Buildings Affect Mobile Patterns: Developing a New Urban Mobility Model

Zimu Zheng
The Hong Kong Polytechnic Univ.

Feng Wang
The Univ. of Mississippi

Dan Wang
The Hong Kong Polytechnic Univ.

Liang Zhang
JD.com

ABSTRACT

Urban Mobility Models (UMMs) are fundamental tools for estimating the population in urban sites and their spatial movements over time. They have great value for such applications as managing the resources of cellular networks, predicting traffic congestion, and city planning. Most existing UMMs were developed primarily in 2D. However, we argue that people's movements and living patterns involve 3D space, i.e., buildings, which can heavily affect the accuracy of UMMs.

In this paper, we for the first time conduct a comprehensive study on the impacts of buildings on human movements, and the effect on UMMs. In particular, we start from an extensive trace analysis of two different real-world datasets. Our key observation is that human patterns of movement among urban sites are affected by buildings, with buildings being able to "temporarily hold" human mobility. We innovatively capture this property by extending Markov processes, which have been widely used in developing UMMs, with *semi-absorbing states*. We then develop a Semi-absorbing Urban Mobility model (SUM) and theoretically prove its properties to capture the intrinsic impacts of buildings with an analysis of SUM on its difference from that of previous UMMs. Our evaluation also demonstrates that, as a basis for supporting mobile applications in an intracity and hourly scale, the SUM is far superior to previous UMMs. Our real-world case study on cellular network resource allocations further reveals the effectiveness of our SUM model. We show that the performance of the resource allocation scheme in a cellular network substantially improves by using SUM, with a reduction in the packet loss probability of 3.19 times.

CCS CONCEPTS

• **Networks → Network mobility**; • **Mathematics of computing → Markov processes**;

KEYWORDS

Urban Mobility Model, Semi-absorbing Markov Process, Building

ACM Reference Format:
Zimu Zheng, Feng Wang, Dan Wang, and Liang Zhang. 2018. Buildings Affect Mobile Patterns: Developing a New Urban Mobility Model. In *Proceedings of ACM BuildSys (ACM BuildSys'18).* ACM, Shenzhen, China, 10 pages. https://doi.org/10.1145/3276774.3276780

ACM BuildSys'18, November 2018, Shenzhen, China

ACM ISBN 978-1-4503-5951-1/18/11.
https://doi.org/10.1145/3276774.3276780

1 INTRODUCTION

Urban Mobility Model (UMM), i.e., estimating the population in urban locations, has long been a research topic with a wide range of real-world applications, such as urban planning [24], transportation engineering [21], epidemic disease control [27], to name but a few. Mobility models were also developed to assist computing applications, e.g., the study and evaluation of mobile ad hoc network protocols [1, 19, 28] as early as 1999. Then, mobility models were developed along with the emergence of such applications, like vehicular networks [40], cellular networks [35] and online social networks [39]. The necessity of a mobility model is two-fold. The first is using the model as a simulator to generate traces to evaluate mobile computing designs [5, 19, 22], e.g., the scalability of a protocol. In contrast, using real-world traces cannot freely set, generate, and change parameters. The second is using the model for mobility prediction [2, 13]. A generative model is useful when data are expensive or even not possible to collect; and as a result, discriminative data-driven models become not suitable. For example, when buildings are planned to be constructed, there is no trace in the area. Thus a generative model can be used, at least as a first mean, to support decision makings.

All previous UMMs have an implicit assumption that people are moving in a 2D space. While the corresponding urban applications such as base station planning in cellular networks, bus route optimization, and traffic congestion prediction may only require the population in a 2D location at a time, the moving and living patterns of people involve the 3D space, i.e., buildings. We argue that the accuracy of UMMs in the 2D space can be heavily affected by such behaviors involving buildings. Elon Musk once stated "having 2D streets and [living in] 3D buildings means bad traffic." [14].

In this paper, we for the first time conduct a comprehensive study on the impact of buildings on UMMs. We start from an analysis on two real-world datasets: a trace of connections from a major cellular carrier and a trace of the indoor population of a building and outdoor traffic status. Our key observation is that human patterns of movement among urban sites are affected by buildings, with buildings being able to "temporarily hold" human mobility, causing both the staying time (which indicates the speed of mobility, i.e., the greater the staying time, the smaller the movements of the people) and the population (which directly shows the number of people in a location) in a building to be significantly different from what would be observed on a road. We innovatively capture the observed properties by extending Markov processes, which have been widely used in developing UMMs, with *semi-absorbing states*. We then develop a Semi-absorbing Urban Mobility model (SUM).

We theoretically analyze SUM. We prove that SUM analytically differs from existing UMMs in two properties of a mobility model,

Spatial \ Temperal	Minutely-scale	Hourly-scale	Monthly-or-beyond-scale
Single-location	[5, 11]	[13, 30]	-
Intracity	mPat [37], [9, 16]	WHERE [18], [6]	[4, 31]
Intercity	-	Gravity [32], [20]	[25, 43]

Table 1: Classification and examples of UMMs of different spatial and temporal granularity.

the staying time and population. We further prove that such difference shrinks and SUM converges to existing UMMs if the time scales and space scales, e.g., a space scale of intercity or a time scale of days or longer. Theoretically this quantifies the scope of SUM; and practically, this nicely reflects the impact of buildings, i.e., buildings may have impact on intracity traffic congestions, but may not have visiable impact on intercity aviation planning.

We evaluate SUM by using SUM as a simulator to generate traces for the cellular network resource allocation, an intracity and hourly application where SUM fits well. In such applications UMMs are widely used in the planning tools such as PLANET of MSI Plc. or PEGASOS of T-Mobile [15] to generate the dynamics of bandwidth demands. We show that SUM can generate traces that are more accurate than two state-of-the-art UMMs, the random waypoint model and weighted random waypoint model by 1.18 times.

We further conduct a real-world case study of our SUM model in the city of Tianjin, China, considering a scenario in which a new building is constructed in the targeted region and using our model to support different phases of the allocation of resources for a cellular network. We show that with our SUM model, the performance of real-world resource allocation substantially improves. For example, in the frequency allocation phase, we see that, assisted by our SUM model, the packet loss probability drops by 3.19 times.

Our contributions can be summarized as follows: (1) By analyzing two real-world datasets, we observe the phenomenon that buildings affect mobility (Section 3). (2) We develop an urban mobility model, SUM, that for the first time captures/explains such a phenomenon. Our model innovatively extends previous UMM models with semi-absorbing states. We analytically prove the properties of our SUM model (Section 4). (3) We study a real-world application, namely, cellular network resource allocation, and demonstrate how it can be better supported by our SUM model (Section 5).

2 RELATED WORK

We can classify UMMs according to spatial and temporal granularities (see Table 1). Along the spatial granularity, there are intercity, intracity, and single-location UMMs. Along the temporal granularity, there are monthly-scale, hourly-scale, and minute-scale UMMs. This naturally classifies different applications. For example, intercity UMMs are useful for the aviation industry, and monthly-scale UMMs are useful for long-term city planning.

Our SUM fits the intracity, hourly-scale UMM [6, 18]. Intrinsically, this applies to human behaviors in buildings, which have an impact on an intracity and hourly scale. In our analysis, we quantify the scope of SUM and prove that the impact of our model can naturally converge into intercity-scale and/or monthly-or-beyond-scale UMMs, as space and time scale up (Section 4.3). Nevertheless, for an UMM in the scope of the intracity and hourly scale, ignoring the impact of buildings would easily lead to inaccuracies in the models, which is also demonstrated in this paper by a comparison of UMMs. Examples of applications that fit this scope include cellular network resource allocation [35], online social network content sharing [39], and bus route optimization [8].

