

From Your Routine to Hotspot Deployment for Data Offloading

Eduardo Mucelli Rezende Oliveira*†

Aline Carneiro Viana†

edumucelli@lix.polytechnique.fr
*Ecole Polytechnique

aline.viana@inria.fr
†INRIA

ABSTRACT

This paper provides the framework to transform mobility and data traffic information of people in an urban scenario into a set of well-positioned places likely to receive wireless hotspots to support data offloading. This is a work in its early stage.

Categories and Subject Descriptors

C.2.1 [Computer-Communication Networks]: Network Architecture and Design—*network topology*

Keywords

Data offloading, complex networks, mobility

1. INTRODUCTION

With the steady growth of sales of smart-phones, the demand for services that generate mobile data traffic has grown tremendously. The growing use of traffic data generated from mobile devices overloads the network infrastructure, which is not always prepared to receive such demand [1].

To tackle this problem, there are typically two approaches proposed in the literature: one may scale the current network structure or shift the traffic from the mobile data network to different networks. The idea of the first approach is to upgrade the network structure in order to scale its resources and thus, better meet the current demands [6]. The latter approach tends to focus on methods to avoid the use of mobile data network as much as possible, for example, avoiding peak hours, or data offloading (i.e., transferring the demands to an inexpensive network [4]). WiFi offloading seems to be a viable solution: carefully deploying more WiFi hotspots can both be cheaper than upgrade the network structure and can concede substantial improvement in the network capacity [4].

Our objective is to define what are the best spots to receive WiFi hotspots in order to maximize the data offloading on a urban scenario. To accomplish our objective, we aim to study the mobile behavior and resource consumptions of people on a metropolitan area in a major city and turn it into a set of well located WiFi hotspots. For this, we propose a data offloading-aware hotspot deployment. It is methodologically divided as follows. First, we create a time

dependent weighted graph to represent people's mobility, and its relation with places/locations able to receive a hotspot, then we measure the locations importance and select the best-ranked ones. Better positioned hotspots are likely to provide better coverage, and therefore, are able to offload more data. It is a desirable behavior both for operators and users: the first can see the traffic being shifted to inexpensive networks while the latter can take advantage of higher data rates and less monetary costs than using cellular networks.

To the best of our knowledge, [6] proposes the only approach to deploy hotspots that takes in consideration user mobility characteristics. It considers a campus scenario with one popular area, and the objective is to provide continuous coverage for a mobile user. Our proposal differs from this one in the following points. First, our objective is not to provide continuous coverage: On an urban scenario, this is prohibitively expensive since it would require the deployment of hotspots over the whole area, where most of them may be underutilized. Second, our approach does not select only one popular spot, but ranks them by their importance.

2. PROPOSAL

Graph creation: let S be a set of n spots of interest $S = \{s_1, s_2 \dots s_n\}$ (e.g., buildings, bus stops, trains stations), i.e., a place physically able to receive a hotspot and that likely has people in its vicinity for a long period of the day. We consider that a place is uniquely identified by its geographical position: $\forall s_{(x,y)} \in S$, where (x, y) is a pair of coordinates of a spot s . Since we are willing to use real data traces containing GPS-based positioning, x and y are, respectively, latitude and longitude. Fig. 1(a) illustrates a set of spots of interest presented as stars on map. Initially, we assign all spots of interest with a hypothetical range illustrated as dashed circles.

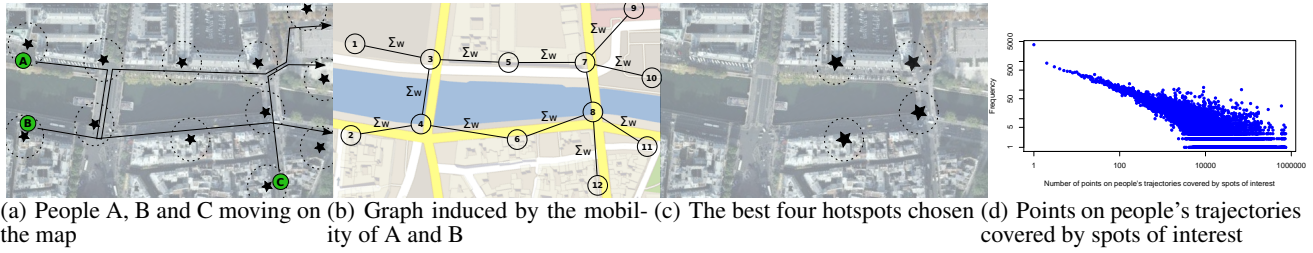
Our graph $G(V, E)$ represents the interaction between people and the urban scenario. The graph should provide a simple and easy-to-see representation of people's mobility. While walking, people create trajectories, represented as a set of geographically positioned points, and their interactions with the urban scenario may point out important places to start planning the deployment of a network infrastructure. Fig. 1(a) illustrates the three trajectories created by three hypothetical people (A, B, and C) and their respective interactions with spots of interest. Let P_i be a set of all geographic coordinates on the trajectory of a person i , i.e., $P_i = \{p_1, p_2 \dots p_o\}$. A vertex $v \in V$ is created on the same coordinates of the spot of interest if the latter covers a person passing by. Note that spots of interest that does not respect this condition will not be considered as vertex in the graph. A vertex coverage of a person i by spot s with a range r is expressed as $\forall p \in P_i, \exists p \mid (p_x - s_x)^2 + (p_y - s_y)^2 \leq r^2$, i.e., i is inside s 's circular range.

