A Simple WiFi Hotspot Model for Cities

Michael Seufert, Tobias Griepentrog, Valentin Burger
Institute of Computer Science
University of Würzburg, Würzburg, Germany
seufert@informatik.uni-wuerzburg.de

Tobias Hoßfeld
Chair of Modeling of Adaptive Systems
University of Duisburg-Essen, Essen, Germany
tobias.hossfeld@uni-due.de

Abstract—**WiFi offloading has become increasingly popular. Many private and public institutions (e.g., libraries, cafes, restaurants) already provide an alternative free Internet link via WiFi, but also commercial services emerge to mitigate the load on mobile networks. Moreover, smart cities start to establish WiFi infrastructure for current and future civic services. In this work, the hotspot locations of ten diverse large cities are characterized, and a surprisingly simple model for the distribution of WiFi hotspots in an urban environment is derived.**

I. INTRODUCTION

With the spread of smartphones, end users access Internet services on the run relying mainly on cellular networks. The amount of total mobile data traffic reached 2.5 exabytes per month in 2014 and this monthly traffic is expected to surpass 24.3 exabytes in the year 2019 [1]. To respond to these trends, WiFi offloading has come to the center of industry thinking and drives the research agenda.

With WiFi offloading, the load on existing cellular infrastructure, and consequently, the expenses for new infrastructure can be reduced. Together with the increased capacity of WiFi hotspots, customer satisfaction can be improved. End users also benefit from WiFi offloading due to reducing the risk of exceeding their data plan volume limitations. Furthermore, in areas with low mobile coverage, WiFi connections are superior to cellular connections and permit usage of bandwidth demanding applications like video streaming or online gaming.

What can a WiFi hotspot model be used for? The performance of traffic management solutions using WiFi networks highly depends on the coverage of WiFi hotspots and on the strength of the received signal. A low signal strength of the WiFi signal results in low throughput, which has an impact on energy consumption [2] and may not meet the requirements of the application [3]. A model for WiFi hotspot locations can facilitate the design and performance evaluation of mobile traffic management solutions, which incorporate WiFi offloading (e.g., [4]), as well as future Internet of Things services for smart cities relying on WiFi infrastructure [5]. A generic model allows generating WiFi hotspot distributions for cities of different size, shape, population density, and number of hotspots to evaluate hypothetical scenarios and the scalability of mechanisms. Thereby, the benefits of such solutions/services can be assessed more accurately depending on the available offloading potential.

In this work, the WiFi hotspot locations of ten large cities were obtained from a public WiFi database[1] and their characteristics are analyzed. As a result, we propose a simple model

for the distribution of WiFi hotspots in an urban environment relative to the city center. Using a transformation into polar coordinates, we show that the hotspot locations can be modeled with a uniform distribution of the angle and an exponential or gamma distribution of the distance. Our proposed model allows to investigate offloading potential and mimics realistic characteristics, e.g., in terms of distance of an arbitrary user to the closest hotspot.

Why propose a simple model? The WiFi hotspot locations could also be modeled with more complex distributions or higher order models. Although these more complicated models might better fit the characteristics of particular cities, they are harder to parametrize and they must be fitted for each city separately. However, in this case, a higher accuracy for a given city can be obtained by taking the actual hotspot locations from public databases. In contrast, our proposed model is general, has an intuitive parameter, and showed to be sufficiently accurate for the desired applications.

This work is structured as follows. Section II describes related work on WiFi hotspot models. Section III shows the applied methodology and Section IV presents the characteristics and model of hotspots distributions in cities. Finally, Section V concludes.

II. RELATED WORK

WiFi offloading/sharing started in specialized communities (e.g., Fon[2]), but public WiFi is now widely available as both free and commercial services. Many cities over the world have comprehensive WiFi coverage in the city centers just by free public WiFi hotspots provided by various cafes, shops, bars, pubs, libraries, public buildings, and government buildings. There are databases providing the locations of these open/public WiFi hotspots. Many of these databases are user based websites with hotspot locations gathered, uploaded, and updated by a huge community (e.g., OpenWiFiSpots[1]). Moreover, also telecommunication operators (e.g., BT[3]) deploy own private/closed WiFi infrastructure to offer their users access to an alternative Internet link.

