{i(N(K-1)\beta + N - i)}.$$

The expression in (8) only depends on the fact that $n_1 = N_{\sup} + 1$, but not on the exact value of $\mathbf{n}$ itself. Hence, the strong Markov property actually implies that $\mathbb{E}[S] = (8)$.

Fig. 11 depicts the evolution of $\mathbb{E}[S]$ in function of α for several values of N.

Let $\alpha_{\min}$ be the value of α such that $\mathbb{E}[S]$ is minimum. For certain values of N, $\mathbb{E}[S]$ is strictly decreasing over $[0,1]$, so that $\alpha_{\min} = 1$, but for some other values of N (like $N = 1000$), $\mathbb{E}[S]$ starts increasing after $\alpha_{\min}$.
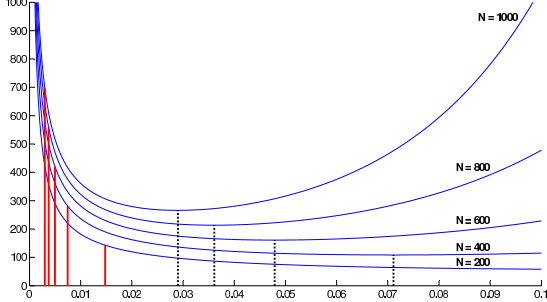

Fig. 11. Mean scattering time [in min] with respect to α. $N_{\sup}/N = 80\%$, $N_{\inf}/N = 20\%$, $K = 3$ and $\mu = 1 \text{ min}^{-1}$. For each curve, the vertical solid line corresponds to the critical value α_c, while the dotted line corresponds to $\alpha_{\min}$ (whenever it is less than the right bound of the axis).

In the same way, we can compute the mean "filling time" F of a zone, which is the dual of the scattering time, and which we propose to define as the time for the population in a given zone to grow from $N_{\inf}$ to $N_{\sup}$. We have:

$$\mathbb{E}[F] = \frac{N}{\mu(1-\alpha)} \sum_{l=N-N_{\sup}}^{N-N_{\inf}+1} \sum_{j=l}^{N} \frac{\prod_{l \le i < j} \epsilon(i)}{j(N\beta + N - j)}, \qquad (9)$$

with

$$\epsilon(i) = \frac{1}{\delta(N-i)} = \frac{(N-i)(N(K-1)\beta + i)}{i(N\beta + N - i)}.$$

Note that it is easy to find some recursive formulas for the computation of $\mathbb{E}[S]$ and $\mathbb{E}[F]$, allowing to decrease its complexity down to a Big-O of N.

V. CALIBRATION WITH TRACES

The aim of this section is to see whether the crowd movements observed in the traces can be described by our "sheep and maverick" mobility model.

We propose to use $\mathbb{E}[S]$ or $\mathbb{E}[F]$ for our calibration. These quantities can be measured. Let us now see that if $\mathbb{E}[S]$ and N are known for instance, we can then deduce α by solving an inverse problem. If the function $\mathbb{E}[S] = f(\alpha)$ is injective, at least for $\alpha < \alpha_c$, this inverse problem can be solved without ambiguity. We know that $f(\alpha)$ is decreasing on $]0, \alpha_{\min}[$, so that a necessary and sufficient condition is $\alpha_c \le \alpha_{\min}$. On Fig. 11, this is always the case, but this is not true in general; a counter-example is $K = 2$, $N = 100$, $N_{\sup}/N = 0.99\%$. In practice, $\alpha_{\min}$ seems to be hard to compute analytically, so that we could not find a simple sufficient condition.

Our experiences show that such counter-examples appear to be extremely rare, especially when $K > 2$. It seems that most choices of $N_{\inf}$ and $N_{\sup}$ satisfy the injectivity of $f(\alpha)$ on $]0, \alpha_c[$. However, if it turned out that it is impossible to choose $N_{\inf}$ and $N_{\sup}$ properly, we advise to complement the measure of $\mathbb{E}[S]$ with a measure of $\mathbb{E}[F]$ for instance.

In contrast, when measuring just $\mathbb{E}[F]$, one has to be more cautious, since the injectivity of $g(\alpha) = \mathbb{E}[F]$ on $]0, \alpha_c[$ seems less common.