The common mathematical tools to formulate a mobility model include gravity models [20, 32], graph-based models [1, 19], and Markov process [3, 9, 23].

Gravity models say that the mobility between two sites is stronger when they are closer or the population in the two sites are greater. This emulates Newton's gravity law. Gravity models work well for the aggregate behavior of massive population. As a result, gravity models are commonly used for intercity UMMs. The mobility of people in individual buildings does not show significant aggregate behavior. Therefore, gravity models are not suitable.

Graph-based models have been used to model vehicle mobility. These models potrait the road system of a city, where links are the roads and nodes are the intersections of roads. In these models, buildings are usually treated as obstacles, and it is not possible to integrate buildings as a combined part in these models.

Our model falls under a stochastic Markov process. A typical Markov process has transient states as well as absorbing states. However, the traditional absorbing state keeps the population in the site forever. In this paper, we innovatively develop semi-absorbing states since both transient states and absorbing states cannot capture the "temporary holding" property of buildings. Our proposed semi-absorbing approach not only captures a new property, but also confirms and naturally captures many important intracity discoveries in earlier studies, e.g., "various average staying times" [3], "urban as a mesh" [12], and "the attraction of highly populated areas" [18].

Besides generative models, there are discriminative models [37, 38]. These models are data-driven and usually achieve a high level of accuracy in predicting mobility for individual scenarios. They rely heavily on the availability of a large amount of data for training to fit different cities. As previously mentioned, these models are less useful for the simulation of traces, especially when data are not available.

3 IMPACT OF BUILDINGS: THE MOTIVATION

There are two important metrics that are usually considered in a typical UMM, namely, *staying time* and *population*. *Staying time* is the length of time from when a person stops in a location until s/he starts to move again, which serves as a direct metric on the tendency to move. *Population* shows the number of people in a location, which reflects the accumulated effect of the movement of people. In this section, we present our extensive trace analysis in terms of staying time and population with two different real-world datasets, so as to better understand the impact of buildings on these two important metrics.

From two datasets, we obtain information on the population of crowds in each location, involving residential and commercial buildings, every hour. In the following sections, we refer to these two datasets as the *CellularD* and *TrafficD* datasets, respectively. The datasets are now briefly introduced below.

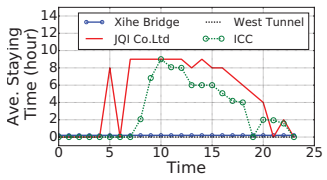

Figure 1: Average staying time for four sites in two datasets as a function of time.

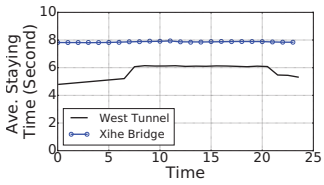

Figure 2: Average staying time on two roads in two datasets as a function of time.

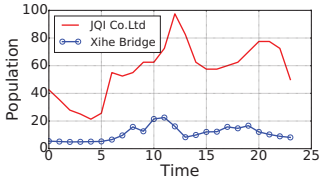

Figure 3: Population of JQI Co.Ltd and Xihe Bridge as a function of time.

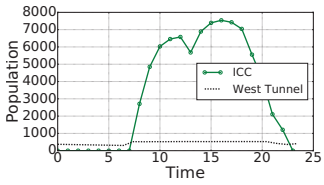

Figure 4: Population of ICC and West Tunnel in *TrafficD* dataset over time.

(a) (b)

Figure 5: Five base stations in *CellularD* dataset.

***CellularD* Datasets.** The first dataset consists of mobile traces collected from five base stations of a dominant carrier of China Mobile Ltd. in the city of Tianjin, China in August 2016. As shown in Fig. 5, the stations cover an area of 2500 × 2500 square meters in the downtown, where a office building of Jinqiao Investments (JQI) Co.Ltd and the Xihe bridge are located. The dataset records time stamps, the MAC of people's mobile devices, connected base stations and the locations of these devices. [1]

***TrafficD* Datasets.** The second dataset includes both outdoor traffic traces and indoor populations from May to August 2016 in the city of Hong Kong SAR, measured with traffic sensing and occupancy sensing systems installed with the help of the local government. The traffic sensing system traces using camera detectors and the occupancy sensing system uses with CO_2 detectors. The dataset covers 617 roads all over the city, including West Tunnel, and covers a 118-story Skyscraper of International Commercial Center (ICC), where more than 10,000 people work.

Next, we show that although the two datasets are for different applications and from different regions with various population densities, we surprisingly obtain consistent observations, which demonstrate: 1) significant statistical differences brought about by the buildings and 2) their impacts. We conducted experiments on other buildings and roads in the two datasets. We see that these observations are not only consistent with common sense, but also with real datasets on different views in different cities. Here we only use the result of the two roads and two buildings to represent that of both types of mobility.

Fig. 1 and Fig. 2 show the average staying time for the *CellularD* and *TrafficD* datasets, respectively. On the one hand, it is quite apparent that people on roads will leave in quite a short time, e.g., several seconds, which is consistent with previous studies on UMMs. We use the term *transient* to describe this situation. On the other hand, we see that a number of people in buildings will stay inside for more than one hour at least. During this period, these people will not show any signs of leaving the building. We will therefore use another term in the next section to denote their status.

Moreover, on the road, people are all transient and the population remains stable; but buildings tend to have two types of people and their populations can change dramatically over time, e.g., during office hours and in off-work hours. Fig. 3 and Fig. 4 show how the population changes over time in the two datasets. It would seem that a certain proportion of the population are "temporarily held" in buildings before they leave, a situation that previous UMMs could hardly capture. In short, we made the following two observations:

Observation 1. *(Staying Time) Both the average and variance of the staying time IN A BUILDING can be much greater than those corresponding values ON A ROAD.*

Observation 2. *(Population) The average and variance of the population IN A BUILDING are much greater than those corresponding values ON A ROAD.*

Due to the large population in buildings, such behaviors can have a significant influence on intracity applications such as cellular and transit networks, as shown in the following sections. To effectively capture such phenomena, we develop our semi-absorbing UMM in the next section.

4 SEMI-ABSORBING UMM

Before presenting the Semi-absorbing Urban Mobility (SUM) model, we first define semi-absorbing sites that capture the "temporal holding" property that previous UMMs could hardly capture. We then develop our SUM model based on the Markov process and translate the semi-absorbing site into a novel semi-absorbing state in the Markov process.

We also present some of the properties of our SUM. In particular, with regard to two important UMM metrics, i.e., staying time and population, significant analytical differences exist between our SUM and traditional UMMs, which consider all sites to be transient (referred to as transient-only UMM in the remainder of this paper). We also present the scope of the SUM. One suggested practice is to use our model in an intracity and hourly scale. In our analysis, we quantify the scope of the SUM and prove that, with regard to impact, our model can naturally converge with that of transient-only UMMs with the scaling up of space and time.

We use the SUM to generate traces and conduct evaluations. The results show that our approach can significantly improve the generation and simulation of traces on a city scale.

4.1 Semi-absorbing Sites

We first introduce a core concept of "semi-absorbing sites" that will be used later. We classify urban sites into three categories: *transient sites*, *absorbing sites*, and *semi-absorbing sites*. Let $p_t^{ij} \in [0, 1]$ denote the transition probability of one person moving from urban site i to site j at time t. Given p_t^{ij}, a *transient site* is an urban site where

[1] The coordinate of the mobile devices is computed as the average coordinates of their connected base stations.

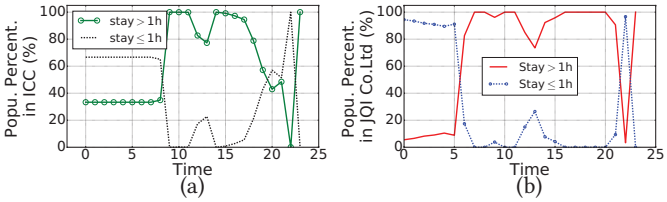

Figure 6: Percentages of two types of population as a function of time, (a) in JQI Co.Ltd of *CellularD*; (b) in ICC of *TrafficD*.