The spatial distribution of WiFi hotspots is measured with a tracking method in [6]. The results show that highest density of WiFi hotspots corresponds to residential areas. The distribution of WiFi hotspots is naturally related to the population density in the city, since WiFi hotspots are deployed in close to every household, offices, shops or public places. A first model of the population density exponential decline from the city center was

---

[1] http://www.openwifispots.com/

[2] http://www.fon.com
[3] http://www.btwifi.co.uk/

(a) Extract of hotspot locations and computed center (51.5126N, 0.1405W). Map source: OpenStreetMap.

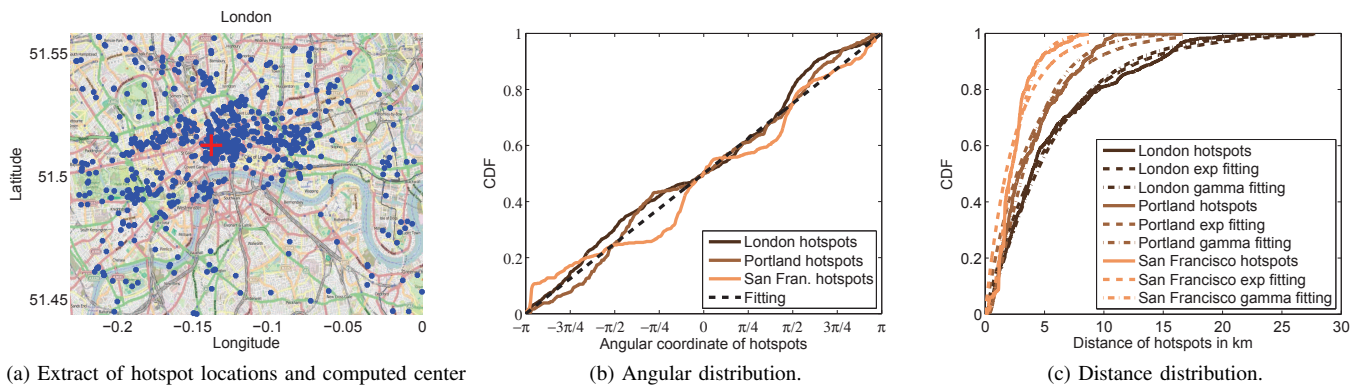(b) Angular distribution.

(c) Distance distribution.

Fig. 1. Visualization of London hotspots. Angular and distance distributions of London, Portland, and San Francisco hotspots with respect to city center.

developed in [7]. A survey on studies of urban population density [8] provides an overview of refined models considering, e.g., lower density in the center due to lower residential land use or polycentric cities [9] induced for example by suburbs. The relation between the spatial structures of wireless networks and population densities has been investiaged in [10]. The authors find that base stations belonging to different mobile operators often cluster according to population density. In [11] different point process models are used to model the density of cellular networks. A survey on the literature related to stochastic geometry models for modeling cellular networks is provided in [12]. The models lack of means to generate hotspot distribution for cities of different shape due to, e.g., natural boarders of a coastline.

III. METHODOLOGY

To characterize the geographic distribution of public WiFi hotspots in cities, geographic information about these hotspots is needed. Considering that every single point on the surface of the earth is uniquely identified by a pair of geographic coordinates (latitude φ and longitude λ), each hotspot location can be described by such an ordered pair (φ, λ).