Let us consider for instance the transfer that happens between *Parc des Princes* and *Longchamp* (see Fig. 2). Observing carefully our data, we conclude that there are three attracting zones:

- *Longchamp* itself
- the *Parc des Princes*, northern part
- the *Parc des Princes*, southern part

so that we take $K = 3$. On Fig. 12, one can see the numbers of users in each zone.
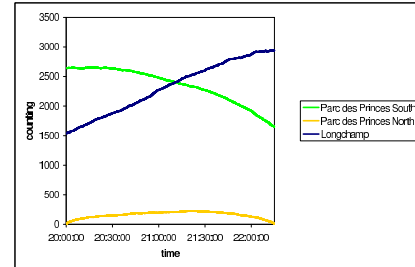
Fig. 12. Traces collected during the *Fête de la Musique* between the three attracting zones.

The total number of traces remains more or less constant, so that we choose an average value of $N = 4830$. This gives $\alpha_c = 6.2 \times 10^{-4}$. Choosing to calibrate first on *Longchamp* filling, we take $N_{\sup}/N = 55\%$ and $N_{\inf}/N = 45\%$. After verifying that $g(\alpha)$ is injective on $]0, \alpha_c[$, we find $\underline{\alpha = 0.6 \times 10^{-4}}$, that is, $\alpha \ll \alpha_c$. This was to expect, since scattering is very clear in our case.

In order to verify this value, we also measure the *Parc des Princes South* scattering, which yields $\alpha = 0.5 \times 10^{-4}$.

Now, if our model is consistent, it has to predict other clumps or scattering over Paris. Let us consider the triangle

Châtelet-Saint Michel-Bastille for instance, in the heart of Paris (see Fig. 13). These zones are attractive. Besides, we consider a fourth zone which is the complementary region inside the triangle. That one experiences scattering, whereas the three attractive zones experience filling at the same time.
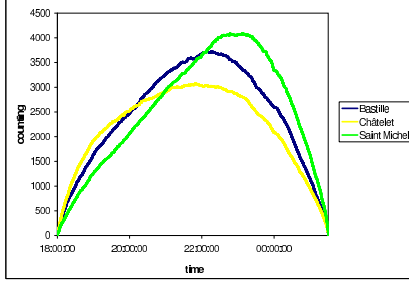
Fig. 13. Traces collected during the *Fête de la Musique* in the triangle *Châtelet-Saint Michel-Bastille*.

Theoretically, a filling of a zone from 5% to 28% for instance should last about 10800 seconds, according to the value of α we found previously. Now, the filling of *Bastille* lasts about 11300 seconds, so does the filling of *Châtelet*, and the filling of *Saint Michel* lasts about 11000 seconds, which is very close to our prediction.

VI. CONCLUSION AND PERSPECTIVES

In this paper, we proposed a new mobility model able to represent and quantify the hot spots found on traces, along with their random time dynamics. Our aim was not to build a generic mobility model, adapted to every kind of situation, but rather to focus on multi-event mass gatherings. We have proved our model to be analytically tractable, by derivating closed forms and asymptotic estimates. Using measurements from a GSM network, we also showed how to calibrate the model to fit reality. In addition, our model was proved to be able to predict the order of magnitude of hot spots in different time-space situations.

Apart from the main application, which consists in building an efficient mobility model for network simulators, we describe here some applications, that rely on the ability to predicting clumping or scattering with a good likelihood.

A. Dynamic Spectrum Allocation and Software Defined Radio

Dynamic Spectrum Allocation (see [7]), consists in allocating variable bandwidth resources, according to differrent zones and different densities of population. As for Software Defined Radio (see [9]), it consists in deploying base stations that are able to switch from a system (like GSM) to another one (like UMTS), depending on the demand.

Even if they have not been implemented in real networks yet, these two domains are today quite an active field of research. It is crucial that implemented algorithms can have an idea of the typical evolution durations of the system.

For instance, if we can predict if a zone that is full will remain so for a long time, or if an empty zone will soon fill, one can decide with much more accuracy whether allocating resources or not.