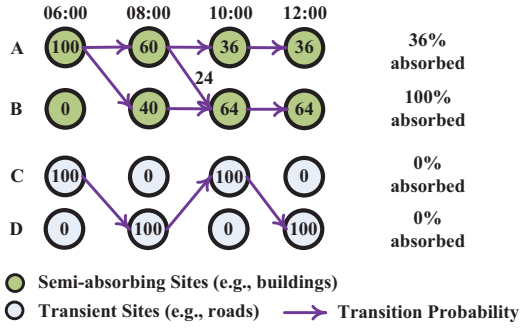

Figure 7: Examples of a semi-absorbing site (A), an absorbing site (B), and transient sites (C and D).

for each person in site i, there exists an urban site j ($j \neq i$) such that $p_t^{ij} > 0$. On the other hand, an *absorbing site* is an urban site where all of the people will stay after arriving and never leave, i.e., $p_t^{ij} = 0$ for all urban sites $j \neq i$ and $p_t^{ii} = 1$.

In addition, we refer to those people who might move to surrounding sites as *transient people* and to those who will stay inside a site for a long period as *absorbed people*. Recall that in Section 3, we showed that the statistics of urban buildings and roads differ, e.g., in terms of staying time and population. In Fig. 6, we further examine the proportion of transient and absorbed people. It is easy to see that for both the *CellularD* and *TrafficD* datasets, such urban sites as buildings usually have both transient people and absorbed people. We thus call these sites *semi-absorbing sites*. In general, for absorbed people, a semi-absorbing site behaves like an absorbing site; while for transient people, it behaves like a transient site.

Fig.7 shows examples of a semi-absorbing site, an absorbing site, and a transient site. Consider a set of sites $\{A, B, C, D\}$ with an initial population of (100, 0, 100, 0), respectively. Whenever people pass the semi-absorbing site of building A in Fig. 7, 36% of them, e.g., the staff, remain in A without any probability of going to another site, while the other 64%, e.g., visitors, do not stay in A but in the end go to building B. In other words, of those people who arrive in building A, 36% are absorbed and the other 64% are transient. However, whenever people pass the absorbing site of building B, all of them are absorbed; while when people pass the transient sites of roads C and D, none of them are absorbed. Note that a single probability, e.g., 36% in the semi-absorbing site A, is sufficient to represent such a "partial temporary holding effect." Merely using transient sites with a looping back probability can easily lead to an "all holding" or "unexpected leaving" situation. For example, either the 64 people who left were also accidentally absorbed in building A or some of the 36 people who stayed left in the middle of the time. Therefore, merely using the transient model will lead to the difficulty of differentiating between those who were absorbed and those who were transient in one site.

4.2 SUM Model

Our SUM model is essentially an extension of Markov processes, and thus can also be defined based on Markov processes, as follows. Consider a set of urban sites $\mathcal{M} = \{1, 2, ..., m\}$ with the total population N. We assume that the whole period of time to be considered is $[0, K]$, which is further divided into L slots with each slot of length K/L.

Site State and Model State. We define a Site State $N_t^i \in \mathbb{N}$ as the urban population in site $i \in \mathcal{M}$ at time slot $t \in \{1, 2, ..., L\}$. The Model State at time t can thus be represented by the sequence vector $\boldsymbol{N}_t = (N_t^1, \cdots, N_t^m)^T$, which gives the population vector among all sites at time t. The initial Model State given by the user is thus denoted by $\boldsymbol{N}_1$.

Transition Matrix. The transition matrix is essential for inferring the total population at different times, considering their different behaviors. It can be defined as $\boldsymbol{\pi}_t = [p_t^{ij}]_{m \times m}$ ($i, j \in \mathcal{M}$), which denotes the probability matrix for people to transit from site i to site j at time t. Note that a site i is considered as absorbed only if $p_t^{ii} = 1$. When $p_t^{ii} \in (0, 1)$, people could remain in site i for a certain period of time, but would still have a non-zero probability of leaving and thus would not be considered to have been absorbed.

Absorbing Proportion and Sub-states. Our SUM model extends the Markov processes by dividing the Site States into the absorbed and transient sub-states and inferring a Site State with different behaviors from those of the two types of populations (i.e., absorbed and transient populations). To this end, we define $\lambda_t^i \in [0, 1]$ as the absorbing proportion, i.e., the probability of a person becoming absorbed when in site i at time t. We denote V_t^i as the number of absorbed people in site $i \in \mathcal{M}$ at time t, and $\overline{V}_t^i$ as the transient population, i.e., $V_t^i = N_t^i \cdot \lambda_t^i$, $\overline{V}_t^i = N_t^i \cdot (1 - \lambda_t^i)$ and $N_t^i = V_t^i + \overline{V}_t^i$.

Model State Transition. In the proposed SUM, the transition matrix and the absorbed and transient sub-states at time t work together to infer the Model State $\boldsymbol{N}_{t+1}$. In particular, let $\boldsymbol{V}_t$ and $\overline{\boldsymbol{V}}_t$ respectively denote the vectors of V_t^i and $\overline{V}_t^i$ at time t in different sites. Given $\boldsymbol{N}_t$ and $\boldsymbol{\lambda}_t$, vectors $\boldsymbol{V}_t$ and $\overline{\boldsymbol{V}}_t$ can be calculated correspondingly. The absorbed people V_t^i at site i will then stay, while the transient people $\overline{V}_t^i$ will move based on the corresponding transition matrix. Finally, the Model State $\boldsymbol{N}_{t+1}$ is computed by:

$$\boldsymbol{N}_{t+1} = \boldsymbol{V}_t + \overline{\boldsymbol{V}}_t \cdot \boldsymbol{\pi}_t$$

Difference between SUM and Transient-only UMM. Existing UMMs merely consider all sites as transient, which we thus refer to as Transient-only UMMs. With the notations defined above, the Model State $\boldsymbol{N}_{t+1}$ in the transient-only UMM model can thus be computed by $\boldsymbol{N}_{t+1} = \boldsymbol{N}_t \cdot \boldsymbol{\pi}_t$. Note that, different from transient-only UMMs, the status of people in our SUM, i.e., transient or absorbed, could be different and correlated with each other. We can further capture buildings of different natures by setting the absorbing proportion of λ_t^i. For example, residential buildings such as apartments are likely to be sparsely populated during normal office hours. On the other hand, in the case of commercial buildings such as offices, it is likely that most of the staff will stay inside the building only during office hours. More specifically, the status changes gradually across different sites $i \in \mathcal{M}$ and at different times $t \leq L$.

4.3 Properties of the SUM Model

1. Difference between semi-absorbed and transient. We next prove that our SUM well captures statistical properties found in Section 3, which existing UMMs cannot capture.