In this work, we use the OpenWiFiSpots[1] database to obtain the addresses of public hotspots. Considering that the website provides no API to request the data, hotspots in different cities were searched manually on the website and the addresses were parsed from the search results. To transform the addresses to geographic coordinates, the MapQuest[4] geocoding API was used. Hotspot locations of ten large cities were obtained, eight in the United States and two in Europe. To give an example, Fig. 1a shows a small map extract of London with those of the hotspots that are contained in this area. Cities with a large number of listed hotspots and different layouts (e.g., grid-based cities, ring-based cities) and characteristics were selected to obtain more general results. The first columns of Table I present some of these characteristics, i.e., the number of gathered hotspots, the total investigated area, and the population of each city. It can be seen that the cities widely differ, e.g., in the number of users per hotspot. Note that the obtained hotspot locations are only a sample of a possibly larger number of WiFi hotspots, as some hotspots might not be listed in the database. As a result of this work, we find a

simple model that fits quite well for all cities independent of the actual characteristics.

To provide general statements for each city, the hotspot distribution is analyzed relative to the city center. Therefore, the city center was computed via a centroid calculation on the WiFi hotspot locations using the k-means algorithm. Then, the geographic coordinates of the WiFi hotspots were transformed into a polar coordinate system, which had the city center (φ_c, λ_c) as reference point and north as reference direction. Thus, coordinates (φ, λ) of each WiFi hotspot could be expressed in terms of polar coordinates (d, θ) with the spherical distance d from the city center and angle θ towards the reference direction[5]. Eq. 2 can be used to calculate the spherical distance between the coordinates (φ_c, λ_c) and (φ, λ) (in radians) by using the haversine formula (term a from Eq. 1) and the mean radius of the Earth r_E. Eq. 3 can be used to compute the angle between (φ, λ) and the reference direction. Both computations use the $atan2$ function[6], which is a two argument version of the arctangent function implemented by many programming languages. Note that negative angles of θ point counterclockwise from north, whereas positive angles point clockwise from north.

$$a = \sin^2(\frac{\varphi - \varphi_c}{2}) + \cos\varphi \cdot \cos\varphi_c \cdot \sin^2(\frac{\lambda - \lambda_c}{2}) \quad (1)$$

$$d = 2 \cdot r_E \cdot atan2(\sqrt{a}, \sqrt{1-a}) \quad (2)$$

$$\theta = atan2(\sin(\lambda - \lambda_c) \cdot \cos\varphi,$$
$$\cos\varphi_c \cdot \sin\varphi - \sin\varphi_c \cdot \cos\varphi \cdot \cos(\lambda - \lambda_c)) \quad (3)$$

IV. MODEL

To model the hotspot distributions in cities, we first analyze their characteristics. Based on the insights from the ten investigated cities, it is possible to generate hotspot locations with similar characteristics.

A. Analysis of Hotspot Distributions

We investigate the hotspot distribution in terms of the distance and angle of the polar coordinates with respect to the city center. As an example, Fig. 1b and 1c show the angle

[4]http://developer.mapquest.com/

[5]http://www.movable-type.co.uk/scripts/latlong.html
[6]http://www.mathworks.com/help/matlab/ref/atan2.html

TABLE I.   General information about investigated cities. Maximum (D) and mean (mae) absolute error for uniform fittings of angular distribution. Mean (μ) and coefficient of variation (c_v) of distribution of hotspot distances, and D and mae for exponential (E) and gamma (G) fittings.

City	Number of hotspots	Total investigated area (in km^2)	Population (in thousands)	D	mae	μ	c_v	D_E	mae_E	D_G	mae_G
Austin	220	220	843	0.1619	0.0677	3.2041	0.6886	0.1262	0.0545	0.0814	0.0319
Berlin	110	250	3502	0.0841	0.0401	5.0306	0.9335	0.0661	0.0270	0.0767	0.0295
Boston	193	173	637	0.0809	0.0274	2.5224	0.7201	0.1884	0.0476	0.1048	0.0195
Brooklyn (NYC)	454	419	2566	0.1142	0.0537	5.5942	0.7948	0.2057	0.0383	0.1674	0.0466
Houston	307	306	2161	0.1023	0.0412	6.7299	0.7294	0.1645	0.0335	0.0608	0.0124
Los Angeles	199	165	3858	0.0757	0.0288	2.7116	0.7239	0.1079	0.0542	0.1027	0.0418
London	668	367	8308	0.0855	0.0308	5.7011	0.8826	0.0488	0.0185	0.0609	0.0177
Portland	419	465	603	0.0713	0.0244	3.8773	0.7394	0.1186	0.0339	0.0640	0.0199
San Francisco	214	241	826	0.1036	0.0370	2.5045	0.6294	0.2732	0.0710	0.0943	0.0179
Seattle	296	202	635	0.0887	0.0353	2.9365	0.7968	0.1136	0.0464	0.1304	0.0456

