Routine-based Network Deployment

Eduardo Mucelli Rezende Oliveira
Ecole Polytechnique and Inria

Aline Carneiro Viana
Inria

***Abstract*—The careful deployment of hotspots in metropolitan areas allow to maximize WiFi offloading, a viable solution to the recent boost up of mobile data consumption. Our proposed strategy considers routine characteristics present on people's daily trajectories, the space-time interaction between them urban locations, and their transportation modes. Using a real-life metropolitan trace, we show our routine-based strategy guarantees higher offload ratio than the current approach in the literature while using a realistic traffic model.**

I. Introduction

WiFi offloading consists in shifting the traffic off of cellular networks to WiFi networks. Carefully deploying WiFi hotspots can both be cheaper than upgrade the current cellular network structure and can concede significant improvement in the network capacity [1]. Nevertheless, one question remains: *how WiFi hotspots should be deployed?*

This work tackles the WiFi hotspot deployment problem in a metropolitan area by leveraging mobile users' trajectories, scenario interaction and traffic demands. Our objective is to define what are the best places to receive WiFi hotspots in order to maximize the offloaded traffic on an urban scenario.

II. Rationale

People are routinary, they have regular circles of actions. Furthermore, people present three mobility characteristics, tendency to use the *shortest-path* to reach their destination, *confinedly* roaming close by their main physical address and *repetitively* visiting the same places. These observations guide our hotspot deployment proposal. For this, we consider GeoLife dataset [2] that describes the mobility and transportation modes of 182 people mostly in Beijing during a 4-year span. To better understand specific behaviors inherent from different periods of the day, each day is divided into four periods of 6 hours starting from 00:00.

A trajectory is a set of geolocalized points describing a route traveled by a person using at least one transportation mode. While passing by, people may "interact" with points of interest. In our scenario, points of interest are geolocalized physical venues, e.g., bar, bus station, supermarket, etc. We have collected information from more than 202 thousand real unique points of interest in Beijing from Google Places, Nokia Maps, and Foursquare. We consider that a person interacts with a point of interest if the former is inside the coverage range (50 meters [3]) of the latter.

III. Proposal

This section presents how mobility is mapped into a time dependent graph, routine characteristics measurement, traffic model and hotspot placement strategy.

This work was supported by the EU FP7 ERANET program under grant CHIST-ERA-2012 MACACO.

978-1-4799-3360-0/14/$31.00 ©2014 IEEE

Graph creation: Our graph $G(V,E)$ represents the interaction between people and points of interest. A vertex $v \in V$ is created on the same coordinates of the point of interest if the latter covers at least one person passing by. An edge $e \in E$ is created between two vertices if their corresponding points of interest sequentially cover a person during its trajectory.

Routine characteristics measurement: The characteristics present on people's routinary mobility are measured on GeoLife dataset using three confirmed intuitions: usage of shortest-paths, confinement and repetitiveness.

On average, 70% of people's trajectories measure, at most, half longer than the respective shortest path. We have used Google Directions API[1] to compute the shortest path. To measure the usage of shortest-paths on the graph, we use *Stress Centrality*. Vertices with high stress are those that lay on most of people's shortest routes and may become well positioned hotspots.

People's mobility is generally confined. In our scenario, 90% of the trajectories have, at most, 10 km maximum displacement regardless of the period of the day. We use *Closeness Centrality* as a way to measure how close a vertex is to the rest of the network. On a city, those are places planned to be close to most of the people's trajectories, e.g., hospitals and markets.

People tend to repetitively visit the same areas. Beijing map was divided into cells of 50 square meters in a grid shape. Then, we have calculated the number of unique cells (NUC) and total cells (NTC) that each trajectory visited. To quantitatively express the repetitiveness of a trajectory, a metric called *Repetitiveness* was conceived: $\frac{NTC-NUC}{NTC} * 100$. The result shows that 91% of the performed trajectories present at most 50% of the visited cells repeated.

Traffic model: To the best of our knowledge, there is not in the literature a freely available dataset describing both people's mobility and their traffic demands on an urban scenario. Therefore, we have created a realistic synthetic traffic model that takes into consideration traffic parameters and measurements from the literature (cf. Table I) to simulate the content generation and offload as if it was done by people participating on GeoLife while performing their trajectories.

Moreover, the traffic model was designed to couple with specific demands inherent in a scenario that contains people moving using different transportation modes. When a person is covered by a point of interest, and the transportation mode being used is "taxi", "bus", "walk", "train", "subway", "car" or "boat" the traffic generated by our model is normally made available to be offloaded. Correspondingly, when the person gets covered and uses other transportations modes, e.g., "run", "motorcycle" and "bike", no content will be available for offloading.