More generally, some optimization algorithms are known to depend massively on the tuning of a key parameter. For instance, [6] proposes an algorithm to optimize radio ressources in a 802.11 network, based on Gibbs sampler. The "temperature parameter" of the sampler has to be tuned very carefully, so that the optimizer can react quicklier than the typical time of evolution of the system itself. More concretely, in our case, supposing that we want to take users mobility into account in a 802.11 network, if we can anticipate a scattering, we are able to adapt the optimizer in a suitable manner.

Also in the case of a genetic algorithm, it can be interesting to adapt the cross-over and the mutation rates of the optimizer to react to a sudden evolution of the system. Indeed, the higher these rates, the stronger the diversity of the genetic population. In our case, a scattering is likely to modify drastically the system to optimize. Thus it is useful broaden preventatively the variety of solutions that are explored.

B. Data Caching in Dense Ad-hoc Networks

In a dense data ad-hoc network, it can be useful to add some fixed servers which keep the most frequently asked data in cache memories. Various algorithms already exist on this subject (see for instance [5][13]). The question is where and when should we cache memory ?

The idea is to select some strategic places where people usually clump together. These are our K zones. Suppose that, using section V, we have previously evaluated α over the study area, possibly for a different number of users and zones. Then, measuring in real-time the actual number N of users and computing the corresponding scattering times, as soon as a zone exhibits clumping, one knows how long on average this will last. In other words, one knows if it is worth caching data at that place.

Conversely, if a zone is almost empty, one knows, using "filling times", if there is a risk that we will soon encounter a peak of population and have to rush to cache memory there.

APPENDIX

A. Resolution of Equation (1)

$\sum_{0 < i < N} q_{mi} h_{(i, N-i)} = -1$ can be rewritten as:

$$\begin{cases} 2h_{(1,N-1)} - h_{(2,N-2)} = \frac{N}{\mu(N-1)} \\ 2h_{(N-1,1)} - h_{(N-2,2)} = \frac{N}{\mu(N-1)} \\ \text{otherwise:} \\ 2h_{\mathbf{n}} - h_{\mathbf{n}+(-1,1)} - h_{\mathbf{n}+(1,-1)} = \frac{N}{\mu i(N-i)}. \end{cases}$$

Thus we have to invert the matrix $U = (u_{i,j})$ with $u_{i,i} = 2, u_{i,i-1} = u_{i-1,i} = -1, u_{i,j} = 0$ otherwise. One can easily check that U^{-1} is given by V/N, where the elements of V are:

$$v_{ij} = \begin{cases} j(N-i) & \text{if } i \geq j \\ i(N-j) & \text{if } i \leq j. \end{cases}$$

Equation (2) follows immediately.

Let us now give an asymptotic estimate of this expression. We introduce $x = n_1/N$, $t = i/N$ and $dt = 1/N$. We have:

$$\mu\, h_{\mathbf{n}} = Nx \sum_{i=n_1+1}^{N} \frac{1}{N} \cdot \frac{1}{i} + N(1-x) \sum_{i=n_2+1}^{N} \frac{1}{N} \cdot \frac{1}{i}.$$

Both sums of the right handside can be interpreted as Riemann integrals on $[x, 1]$ and $[1 - x, 1]$. Let us work out the first one for example:

$$\int_{x+1/N}^{1+1/N} \frac{1}{t}dt \leq \sum_{i=n_1+1}^{N} \frac{1}{N} \cdot \frac{1}{i} \leq \int_{x}^{1} \frac{1}{t}dt$$

A fortiori:

$$-\log(x) - \frac{1}{Nx} \leq \sum_{i=n_1+1}^{N} \frac{1}{N} \cdot \frac{1}{i} \leq -\log(x).$$

We would obtain the same inequality replacing x by $1 - x$. Defining the remainder:

$$r_{\mathbf{n}} = \left| h_{\mathbf{n}} + \frac{N}{\mu}\big(x\log(x) + (1-x)\log(1-x)\big) \right|,$$

we have $\mu r_{\mathbf{n}} \leq 2$. Thus, our estimate is $(-Nx\log(x) - N(1-x)\log(1-x))/\mu$, and the error is uniformly bounded by $2/\mu$. Note that even for $x = 1/N$ or $x = N - 1/N$, which yields the minimum nonzero value of $h_{\mathbf{n}}$, the remainder still remains a little-o of $h_{\mathbf{n}}$.