1.1. Staying Time. We first present two lemmas and the notation system, which are used to prove our theorem. Let $\bar{h}^m$ denote the average staying time in semi-absorbing urban sites using mobility model m. Let $\sigma^2(h)^m$ denote the variance in the staying time in semi-absorbing urban sites using mobility model m. Let $\varepsilon \in [0, L]$ denote the upper bound of the staying time in transient sites, which, in practice, is a small value, e.g., 0.00278 hour (ten seconds) on a road. Let $\bar{\boldsymbol{\lambda}}$ denote the average value of matrix $\boldsymbol{\lambda} = [\lambda_t^i]$. Let $|\phi_{on}|, |\phi_{off}|, |\phi_{on-off}| \in [0, L]$ denote the average length of the time interval for semi-absorbing site i to stay absorbing ($\lambda_t^i = 1$), transient ($\lambda_t^i = 0$), and semi-absorbing ($\lambda_t^i \in (0, 1)$), respectively. Let ρ denote the *semi-absorbing density*, which is the proportion of semi-absorbing sites that currently are not transient during the considered time interval, i.e., for site $i, \exists \lambda_t^i > 0$. Let $\delta > 0$ denote the average staying time in sites of SUM when $\lambda_t^i \in (0, 1)$, which is computed with the parameter of absorbing proportion λ_t^i multiplied by a constant, i.e., $\delta = |\phi_{on}|/L \sum_{t\in[1,L]} \lambda_t^i$. We thus have the following two lemmas:

Lemma 1. *In the semi-absorbing site of SUM, the average staying time satisfies $\delta \cdot |\phi_{on}|/L < \bar{h}^{SUM} < L - |\phi_{off}| + \varepsilon \cdot |\phi_{off}|/L$. In the transient site of SUM, the average staying time is smaller than ε.*

Proof. Due to page limitations, all missing proofs and lemmas can be found in the technical report [42]. □

Lemma 2. *In the semi-absorbing sites of SUM, $\sigma^2(h)$ satisfies $(\delta - \varepsilon)^2 \cdot |\phi_{off}|/4L < \sigma^2(h)^{SUM} < (L-\varepsilon)^2 \cdot |\phi_{on}| \cdot 4/L$. In the transient sites, the variance is 0.*

From Lemmas 1 and 2, we derive the following theorem:

Theorem 1. *Given the same urban sites $\mathcal{M}$, the average and variance of the staying time of a SUM model are strictly and significantly larger than those of a transient-only UMM model. Both differences are lower bounded by the semi-absorbing density ρ or by the average absorbing proportion $\bar{\boldsymbol{\lambda}}$.*

1.2. Population. We introduce the following lemmas for the next theorem. $\bar{\boldsymbol{N}}^m$ denotes the average population in semi-absorbing sites with mobility model m. $\sigma^2(\boldsymbol{N})^m$ denotes the variance of the population in semi-absorbing sites with mobility model m. n^i denotes the population in site i. The proportion of the population in semi-absorbing sites in $\phi_{off}, \phi_{on-off}, and \phi_{on}$ are denoted as r_{off}, r_{on-off}, and r_{on}, respectively. We thus have the following two lemmas:

Lemma 3. *In a semi-absorbing site i of SUM, the average population n^i satisfies $n^i > \lfloor r_{on} \rfloor \cdot |\phi_{on}| \cdot N/L$. In transient sites, the average population n^i_{tran} satisfies, $0 \le n^i_{tran} < N \cdot \varepsilon \cdot m^t_{tran}/(\delta \cdot m^t_{on} + \varepsilon \cdot m^t_{tran})$.*

Lemma 4. *In a semi-absorbing site i of SUM, the variance of the population satisfies $\sigma^2(n^i) > [(r_{off} - \lfloor r_{on} \rfloor \cdot |\phi_{on}|/L)^2 |\phi_{off}| + 2(r_{on-off} - \lfloor r_{on} \rfloor \cdot |\phi_{on}|/L)|\phi_{on-off}|] \cdot N^2/L$. In transient sites of SUM, the variance of the population can be taken as $\sigma^2(n^i) = 0$.*

From Lemmas 3 and 4, we can derive the following theorem:

Theorem 2. *Given the same urban sites $\mathcal{M}$, the average and variance of the population of a SUM model are strictly and significantly larger than that of a transient-only UMM model. Both differences are lower bounded by the semi-absorbing density ρ or by the average absorbing proportion $\bar{\boldsymbol{\lambda}}$.*

Brief Summary. From Theorem 1 and Theorem 2, we can infer that the difference between SUM and transient-only UMM on the average and variance of the staying time and population will both increase monotonically as the semi-absorbing density ρ or the average absorbing proportion $\bar{\boldsymbol{\lambda}}$ increases. In general, by adjusting λ_t^i in different scenarios, we can control different behaviors, especially semi-absorbed behaviors, which is useful when SUM is integrated with an urban mobility simulator to generate synthetic data traces. As an illustration, one simulation of urban sites will be shown in Section 4.4.

Moreover, both theorems demonstrate that the behaviors of SUM well match our Observation 1 and Observation 2 in Section 3, where great differences can be seen in the average and variance between semi-absorbing sites (e.g., buildings) and transient sites (e.g., roads), in terms of staying time and population.

1.3. Clustering Phenomenon. The SUM also captures a clustering phenomenon. A cluster Φ denotes a connected subgraph of $\mathcal{M}$, which is filled with semi-absorbing sites and surrounded with transient border sites $B \in \Phi$, and N^Φ denotes the original population in Φ. P^{ib} denotes nodes on the path of a random walk from a site $i \in \Phi$ in the cluster to border site $b \in B$; p^{ib} denotes the probability that path P^{ib} will be taken, and $||P^{ib}||$ denotes the distance to the border site through path P^{ib}. p^{iB} denotes the probability that a random walk will start from i and end at border sites B, i.e., $p^{iB} = \sum_{b\in B}(1-\rho)^{||P^{ib}||} p^{ib}$. Let p_o denote the probability that the transient will leave from the border to sites out of Φ. Let I denote the identity matrix. We thus have:

Theorem 3. *For N^Φ people in a cluster Φ, the staying population is $N^\Phi \cdot [1 - p_o \cdot (I - \bar{\boldsymbol{\lambda}}) \cdot \sum_{b\in B} p^{ib}(1-\rho)^{||P^{ib}||}]$.*

Proof. Only those people who are not absorbed can leave a semi-absorbing site. Some of them will reach the border sites B without being absorbed, i.e., $p^{iB} = \sum_{b\in B}(1-\rho)^{||P^{ib}||} p^{ib}$. The probability of the population leaving the cluster is $p^{iB} \cdot p_o \cdot (I - \bar{\boldsymbol{\lambda}})$ and that of the population staying in the cluster is $[1 - p^{iB} \cdot p_o \cdot (I - \bar{\boldsymbol{\lambda}})] \cdot N^\Phi$. □

Brief Summary. Theorem 3 reflects the fact that in SUM, people start from the building and tend to move or stay in the surrounding area. The population can vary slightly within the area due to the stable $\bar{\boldsymbol{\lambda}}$, while dropping sharply outside the area as the $\bar{\boldsymbol{\lambda}}$ is reduced to 0 with transient sites on the border. In other words, people tend to stay in a certain urban area and are unlikely to embark on long-distance travel, and unlikely to propagate to the end of the world. But a transient-only UMM does not have such a property, which will be demonstrated in experiments presented in Section 4.4.

To sum up, Theorem 1 - 3 capture the difference between absorbed and transient behaviors, which significantly affects application designs; e.g., absorbed people in buildings tend to use WiFi and transient people tend to use the cellular network. Application

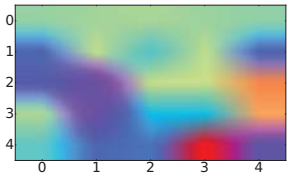

Figure 8: Real-world population.

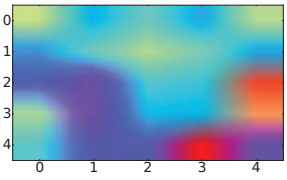

Figure 9: SUM population.

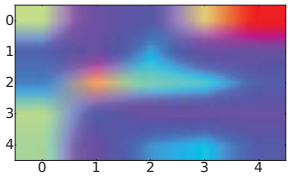

Figure 10: RWM population.