and distance distribution of hotspots of three diverse cities, namely, London, Portland, and San Francisco.

In Fig. 1b, the cumulative distribution functions (CDF) of the angular coordinates of the hotspots (solid) are compared to a uniform distribution $F(x) = \frac{x+\pi}{2\pi}, x \in [-\pi, \pi)$ (black dashed). We observed for each of the ten cities that the angular distributions are not perfectly uniform with some minor deviations due to city-specific geographic conditions like water areas or parks, which caused hotspot-free spaces at the corresponding angles. Nevertheless, still a high similarity to a uniform distribution is visible.

For assessing the goodness of fit, we apply two standard methods for comparing distributions[7], namely, the maximum absolute error, i.e., the Kolmogorov-Smirnov statistic D, and the mean absolute error (mae), indicating how far the model is from reality at most (D) and on average (mae), respectively. The fifth column of Table I shows D values for fitting the angular distributions with a uniform distribution. It can be seen that all fittings have a rather high D due to particular geographic characteristics of the different cities. For example, the shape of the city of Austin contributed to a slightly elliptic hotspot distribution causing the highest D value. However, the mae values in column six show low values and indicate that the angular distribution of hotspots in a city can nevertheless be well approximated by a uniform distribution, which is sufficiently accurate for practical applications (see below).

In Fig. 1c, the cumulative distribution functions of hotspots (solid) are shown, i.e., the relative frequency of hotspots having a distance to the city center smaller than d. In this case, a high similarity to an exponential distribution $F(x, \mu) = 1 - \exp(-\frac{x}{\mu}), x \geq 0$ (dashed) with mean μ can be observed. Estimating the mean μ of the exponential distribution from the hotspot data (cf. seventh column of Table I) in a maximum likelihood sense, a good approximation is reached. As the coefficients of variation c_v in the eighth column indicate that exponential fitting might not be perfectly accurate ($c_v \approx 1$), we also compare to a more general gamma distribution $F(x, \alpha, \beta) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}, x \geq 0$, whose parameters α and β can be estimated from μ and c_v.

The ninth and tenth column of Table I indicate D and mae of the fitting of distance distributions of each city with exponential distributions, while columns eleven and twelve refer to the respective goodness of fit values for Gamma distributions. The D values illustrate that the distributions

[7]http://www.mathworks.com/matlabcentral/fileexchange/ 22020-goodness-of-fit--modified-/content/gfit2.m

are not perfectly exponential. For example, the highest D value in San Francisco is caused by the high hotspot density along the northeast waterfront, which cannot be accurately reproduced by an exponential distribution ($c_v \ll 1$). Again for all cities, the generally low mae values indicate that yet a good approximation is possible. It is also noteworthy that the exponential fitting works well for cities of different sizes, although small cities are more prone to inaccuracies caused by geographical peculiarities. Fitting with the more general, two parametric gamma distribution, in most cases a better approximation in terms of D and mae can be reached. As expected, the smaller the c_v values, the better the goodness of the gamma fitting compared to the exponential fitting. This means, especially for cities with low c_v (e.g., Austin or San Francisco), the additional parameter of the gamma fitting helps to decrease the D and mae values, and thus improves the approximation of the actual distance distribution. The closer c_v to 1 (e.g., Berlin or London), the less the gain of using a gamma distribution is visible.