B. Proof of equation (5)

We are now looking for the solution of equation (1) in K dimensions. In fact, we assume that (5) is the solution, and we prove it. First remark: if $\mathbf{n}$ is an absorbing state, (5) is equal to 0. Then, let us introduce $\epsilon_{\mathbf{nn'}} = h_{\mathbf{n'}} - h_{\mathbf{n}}$. One can easily check that $\forall \mathbf{n}$, $\sum_{\mathbf{n'}} q_{\mathbf{nn'}} h_{\mathbf{n'}} = \sum_{\mathbf{n'} \neq \mathbf{n}} q_{\mathbf{nn'}} \epsilon_{\mathbf{nn'}}$.

- let us determine $\epsilon_{\mathbf{nn'}}$; if $\mathbf{n'}$ is not an absorbing state, we have:

$$h_{\mathbf{n'}} = h_{\mathbf{n}} + \underbrace{\frac{1}{\mu(K-1)}\left(\sum_{n_{k'} < i < N} \frac{1}{i} - \sum_{n_k \leq i < N} \frac{1}{i} \right)}_{\epsilon_{\mathbf{nn'}}}.$$

Now, if $\mathbf{n'}$ is an absorbing state, it means that $n_k = 1$ and $n_{k'} = N - 1$. Then equation (5) gives:

$$h_{\mathbf{n'}} - h_{\mathbf{n}} = -h_{\mathbf{n}} = -\frac{1}{\mu(K-1)} \sum_{1 \leq i < N} \frac{1}{i},$$

so that it is easy to check that the previous expression of $\epsilon_{\mathbf{nn'}}$ is still valid.

- now, let us prove that expression (5) verifies (1). In a first time, let us focus on the most general case, *ie* when none of the n_ks is zero. We can write:

$$\sum_{\mathbf{n'} \neq \mathbf{n}} q_{\mathbf{nn'}} \epsilon_{\mathbf{nn'}}$$

$$= \underbrace{\sum_{(k,k')}^{\neq} \frac{n_k n_{k'}}{N(K-1)}\left(\sum_{n_{k'} < i < N} \frac{1}{i} - \sum_{n_k < i < N} \frac{1}{i} \right)}_{=0 \text{ by symmetry}}$$

$$-\sum_{(k,k')}^{\neq} \frac{n_k n_{k'}}{N(K-1)} \cdot \frac{1}{n_k}$$

$$= -\frac{1}{N(K-1)} \cdot (K-1)\sum_{k'} n_{k'} \;=\; -1,$$

which completes the proof of the general case.

At last, if a n_k is zero, the $(k \to k')$ transition is forbidden, so that the corresponding term above should not be taken into account. But fortunately, in that case we have $n_k n_{k'}/N(K-1) = 0$, so that previous calculation still holds.

C. Steady-state Computation

In this section, we will explain how to find progressively expression (7). First let us take $K = 2$ and keep notations of section III-B. For more simplicity, let us also introduce successively:

$$\beta = \frac{\alpha}{2(1-\alpha)},$$
$$u(x) = (N-x)(N\beta + x),$$
$$v(x) = u(N-x),$$
$$\pi(m) = \pi\big((m, N-m)\big).$$

Then balance equations give us, if $m \notin \{0, N\}$:

$$q_{m-1,m}\pi(m-1) + q_{m+1,m}\pi(m+1) = q_{m,m}\pi(m).$$

Multiplying by $\frac{N}{\mu(1-\alpha)}$ yields:

$$u(m-1)\pi(m-1) + v(m+1)\pi(m+1)$$
$$= (u(m) + v(m))\pi(m).$$

So that if we introduce:

$$\gamma(m) = v(m)\pi(m) - u(m-1)\pi(m-1),$$

we obtain: $\forall m \notin \{0, N\}, \gamma(m) = \gamma(m+1)$.
Hence $\gamma(m) = \gamma(1) = 0$. Finally:

$$\pi(m) = \pi(0)\prod_{i=1}^{m} \frac{(N-i+1)(N\beta + i - 1)}{i(N\beta + N - i)}.$$

In order to generalize this formula, we would like to rewrite it in a symmetric manner w.r.t. n_1 and n_2. In fact, we have:

$$\pi(m) = \pi(0)\frac{\prod_{i=1}^{n_1} \frac{N\beta+i-1}{i} \prod_{i=1}^{n_2} \frac{N\beta+i-1}{i}}{\prod_{i=1}^{N} \frac{N\beta+i-1}{i}}.$$

This is equation (7) in case $K = 2$.