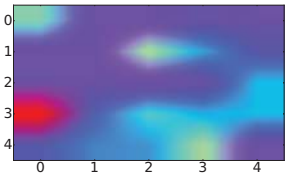

Figure 11: WRWM population.

designers can leverage SUM to facilitate differentiation of cellular resource allocations, which will be demonstrated in Section 5.

2. Scope of SUM. We also present the scope of our SUM model to investigate how the impact of semi-absorbing sites changes according to the scaling of time and space.

First, the SUM model was originally designed for application on an intracity-scale with sites as the unit. $\mathcal{M}'$ denotes a set in which the unit has been scaled up to regions, with each single region in $\mathcal{M}'$ covering multiple sites in $\mathcal{M}$, i.e., $\mathcal{M}' \preceq \mathcal{M}$. ρ' denotes the average proportion of semi-absorbing sites (e.g., buildings) in each aggregated site (or region) of $\mathcal{M}'$. $\bar{h}_M^m$ denotes the average staying time in semi-absorbing urban sites using mobility model m with urban sites M.

Theorem 4. *Given the same urban site set $\mathcal{M}' \preceq \mathcal{M}$, for a SUM model and a transient-only UMM model, we have $\bar{h}_{\mathcal{M}'}^{SUM} - \bar{h}_{\mathcal{M}'}^{UMM} < \rho'(L-\varepsilon)(L-|\phi_{off}|)/L + \varepsilon$.*

Proof. Theorem 4 can be inferred with Lemma 1. □

Second, the SUM model was originally designed for application on an hourly scale. When we divide time slots with an interval of $K/L' \geq K/L$, the whole time period $[0, K]$ is then divided into $L' < L$ slots.

Theorem 5. *Given the same urban site set $\mathcal{M}$ and the sampling interval of K/L', $L' \to 1$, for a SUM model and a transient-only UMM model, we have $(\bar{h}_{L'}^{SUM} - \bar{h}_{L'}^{UMM}) \to (\bar{h}_1^{SUM} - \bar{h}_1^{UMM}) < L' - (1-\varepsilon/L') \cdot |\phi_{off}|$.*

Proof. Theorem 5 can be inferred with Lemma 1. □

Brief Summary. With the two theorems, we clearly justify whether and to what extent SUM shows that it is different from transient-only UMMs given any sampling granularity in specific datasets. Theorem 4 indicates that, when scaled up from a site-as-unit design to a region-as-unit one, if $\rho' \to 0$ or $\lambda_t^I \to 0$ in the given $\mathcal{M}'$, SUM then converges to a transient-only UMM. Theorem 5 indicates that, when scaled up from an hourly design to a multi-hourly one, if $L' \to 1$ and $\lambda_t^i \to 0$, SUM also converges to a transient-only UMM. A suggested practice is to use SUM on an intracity and hourly scale. Considering that there are constant changes in the number of buildings and the population inside them in a city, the difference resulting from the use of SUM will be clear.

4.4 Evaluation on Urban Trace Generation

Having only limited access to real-world data traces can be a frustrating issue for many researchers. Therefore, the ability to create synthetic, yet realistic, urban mobility traces will have a wide range of implications for the scientific community. Our SUM will enable a larger community to perform a wider range of experiments, with the assurance that the results will mimic those obtained from real-world data traces. Towards this aim, we evaluate our SUM model.

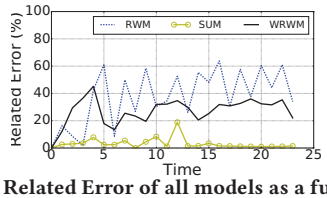

Figure 12: Related Error of all models as a function of time.

We compare our model with two typical UMMs widely used in intracity scenarios, namely, the Random Waypoint Model (RWM) [3], which serves as a baseline, and the state-of-the-art Weighted Random Waypoint Model (WRWM) used in [18]. Both models are transient-only UMM with staying time parameters. To ensure fair comparisons, we extend the RWM and WRWM a little bit by setting different parameters on roads and buildings. All of the evaluated models have been calibrated using historical data.

Real-World Similarity. With effect of multiple buildings, the *CellularD* Dataset is used for the evaluation of real-world similarity. It covers the central city area with 5×5 grids as presented in Section 3. We adopt the most commonly used evaluation criterion of related error to compare the performance of these models in generating urban traces. Related error is defined as $\frac{1}{L}\sum_{t=1}^{L} |\boldsymbol{N}_t - \hat{\boldsymbol{N}}_t| / \boldsymbol{N}_t$, where $\hat{\boldsymbol{N}}_t$ denotes the synthesized population of people at time t and $\boldsymbol{N}_t$ denotes the corresponding real population of people.

To demonstrate how similar the various UMMs are to the real-world dataset, in Fig. 8 to Fig. 11 we first show the heat map of the real-world, SUM, RWM, and WRWM populations for ICC and the nearby roads at 22:00. The warmer the color of a location, the larger the population that it holds. We see that our SUM model can better capture the concentrated population of the building in the middle without too much noise from the nearby roads, which matches well with the real situation shown in Fig. 8. More specifically, Fig. 12 shows the related error of the urban traces generated by different UMMs. On average, SUM has a related error that is a 73.68% improvement over RWM and 63.10% over WRWM. This clearly shows the effectiveness of SUM, thanks to the additional modeling of semi-absorbing sites.

City Clustering. We also examined the correlation between the population in a building and those in the surrounding roads, where the roads have different distances from the building. We compute the correlations for ICC and the surrounding roads as shown in Fig. 13 (a); the results for SUM, RWM, and WRWM are shown in Fig. 13 (b) - (d). The results show the existence of the clustering phenomenon. In essence, the correlation is stable within a range of area, which matches well with the clustering phenomenon derived from our theoretical analysis in Theorem 3. We also note that the average correlation drops significantly to below 30% over 12 km away. This indicates that the dynamics of the population from ICC have little effect on the population of roads over 12 km away. Such an observation is also confirmed in the *CellularD* Dataset. However, this phenomenon cannot be well modeled in the RWM

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

Figure 13: Correlation of between the mobility at present and after 24 hours in (a) real-world *TrafficD* dataset; (b) SUM; (c) RWM; (d) WRWM.

and WRWM due to their homogeneous nature and steady states, i.e., if a sufficiently long time is given, a citizen can stay in any site on the map. The clustering phenomenon reveals that this conclusion about existing models might not be true in reality.

5 CASE STUDY ON CELLULAR NETWORK RESOURCE ALLOCATION

Our SUM model not only demonstrates a great improvement in performance, but is also of much practical value for applications that are sensitive to absorbed and transient crowds. In this section, we conduct a real-world case study on cellular network resource allocation, to further illustrate the practical usefulness of our SUM model. To this end, we first use the SUM model as a simulator to evaluate network demand and cross-boundary frequency for the allocation of resources in cellular networks. More specifically, we generate synthetic traces using both SUM and previous transient-only UMMs. In addition, we show that when SUM is used, the design of base station positions and channel allocations can differ from those devised using a transient-only UMM model. We demonstrate that the differentiated or even dynamic allocation of urban resources is a significant improvement over previous homogeneous over-allocated designs. This shows that more caution should be exercised towards the design of resource allocations in cellular networks, as a design that is evaluated as successful using the transient-only UMM model may not necessarily represent the real-world case.

5.1 Overview

Briefing on Resource Allocation in Cellular Networks. As more and more new mobile cellular network services are becoming available to larger audiences, there is an ever-increasing demand for high-quality wireless access under different scenarios. This demand is usually met via *cellular network resource allocation*, which is the process of ensuring that sufficient network resources, e.g., channels, are provisioned, such that the committed core network targets of delay, jitter, loss, and availability can be achieved [10].