An edge $e \in E$ will be created between two vertices if their respective spots of interest sequentially cover a person during its tra-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CoNEXT Student'12, December 10, 2012, Nice, France.

Copyright 2012 ACM 978-1-4503-1779-5/12/12 ...\$15.00.



jectory. Fig. 1(b) shows the graph induced when three people were covered by spots of interest. The weight of an edge is the total amount of data people transferred while walking on the path corresponding to that edge enhances the importance of spots of interest laying on trajectories in which people transfer more data.

Hotspot selection: a centrality metric herein proposed will be calculated on the graph, being guided by two intuitions: "Desire Lines" [5] and confinement of people's mobility. The vertices with high centrality values are likely to be places for hotspot deployment. Indeed, centrality metrics are widely used on the literature, but not yet explored in the problem we are considering.

The concept of "Desire Lines" states that people tend to choose the shortest-paths to arrive on their destinations. To couple with that, vertices are ranked by the betweenness centrality metric, since vertices with high betweenness are those that lay on most of people's shortest routes and likely can become well positioned hotspots.

Moreover, people's mobility is generally periodic and confined [2]. Even if people are not using the shortest routes, they are at least not going far from their home location. Closeness centrality is calculated based on the geodesic distance between all pairs of vertices on the network. It assigns higher values for vertices closer from the rest of the network, that is, on a city those are probably hospitals, and markets, i.e., places planned to be close to most of the people.

Finally, when planning hotspot deployment for data offloading, those points are also important because they are part of people's daily routes and give opportunity to data offloading-aware applications, protocols, and architectures. Therefore, the amount of data transferred by a person should be also taken in consideration in order to enhance the outflow potential to nearby hotspots.

To couple with all the above intuitions, we introduce Routine centrality (ρ), a centrality metric that aims to rank spots based on their potential to offload data: $\rho(v) = Cc(v) + Cb(v) + D(v)$ where $v \in V$, Cc is the normalized closeness centrality of v , Cb is the normalized Shortest-path Betweenness of v , and D is the total normalized amount of data w transferred by people on edges (paths) connecting the vertex v and its neighbors t , i.e., $D(v) = \frac{d(v)}{\sum_{v \in V} d(v)}$ where $d(v) = \sum_{t \in T} w(v, t)$. All normalized values are on the interval $[0, 1]$. Fig. 1(c) show the best four hotspots (stars) chosen from ρ .

3. CONCLUSIONS AND NEXT STEPS

This paper presents the initial guide-lines of a proposal for the deployment of WiFi hotspots for data offloading. The main challenges are:

- **Interesting spots identification:** after selecting an urban scenario to be used as case study, we have to build our graph based on the set of interesting spots in the city. Database of places (e.g., Google Places¹) are growing and are the most accurate source of public information about points of interest. To avoid a data-centric evaluation, we have collected data from multiple sources namely Nokia Maps, Google Places, and Foursquare counting

more than 200 thousand real points of interest with their respective latitude and longitude.

- **Trace aggregation:** to capture the timing information in a dynamic graph it would be necessary to store a graph for every time instant in the past. However, this approach leads to scalability issues, and consequently the data summation becomes required. One promising way is to aggregate all the data of a certain period and generate a graph for each one of them. It is still an open problem how to identify the perfect size for this period [3], but solutions can be proposed according to the considered scenario. Due to the routine people have, initially, data will be aggregated using daily intervals, i.e., all the data of one day will be converted into a graph using the method presented in Section 2.
- **Time-space traffic variation:** traffic volume varies during time, and our proposal couples with that setting weights on the edges to measure it. Besides, a preliminary evaluation shows that the coverage of points on users trajectory by spots of interest is power-law distributed (Fig. 1(d)), i.e., the majority of spots are able to cover a small set of points where people have been, while a small amount of spots are able to cover a high amount of points. This enhances the fact that the data traffic will be different on each hotspot, and modeling the traffic variation is a hard challenging. Moreover, there are not publicly available data traces that provide a regularly/frequently sampled data on that. We are finishing a smart-phone application to regularly/frequently collect such data as part of an experimental study we plan to conduct in a major European city.

4. REFERENCES

- [1] AT&T. AT&T Launches Pilot Wi-Fi Project in Times Square, May 2010.
- [2] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, Jan. 2006.
- [3] T. Hossmann, T. Spyropoulos, and F. Legendre. Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, March 2010.
- [4] K. Lee, I. Rhee, J. Lee, Y. Yi, and S. Chong. Mobile data offloading: how much can wifi deliver? *SIGCOMM Comp. Comm. Rev.*, 41(4), Aug. 2010.
- [5] J. A. Throgmorton and B. Eckstein. Desire lines: The Chicago area transportation study and the paradox of self in post-war America, Nov. 2000.
- [6] T. Wang, G. Xing, M. Li, and W. Jia. Efficient wifi deployment algorithms based on realistic mobility characteristics. In *IEEE MASS*, pages 422–431, Nov. 2010.

¹<https://developers.google.com/maps/documentation/places>