Now we can reasonably suppose that (7) still holds in the general case. In fact, this is quite immediate. Indeed, previous calculation assures us that in case $K = 2$, global balance equations are satisfied. Now, thanks to the product form of (7), we know by the same calculation that in the general case, the partial balance equations along one transition are satisfied. In other words, $\forall(k, k')$, if we denote $\mathbf{n} + \mathbf{e}_{k'} - \mathbf{e}_k$ by $\mathbf{n'}$ and $\mathbf{n} - \mathbf{e}_{k'} + \mathbf{e}_k$ by $\mathbf{n''}$, we have $q_{\mathbf{n'},\mathbf{n}}\pi(\mathbf{n'}) + q_{\mathbf{n''},\mathbf{n}}\pi(\mathbf{n''}) = \pi(\mathbf{n})(q_{\mathbf{n},\mathbf{n'}} + q_{\mathbf{n},\mathbf{n''}})$.

This achieves the proof. Note that we can also find an asymptotic approximate of this expression. We have:

$$\log \prod_{i=0}^{n_k-1} (N\beta + i) = n_k \log N + \underbrace{\sum_{i=0}^{n_k-1} \log\left(\beta + \frac{i}{N}\right)}_{(*)}.$$

Approximating $(*)$ by a Riemann's integral, one can find:

$$\frac{1}{n_k!} \prod_{i=0}^{n_k-1} (N\beta + i) \sim \frac{N^{n_k}}{n_k!} \left(\frac{(\beta + x_k)^{(\beta + x_k)}}{\beta^\beta e^{x_k}}\right)^N \sqrt{\frac{\beta}{\beta + x_k}}.$$

After multiplying K such members and removing terms that only depend on N or β, we obtain:

$$\pi(\mathbf{n}) \sim A' \prod_{k=1}^{K} \frac{1}{(Nx_k)!} (\beta + x_k)^{N(\beta + x_k) - \frac{1}{2}},$$

where A' is another normalizing constant. If we introduce the relative error $r_\mathbf{n}$, using the mean value theorem, one can obtain the following bound:

$$r_\mathbf{n} \le e^{\frac{K}{4N\beta} + \frac{K}{4N^2\beta^2}} - 1 \sim \frac{K/4\beta}{N}.$$

D. Proof of equation (8)

As previously, let us begin with the case $K = 2$. We are in state $(m, N - m)$ and we are looking for the hitting time h_m of $\{\mathbf{n}/n_1 < N_{\inf}\}$.

We want to obtain an analytical formula of h_m (let us say for $N_{\inf} \le m \le N$, since otherwise $h_m = 0$). If we introduce $u_m = h_m - h_{m-1}$, we have the following recursion ($N_{\inf} + 1 \le m \le N - 1$):

$$u_{m+1} = \frac{m(N\beta + N - m)}{(N - m)(N\beta + m)} u_m - \frac{N}{\mu(1 - \alpha)(N - m)(N\beta + m)}.$$

Then the idea is to introduce:

$$\delta(i) = \frac{(N - i)(N\beta + i)}{i(N\beta + N - i)} \text{ and } v_m = \frac{u_m}{\prod_{i=m}^{N-1} \delta(i)}.$$

We obtain the following value of $v_m - v_{m+1}$:

$$\frac{N}{\mu(1 - \alpha)m(N\beta + N - m)\prod_{i=m}^{N-1} \delta(i)},$$

so that, for $N_{\inf} + 1 \le m \le N$, v_m is equal to:

$$v_N + \frac{N}{\mu(1 - \alpha)} \sum_{j=m}^{N-1} \frac{1}{j(N\beta + N - j)\prod_{i=j}^{N-1} \delta(i)}.$$

Since $q_{N,N}h_N + q_{N,N-1}h_{N-1} = -1$, we also have: $u_N = -1/q_{N,N} = 2/\mu\alpha N = v_N$. Finally:

$$v_m = \frac{N}{\mu(1 - \alpha)} \sum_{j=m}^{N} \frac{1}{j(N\beta + N - j)\prod_{i=j}^{N-1} \delta(i)},$$

and:

$$u_m = \frac{N}{\mu(1 - \alpha)} \sum_{j=m}^{N} \frac{\prod_{m \le i < j} \delta(i)}{j(N\beta + N - j)}.$$