The growth of the mobile cellular market imposes a difficult task on designers of mobile systems. The performance of a cellular network can be noticeably affected by changes in the urban environment. In particular, cellular networks are sensitive to the user population within each cell and to their mobility across cells, which can be better captured by our proposed SUM. In particular, the conventional approach used by cellular network designers to evaluate performance, which is widely used in today's commercial cellular planning tools such as PEGASOS of T-Mobil, or PLANET of MSI Plc., is based on the so-called analytical approach [15]. To this end, we implement a SUM-based simulator that has the potential to be integrated into these tools and to help with the production of more practical designs relating to network and capacity issues.

Application Scenario. The growth of the mobile cellular market imposes a difficult task on designers of mobile communication systems. The performance of the cellular network can be noticeably affected by human behaviors in different urban environments. From this point onwards, our SUM model fits such applications well, due to its ability to differentiate between urban mobilities.

For example, in cellular network applications, the absorbed people in a building tend to use WiFi more than transient people; large buildings and high-rise buildings are also examples of structures where mobile phones cannot properly reach the macro or outdoor network of a carrier [29]. In both cases, the result can be a drop in cellular data traffic in buildings. Considering human behaviors in different environments, it is apparent that the same allocation of cellular resources can lead to completely different performances, e.g., in terms of average throughput and signal strength, even with the same population in the cellular network.

In another example, when buildings are newly constructed, the movement of the surrounding population will change accordingly, causing additional people to possibly enter the cell and thereby affect the performance of the original network, e.g., signal strength. Thus, the performance of the original network needs to be re-evaluated so that resources can be reallocated accordingly.

Three Phases in a Case Study. In principle, the analytical approach consists of three phases, namely, *Traffic Demand Analysis*, *Radio Network Definition*, and *Frequency Allocation*, which passes through several turns iteratively. In the next subsection, we will show the results of a network simulation of a traffic demand analysis. This is followed by a demonstration of the further use of the other two phases of our SUM model. All are verified by our real-world datasets.

Traffic Demand Analysis: First, the tele-traffic demand within the planning region is derived based on rough estimates of the targeted area. Due to their ability to describe the behaviors of users in detail, *Traffic Source Models*, also referred to as mobility models, are usually applied to estimate the traffic demands in an individual cell of a mobile network. The traffic scenario is represented as a population of individual sources of traffic performing random walks through the service area and randomly generating demand for resources, i.e., the radio channel. Based on the Traffic Demand Analysis, there are two phases in the allocation of cellular network resources: the *Radio Network Definition Phase* to choose cell sites, and the *Frequency Allocation Phase* to allocate channels to stations. Both will be covered in this case study.

Baseline Approach. As for the simulation of the analysis of demand, as in Section 4 we use the state-of-the-art WRWM and the

baseline RWM with the extended setting to determine the interactions of the population between buildings and roads.

For a real-world application of network definition and frequency allocation, we compare the following: a real-world solution, if any (e.g., a solution provided by China Mobile Ltd. (denoted as Real)), the best possible solution based on the estimated traces if the building is still under construction (denoted as Ideal-Before), and the best possible solution to achieve based on real traces (denoted as Ideal). We also compare the performance of SUM with that of the baseline RWM. WRWM cannot be evaluated due to the lack of status and trace information.

Settings. The radio transmission range of a base station is about 200 meters. The fixed capacity of the base station is set as large as 2.87 Gbps as in [36], which is collected for the busiest period of the day, when the load on a base station is at its highest. The average data traffic per person is 2 Mbps as in [33]. We will deploy two more cell stations with all 25 locations feasible.

5.2 Traffic Demand Analysis Phase

Using the baseline and our mobility model, local performance measures, such as fresh call or handover blocking probability, can be derived from the mobility pattern and mobile environments, e.g., as indicated by the network demand and cross-boundary frequency.

Evaluation Metrics. We illustrate how SUM can better emulate different urban mobility environments in simulations for evaluating a cellular network demand analysis. Without loss of generality, we consider two typical metrics widely used for evaluating cellular network resource allocations, namely, network demand and cross-boundary frequency.

Network demand: We evaluate the performance of a network using throughput, which involves measuring the volume of network traffic per second (gigabits/second). With mobility, physically available routes may become invalid (i.e., may no longer be found in the transmission range), causing packets to be dropped and leading to throughput degradation and an increase in control overhead.

Cross-boundary frequency: Cross-boundary frequency refers to the link up/down events caused by user movements across cell boundaries, which mainly happens when a user moves across cells and triggers the hand-off mechanism.

When one person moves into the transmission range of a cell station, a connection is gained. This is a link-up case. Accordingly, when two people previously within the transmission range (assuming that they have same transmission range) move out, the connection is lost. This event increments a link down counter. We evaluate how types of sites affect the link up/down dynamics.

With user mobility, the resources allocated for the network traffic may vary, in that if the allocated resources do not match the demands of the network, packets may be dropped, further leading to throughput degradation and increasing the control overhead.

Experimental Results. Fig. 14 shows the results on network demand, where each person is assumed to generate the same amount of demand. We find that (a) in transient sites, SUM has lower demands than other models in office hours; (b) in semi-absorbing sites, the network demand is higher in SUM for office hours as expected. For example, the network demand in SUM can be 1.18 times of that in WRWM for semi-absorbing sites when the population reaches three thousand. This matches well with the reality

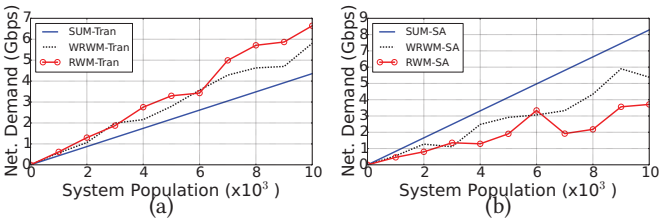

Figure 14: Network Demand vs. Population for office hours (a) in transient sites; (b) in semi-absorbing sites.

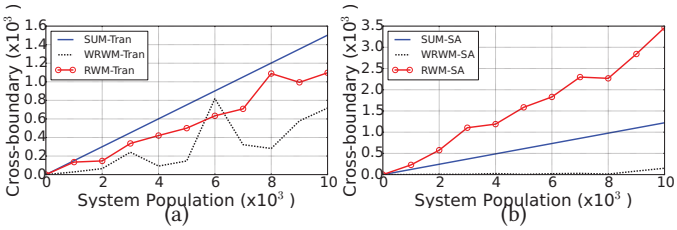

Figure 15: Cross-Boundary Frequency vs. Population for commuting hours (a) in transient sites; (b) in semi-absorbing sites.

that the population is more concentrated in semi-absorbing sites during office hours, although the majority of sites are transient sites. If we consider more realistic scenarios, e.g., people tend to use WiFi when absorbed and use the cellular network when transient, SUM demonstrates even better performance thanks to its ability to differentiate between absorbed and transient, a capability that will be further investigated in Section 5.3 and Section 5.4.

Fig. 15 shows those results relating to the cross-boundary frequency for commuting hours. As expected, (a) SUM shows slightly higher cross-boundary frequencies in transient sites, because staying times in those sites are shorter and people are not likely to stop in those sites; (b) SUM also has a 61.2% lower cross-boundary frequency than RWM in semi-absorbing sites, because the average staying time in semi-absorbing sites is much longer than the arbitrary random transient setting of RWM; (c) the cross-boundary frequency in SUM is 12.4 times of that in WRWM. That is because, different from the rigid settings of WRWM, which keeps holding people during commuting hours, SUM shows a greater variance in staying time during commuting hours, e.g., people can stop staying in buildings and move to roads.