All in all, after transforming the hotspot locations in polar coordinates with respect to the city center, it could be observed that the angular distribution can be decently approximated by a uniform distribution, whereas the distance distribution can be fitted by an exponential or gamma distribution. Next, hotspot distributions with these characteristics can be created.

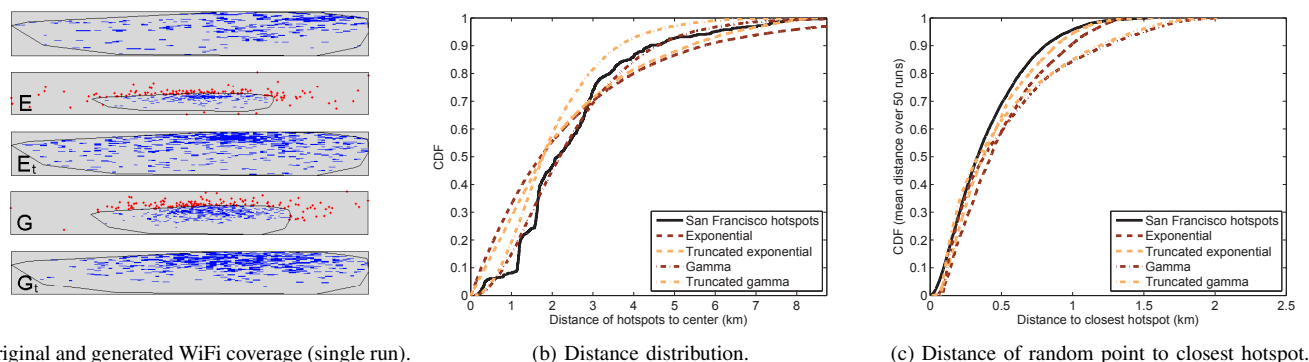### B. Generation of a Hotspot Distribution for a Generic City

First, the coordinates of the city center (φ_c, λ_c) (latitude/longitude) have to be determined. Then, random hotspot locations will be computed in polar coordinates by generating a uniformly distributed angle θ, and a distance d, which follows the desired exponential or gamma distribution. For example, for a uniformly distributed angle and an exponential distance distribution with mean μ, two random numbers (d, θ) can be easily obtained by inverse transform sampling. Eq. 4 and 5 use the trigonometrical functions to transform the polar coordinates (d, θ) back to latitude/longitude coordinates (φ, λ) (in radians)[5] taking into account the city center (φ_c, λ_c) and the spherical Earth with radius r_E:

$$\varphi = \arcsin(\sin \varphi_c \cdot \cos \frac{d}{r_E} + \cos \varphi_c \cdot \sin \frac{d}{r_E} \cdot \cos \theta) \quad (4)$$

$$\lambda = \lambda_c + atan2(\sin \theta \cdot \sin \frac{d}{r_E} \cdot \cos \varphi_c, \cos \frac{d}{r_E} - \sin \varphi_c \cdot \sin \varphi) \quad (5)$$

The limitation of this naive approach is that a circular and possibly unlimited area will be covered with hotspots.

(a) Original and generated WiFi coverage (single run).   (b) Distance distribution.   (c) Distance of random point to closest hotspot.

Fig. 2. Impact of models (exponential, gamma) and truncation to city limits (convex hull of hotspots) on generated hotspot distribution in San Francisco.