To determine the h_ms, we still miss an equation. This is the last one before reaching $\{\mathbf{n}/n_1 < N_{\inf}\}$:

$$q_{N_{\inf}, N_{\inf}} h_{N_{\inf}} + q_{N_{\inf}, N_{\inf}+1} h_{N_{\inf}+1} = -1.$$

Together with $h_{N_{\inf}+1} - h_{N_{\inf}} = u_{N_{\inf}+1}$, this leads successively to:

$$h_{N_{\inf}} = \frac{N}{\mu(1 - \alpha)} \sum_{j=N_{\inf}}^{N} \frac{\prod_{N_{\inf} \le i < j} \delta(i)}{j(N\beta + N - j)},$$

and, more generally, to:

$$h_m = \frac{N}{\mu(1 - \alpha)} \sum_{l=N_{\inf}}^{m} \sum_{j=l}^{N} \frac{\prod_{l \le i < j} \delta(i)}{j(N\beta + N - j)}.$$

In particular, for $m = N_{\sup} + 1$, we obtain the claimed formula. Now suppose we are in the general case. It is perfectly correct to consider the union of the $K - 1$ last zones as one big zone. A sheep will stay in this "macro-zone" with probability $\frac{n_2 + ... + n_k}{N}$, and a maverick will stay with probability $\frac{K-1}{K}$. This last probability is the only adaptation we have to make in the previous calculation. Introducing $\beta' = (K - 1)\beta$, we obtain the following new recursion:

$$u_{m+1} = \frac{m(N\beta' + N - m)}{(N - m)(N\beta + m)} u_m - \frac{N}{\mu(1 - \alpha)(N - m)(N\beta + m)},$$

The rest of the calculation is exactly the same. In particular, for a fixed n, the result is the same for each $\mathbf{n}$ such that $n_1 = n$.

REFERENCES

[1] Camp T., Boleng J., and Davies V., *A Survey of Mobility Models for Ad Hoc Network Research*, WCMC, Special Issue on Mobile Ad Hoc Networking - Research, Trends and Applications, vol. 2, no. 5, pp. 483502, 2002.

[2] Cover T.M., Thomas J.A., *Elements of Information Theory, Second Edition*, Wiley, 2006.

[3] Dembo A. and Zeitouni 0., *Large Deviations Techniques and Applications*, Jones and Bartlett, 1993.

[4] Garcia N.L. and Kurtz T.G., *Spatial birth and death processes as solutions of stochastic equations*, Alea 1, 281-303 (2006).

[5] Hara T., *Replica Allocation Methods in Ad Hoc Networks with Data Update*, Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, 2003.

[6] Kauffmann B., Baccelli F., Chaintreau A., Mhatre V., Papagiannaki K. and Diot C., *Measurement-Based Self Organization of Interfering 802.11 Wireless Access Networks*, Proceedings IEEE Infocom'07, Anchorage, Pages:1451 - 1459, May 2007.

[7] Leaves P. et al., *Dynamic Spectrum Allocation in Composite Reconfigurable Wireless Networks*, IEEE Communications Magazine, May 2004.

[8] Le Boudec J.Y. and Vojnović M., *Perfect Simulation and Stationarity of a Class of Mobility Models*, Proceedings IEEE Infocom'05.

[9] Mitola J., *Cognitive Radio: an Integrated Agent Architecture for Software Defined Radio*, PhD Thesis, Royal Institute of Technology (KTH), 2000.

[10] Norris J.R., *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, 1997.

[11] Peres Y., Virág B., Ben Hough J. and Krishnapur M., *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*, draft version of April 27, 2007. Available at Y. Peres' homepage: http://www.stat.berkeley.edu/~peres/.

[12] Preston C.J., *Spatial birth-and-death processes*, Bull. Int. Statist. Inst. 46 (2) 371391, 1975.

[13] Yin L. and Cao G., *Supporting Cooperative Caching in Ad Hoc Networks*, IEEE Transactions on Mobile Computing, vol. 5, no. 1, pp. 77-89, Jan. 2006.