We summarize our findings below. First, in general, more people aggregate in semi-absorbing sites, which might generate more network traffic. Second, cross-boundary behaviors are more likely to happen in transient sites than in semi-absorbing sites. Thus, homogeneous designs based on previous UMMs for different regions and times slots either overestimate link quality or waste resources. Differentiated resource allocation can be conducted among regions and time slots according to a factor of ρ and λ_t^i. For example, to support possible voice call demands, base stations with a smaller capacity and a larger transmitting range should be deployed to a region with a low building density, i.e., with more transient sites; while base stations with dynamic capacities and transmitting ranges can be deployed for regions with a high building density, which again leads to a differentiated design for a cellular network that takes into account the two types of regions. We will show more items that are taken into consideration and the resulting gain in performance based on our SUM, as case studies in Section 5.3 and 5.4.

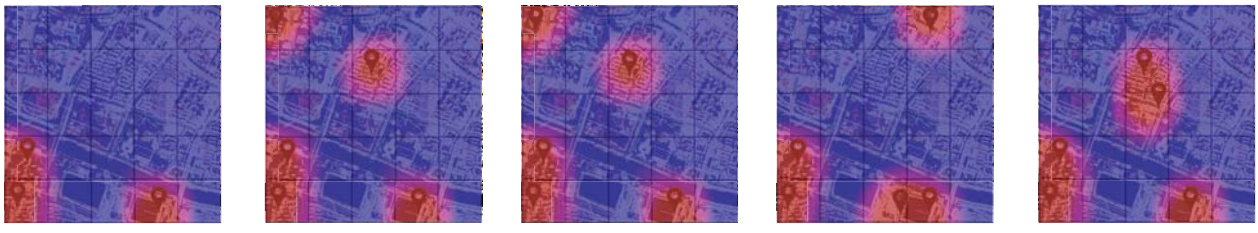

Figure 16: Base Station Positioning (a) Ideal-Before; (b) Ideal; (c) SUM; (d) RWM; (e) Real.

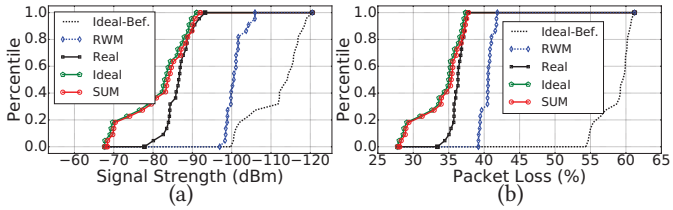

Figure 17: Phase II: CDF of (a) signal strength; (b) packet loss.

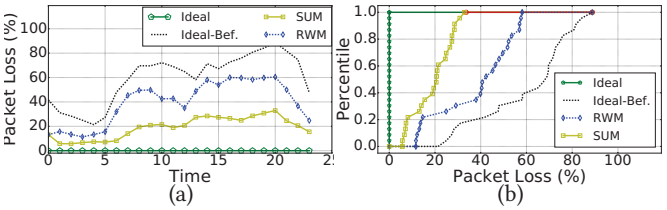

Figure 18: Phase III: (a) Packet loss probability as a function of time; (b) CDF: Percentile as a function of packet loss.

5.3 Radio Network Definition Phase

With the estimated demand, a human expert chooses the cell sites. In order to obtain a regular structure, the popular algorithms of base station positioning [35] [26], which distribute the transmitters on a hexagonal grid, are usually used in this step. Using these transmitter configurations, the expert evaluates the radio coverage using field strength prediction methods. Here, stochastic channel models are applied. Usually, several field strength prediction methods are implemented, but the tools offer little if any support in the task of choosing the appropriate propagation model. If the planning expert decides that the coverage is insufficient, new transmitter positions have to be chosen and the propagation has to be analyzed again.

Evaluation Metric. In the Radio Network Definition Phase, we evaluate two factors that typically affect the deployment of base stations: a) the varying strength of radio signals; b) the heterogeneously distributed network quality in the coverage area of the base stations.

In telecommunications, signal strength refers to the power output of the transmitter as received by a reference antenna at a distance from the transmitting cell station. The received signal strength in a mobile device can be estimated as follows [41]:

$$dBm_e = -113.0 - 10.0 \times \gamma \log_{10}(r/R),$$

where dBm_e denotes the estimated received power in a mobile device; r denotes the distance from the mobile device to the cell station; R denotes the mean radius of the cell station; and γ denotes the path loss exponent.

As in [7], we also apply the metric of packet loss per second, with a packet size of 12 Kb. The relationship between packet loss and signal strength is also studied in [41], as it is generally true that high signal strength corresponds to low packet loss.

Here, the deployment is to minimize the packet loss under a base station positioning design. We apply the state-of-the-art positioning algorithm as in [35], where network demand can be simulated with UMMs when confronted with insufficient information, e.g., when a building is newly constructed.

Experimental Results. Fig. 16 shows the positioning results of base stations, with the stations marked in red and two more base stations to be allocated based on Ideal-Before (i.e., when the building has not yet been established). The RWM gives a more random allocation. SUM provides exactly the same solution as Ideal, which places cell stations in regions where the density of buildings is high, in order to enable more differentiating and dynamic resource allocations than the allocations of RWM and Real.

We next evaluate the signal strength, which is shown in Fig. 17 (a). We can see that the positioning of base stations based on SUM reaches Ideal and always outperforms other schemes, including Real. We also examine the packet loss ratio for each scheme, and the CDF is shown in Fig. 17 (b). The SUM solution again reaches the Ideal, and both the SUM solution and the Ideal outperform Real and RWM. That is because SUM does a good job of differentiating between the absorbed and the transient in different urban sites, and is better at distinguishing and estimating the corresponding users and demands of WiFi and cellular networks. By filtering noises generated by WiFi users, the corresponding base station positioning can, in particular, serve consistent users of the cellular network.

5.4 Frequency Allocation Phase

Dynamic capacity balancing is conducted to allocate channels under fixed locations of base stations, which are given in Fig. 16 (e), so as to minimize the packet loss. There are several ways to allocate capacity for these stations. In this part, we use an approach proposed in [34], which allocates more bandwidth to sites with higher cellular demands, as indicated in the correspondingly used UMM. We conducted experiments taking both data and voice traffic into consideration and observed consistent results. Here we only use the result of data traffic to represent that of both types of traffic.

As in the previous subsection, we use the metric of packet loss per second, with a packet size of 12 Kb. We measure the packet loss in Fig. 18 (a). The capacity balancing solution with SUM is the closest to that with Ideal, and always better than with RWM and Ideal-Before. Specifically, at 10:00, Ideal-Before and RWM have a packet loss probability of 4.18 and 3.19 times that of our SUM. We also measure the corresponding CDF of the packet loss probability under each UMM in Fig. 18 (b). We see that SUM has a similar performance to that of Ideal and is closest to it, outperforming

RWM and Ideal-Before. The reason for this is that the SUM model does a good job of differentiating between the absorbed and the transient in different urban sites at different times. This leads to a better estimation of the users and the demands on WiFi and the cellular network, so that the corresponding dynamic frequency allocation can serve users of cellular network adaptively, e.g., in an hourly manner.

6 DISCUSSION ON PARAMETER ESTIMATION

Parameter calibration for a model is challenging. It depends on the objective, whether to be generic so as to fit more scenarios or to be specific so as to best fit one scenario. It also depends on the data available that can be used for calibration, i.e., with partial status information (e.g., traces on partial sites), without status information but with alternative data source (e.g., map or population). It is our future work to study the parameter calibration of SUM comprehensively.

In this paper, we conduct parameter calibration given urban mobility traces on each site and time slot. Parameters of absorbing proportion can be automatically selected by maximizing the estimation accuracy using dynamic programming. Parameters of transition probability can be computed as the ratio between the average number of traces moving between sites and the number of all feasible traces. Similar techniques have also been used in [17, 18].