To create a hotspot distribution for a city with a given shape (or any arbitrary area), additionally an accept-reject method can be applied, accepting only hotspot locations within the city limits. However, the rejection sampling leads to a truncated distribution, which has different characteristics than the modeled distribution. In Fig. 2a, we illustrate this effect for the city of San Francisco. The top subplot shows the original WiFi coverage (in blue) within the city limits (convex hull of hotspots, black) assuming a WiFi range of 100m. In the subplots below, the same number of hotspot locations were generated using the fittings presented in Table I. Generating hotspot distributions with the naive exponential (E) or gamma (G) model places many hotspots outside the city limits (red locations). This can be avoided by rejecting hotspot locations outside the convex hull of the real hotspots resulting in truncated exponential (E_t) or truncated gamma (G_t) models. Fig. 2b depicts the CDF of the distance d of a hotspot to the center for the different generation approaches, showing that the truncation leads to smaller distances than in reality. This means, the model parameters need to be adjusted to take the truncation into account. Fig. 2c shows the CDF of the mean distance from a random point within the city to the closest hotspot over 50 generated hotspot distributions. This constitutes an exemplary application and can be used, for example, to calculate coverage, signal strength, and handovers. Here again, the impact of the truncation is visible. It can be seen that the exponential models produce more realistic results, which is due to the higher variance of the distances. This effect could be observed for all ten investigated cities, therefore, we infer that the exponential model is sufficient to create practical hotspot distributions for truncated areas.

To sum up, we showed that a simple model can be used to generate hotspot distributions for generic cities. Using a uniformly distributed angle and an exponentially distributed distance provides an easy generation of hotspot distributions (only mean of distances is needed as parameter), which have a good applicability and a high accuracy. To increase the accuracy of the generated distributions, it is not sufficient to only improve the fitting of the angular/distance distributions, but more sophisticated generation processes are needed, which also take city shape and geographical peculiarities into account.

V. CONCLUSION

This work presented the characteristics of the distribution of WiFi hotspot locations in cities. When looking at the polar coordinates of the hotspots with respect to the city center, a uniform distribution of the angle and an exponential or gamma distribution of the distance could be observed. Thus, a simple but accurate model of WiFi hotspot locations could be derived, which can be used to create spatial distributions of WiFi hotspots in arbitrary cities, e.g., for performance evaluation of mechanisms that rely on the coverage and throughput of WiFi hotspots in cities. In future work, additional characteristics of hotspot locations have to be investigated, which could necessitate more complex models.

REFERENCES

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014-2019," Cisco, Tech. Rep., 2015.

[2] V. Burger, F. Kaup, M. Seufert, M. Wichtlhuber, D. Hausheer, and P. Tran-Gia, "Energy Considerations for WiFi Offloading of Video Streaming," in *EAI MONAMI*, Santander, Spain, 2015.

[3] V. Burger, M. Seufert, F. Kaup, M. Wichtlhuber, D. Hausheer, and P. Tran-Gia, "Impact of WiFi Offloading on Video Streaming QoE in Urban Environments," in *IEEE ICC QoE-FI Workshop*, London, UK, 2015.

[4] M. Seufert, V. Burger, and T. Hoßfeld, "HORST - Home Router Sharing based on Trust," in *CNSM SETM Workshop*, Zurich, Switzerland, 2013.

[5] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, 2014.

[6] P. Rouveyrol, P. Raveneau, and M. Cunche, "Large Scale Wi-Fi tracking using a Botnet of Wireless Routers," in *Workshop on Surveillance & Technology*, Philadelphia, PA, USA, 2015.

[7] C. Clark, "Urban Population Densities," *Journal of the Royal Statistical Society. Series A (General)*, vol. 114, no. 4, 1951.

[8] J. F. McDonald, "Econometric Studies of Urban Population Density: A Survey," *Journal of Urban Economics*, vol. 26, no. 3, 1989.

[9] D. A. Griffith and D. W. Wong, "Modeling Population Density Across Major US Cities: A Polycentric Spatial Regression Approach," *Journal of Geographical Systems*, vol. 9, no. 1, 2007.

[10] M. Michalopoulou, J. Riihijarvi, and P. Mahonen, "Studying the Relationships between Spatial Structures of Wireless Networks and Population Densities," in *IEEE GLOBECOM*, Miami, FL, USA, 2010.

[11] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.

[12] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic Geometry for Modeling, Analysis, and Design of Multi-tier and Cognitive Cellular Wireless Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, 2013.