7 CONCLUSION

Developing Urban Mobility Models (UMMs) has long been an important research topic. In this paper, we demonstrated that buildings heavily affect human patterns of movement. We developed a new Semi-absorbing Urban Mobility model (SUM). SUM captures intrinsic properties of the impact of buildings and analytically differs from previous UMMs. We carefully clarified the scope of SUM on an intracity and hourly scale. We showed that SUM converges to other UMMs when the scale of space and time changes. Our evaluation showed that as a foundation to support mobile applications, SUM is far superior to previous UMMs. Its performance was further confirmed by a case study of a typical real-world application for cellular network resource allocation.

ACKNOWLEDGMENTS

Parts of the work were supported by the Hong Kong Polytechnic University under Grant No.: 1-BBYX, and by NSF I/UCRC under Grant No.: 1539990.

REFERENCES

[1] Fan Bai, Narayanan Sadagopan, and others. IMPORTANT: A framework to systematically analyze the Impact of Mobility on Performance of Routing protocols for Adhoc Networks. In *IEEE INFOCOM'03*. 825–835.
[2] Paul Baumann and others. How long are you staying?: predicting residence time from human mobility traces. In *ACM MobiCom'13*. 231–234.
[3] Christian Bettstetter, Hannes Hartenstein, and others. 2004. Stochastic properties of the random waypoint mobility model. *Wireless Networks* 10, 5 (2004), 555–567.
[4] Thomas P Boehm and others. 1991. Intra-urban mobility, migration, and tenure choice. *The Review of Economics and Statistics* (1991), 59–68.
[5] Federica Bogo and Enoch Peserico. Optimal throughput and delay in delay-tolerant networks with ballistic mobility. In *ACM MobiCom'13*. 303–314.
[6] I Cameron and others. 2003. Understanding and predicting private motorised urban mobility. *Transportation research part D* 8, 4 (2003), 267–283.
[7] Andrew T Campbell, Javier Gomez, Sanghyo Kim, András Gergely Valkó, Chieh-Yih Wan, and others. 2000. Design, implementation, and evaluation of cellular IP. *IEEE personal communications* 7, 4 (2000), 42–49.
[8] Chao Chen, Daqing Zhang, Zhi-Hua Zhou, and others. B-Planner: Night bus route planning using large-scale taxi GPS traces. In *IEEE PerCom'13*. 225–233.
[9] Xiqun Chen, Li Li, and others. 2010. A Markov model for headway/spacing distribution of road traffic. *IEEE TITS* (2010).
[10] Cisco. 2013. White Paper: Best Practices in Core Net Capacity Planning. (2013).
[11] L da F Costa, BAN Travencolo, MP Viana, and others. 2010. On the efficiency of transportation systems in large cities. *Europhysics Letters* 91, 1 (2010), 18003.
[12] Vanessa Ann Davies and others. 2000. *Evaluating mobility models within an ad hoc network*. Master's thesis. Colorado School of Mines.
[13] Trinh Minh Tri Do and Daniel Gatica-Perez. Contextual conditional models for smartphone-based human mobility prediction. In *ACM UbiComp'12*. 163–172.
[14] Elon Musk. 2015. Twitter: Elon Musk is worried about urban traffic. https://twitter.com/elonmusk/status/580482462612135936. (2015).
[15] A Gamst, E-G Zinn, R Beck, and R Simon. 1986. Cellular radio network planning. *IEEE Aerospace and Electronic Systems Magazine* 1, 2 (1986), 8–11.
[16] Raghu Ganti and others. Inferring human mobility patterns from taxicab location traces. In *ACM UbiComp'13*.
[17] Wei-jen Hsu and others. 2005. Weighted waypoint mobility model and its impact on ad hoc networks. *ACM SIGMOBILE Mobile Computing and Communications Review* 9, 1 (2005), 59–63.
[18] Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, and others. Human mobility modeling at metropolitan scales. In *ACM MobiSys'12*.
[19] Amit Jardosh, Elizabeth M Belding-Royer, Kevin C Almeroth, and others. Towards realistic mobility models for mobile ad hoc networks. In *ACM MobiCom'03*.
[20] Krings, Calabrese, and others. 2009. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment* (2009).
[21] Neal Lathia and Licia Capra. How smart is your smartcard?: measuring travel behaviours, perceptions, and incentives. In *ACM UbiComp'11*. 291–300.
[22] Kyunghan Lee, Seongik Hong, Seong Joon Kim, Injong Rhee, and others. Slaw: A new mobility model for human walks. In *IEEE INFOCOM'09*. 855–863.
[23] Ben Liang and Zygmunt J Haas. Predictive distance-based mobility management for PCS networks. In *IEEE INFOCOM'99*. 1377–1384.
[24] Yu Liu and others. 2012. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning* (2012).
[25] Laudo M Ogura. 2005. Urban growth controls and intercity commuting. *Journal of Urban Economics* (2005).
[26] Fred Richter and others. 2009. Energy efficiency aspects of base station deployment strategies for cellular networks. In *IEEE Vehicular Technology Conference*.
[27] Steven Riley. 2007. Large-scale spatial-transmission models of infectious disease. *Science* 316, 5829 (2007), 1298–1301.
[28] Elizabeth M Royer and Charles E Perkins. Multicast operation of the ad-hoc on-demand distance vector routing protocol. In *ACM MobiCom'99*. 207–218.
[29] Adel AM Saleh and others. 1987. Distributed antennas for indoor radio communications. *IEEE Transactions on Communications* 35, 12 (1987), 1245–1251.
[30] Matthias Schwamborn, Nils Aschenbruck, and others. A realistic trace-based mobility model for first responder scenarios. In *ACM MSWiM'10*. 266–274.
[31] Sevtsuk and others. 2010. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology* (2010).
[32] Filippo Simini, Marta C González, and others. 2012. A universal model for mobility and migration patterns. *Nature* 484, 7392 (2012), 96–100.
[33] Statista Inc. 2016. Average Global Mobile Network Connection Speeds from 2015 to 2020 (in Mbps). (2016).
[34] Kurt Tutschku. Demand-based radio network planning of cellular mobile communication systems. In *IEEE INFOCOM'98*. 1054–1061.
[35] Kurt Tutschku and Phuoc Tran-Gia. 1998. Spatial traffic estimation and characterization for mobile communication network design. *IEEE Journal on selected areas in communications* 16, 5 (1998), 804–811.
[36] Unwired Insight Ltd. 2016. What is the Realistic Capacity of a 3G HSPA, HSPA+ or LTE Network. (2016).
[37] Desheng Zhang and others. 2014. Exploring human mobility with multi-source data at extremely large metropolitan scales. In *ACM MobiCom'14*. 201–212.
[38] Desheng Zhang, Juanjuan Zhao, and others. coMobile: real-time human mobility modeling at urban scale using multi-view learning. In *ACM SIGSPATIAL'15*. 40.
[39] Lei Zhang, Feng Wang, and others. SoCrowd: When Social Content Sharing Meets Mobile Crowd. In *IEEE ICDCS'17*.
[40] Xiaolan Zhang and others. Study of a bus-based disruption-tolerant network: mobility modeling and impact on routing. In *ACM MobiCom'07*. 195–206.
[41] Jerry Zhao and Ramesh Govindan. 2003. Understanding packet delivery performance in dense wireless sensor networks. In *ACM Sensys'03*. 1–13.
[42] Zimu Zheng, Feng Wang, Dan Wang, and Liang Zhang. 2017. Technical report - buildings affect mobile patterns: developing a new urban mobility model. https://goo.gl/hwCc4b. (2017).
[43] George Kingsley Zipf. 1946. The P_1P_2/D hypothesis: on the intercity movement of persons. *American Sociological Review* 11, 6 (1946), 677–686.