Spots selection: Consider a graph $G_t(V_t, E_t)$ constructed as described in Section III for a period of time $t \in T = \{1, 2, \ldots, u\}$. Let ϕ be the number of period of times per day $d \in D$, which

[1]https://developers.google.com/maps/documentation/directions

TABLE I
TRAFFIC MODEL PARAMETERS

Parameter	Value
File size	Gamma (shape=2, scale=8.5 MB)
Inter-arrival time (IAT)	Exponential
Amount per day	94.54 MB
Transfer rate	3.9 Mbps
Connection time	1.83 second

represent the set of days in which we analyze mobility data. Therefore, $|T| = |D| * \phi$. G_t represents all the interactions between people and urban scenario for the hours contained in t. Therefore, $V_t \subseteq V$ is the the set of vertices that covered at least one person during t and $E_t \subseteq E$ is the subset of edges that represents the mobility of people during t. Our objective is to maximize data offloading. We advocate that in order to deploy a WiFi network for data offloading on an urban scenario, it is important to take into consideration both traffic and mobility characteristics. For instance, consider two points of interest h_i and h_j that receive the same amount θ of traffic to be offloaded during a period of time t. If we consider only offloaded traffic, both would be equally selected. Our intuition is that if we also consider people's mobility it is possible to provide a better network deployment for data offloading envisioning future traffic.

Consider offloaded traffic (tr) and mobility characteristics metrics Stress (st), Closeness (cl) and Repetitiveness (re). Consider $U_{i,t,m}$ as the value of the metric $m \in M = \{st, cl, re, tr\}$ on the period t for the vertex i. The traffic on a period t for the vertex i is calculated as the sum of all b bytes of content offloaded from all people covered by i during t, i.e. $U_{i,t,tr} = \sum_b b_{i,t}$. Due to the different magnitudes of the metrics, we have normalized $U_{i,t,m}$ between 0 and 1 by the maximum value of m on the period t.

Let $\overrightarrow{R_{i,t}}$ be a vector of metrics calculated for the vertex $i \in V_t$. Due to the mobility, a vertex i may appear on the graph on a certain period $t = k$, but not appear on a period $t = k+1$. Therefore, $[R_{i,t,m} = U_{i,t,m} \cdot z_{i,t}]$, where $z_{i,t} = 1$ if i appears on t and $z_{i,t} = 0$ if i does not appear in t. The result of metrics attributed to the same vertex on different periods is aggregated through a sum, i.e., $\sum_{i,t} R_{i,t,m} = Total_m \forall m \in M$. Then, our problem is to find the set of vertices that maximizes the metrics with a limited number of hotspots λ, formally:

$$\text{Maximize} \sum_{i,t} R_{i,t,m} \quad \text{Subject to} \sum z_{i,t} < \lambda$$

Besides, our model can be tunned in order to prioritize either mobility characteristics or traffic: $\alpha * (\sum_{i,t} R_{i,t,st} + \sum_{i,t} R_{i,t,cl} + \sum_{i,t} R_{i,t,re}) + \beta * \sum_{i,t} R_{i,t,tr}$ by choosing weights α and β depending on operators' demands. Considering $\alpha \to 0$ and $\beta \to 1$ results in a deployment that prioritizes traffic demands. On the other hand, when $\alpha \to 1$ and $\beta \to 0$ hotspots will likely be placed on more important areas of the city regardless their traffic offload. Initially we consider both traffic and mobility characteristics with the same importance.

IV. PERFOMANCE EVALUATION

We have compared our approach, routine-based, with the current work on the literature that proposes WiFi deployment considering user mobility characteristics, grid-based[4]. Grid-based uses only the mobility of taxis in San Francisco, which is divided in a grid with one hotspot in the center of each cell. Its traffic model generates a data request every 5 seconds.

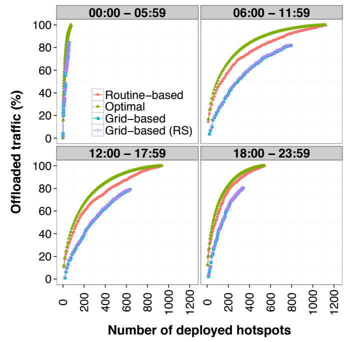

Fig. 1. Offloaded traffic per # of deployed hotspots by period of the day.

Figure 1 shows the offloaded traffic by period of the day. Each period of the day has its own number of trajectories which in turn generates a respective amount of traffic. If it is necessary to guarantee an offload ratio per period, different number of hotspots are needed to offload the traffic on each period of the day. Indeed, in order to achieve 80% of offloaded traffic on each period of the day, the routine-based needs, on average, 34% less hotspots than the grid-based approach.

V. CONCLUSIONS AND NEXT STEPS

In this work in progress, we have presented to the best of our knowledge, the first analyses of a metropolitan-wide hotspot deployment which employs a realistic traffic model. We have proposed (1) a graph model to represent the relationship between people and the city infra structure, (2) a realistic synthetic traffic model that considers transportation modes, and (3) a metric to turn points of interest into well positioned hotspots capable of providing data offload based on people's real routines and their traffic demands. Our initial results show that our routine-based deployment can offload more data with less hotspots than the current work in the literature while using a realistic traffic model. As future work, we intend to investigate how good is our deployment configuration for a different period of the dataset. Additionally, we are going to further investigate the monetary trade-off of a deployment in which the cost of the hotspots is not unique and the cost of offloaded traffic is considered. Finally, we envision to study the impact of changes on the mobility and traffic parameters α and β on the deployed configuration.

REFERENCES

[1] K. Lee, I. Rhee, J. Lee, Y. Yi, and S. Chong, "Mobile data offloading: how much can wifi deliver?" *SIGCOMM Comp. Comm. Rev.*, vol. 41, no. 4, Aug. 2010.
[2] Y. Zheng, X. Xie, and W.-Y. Ma, "Geolife: A collaborative social networking service among user, location and trajectory," *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–39, 2010.
[3] S. Dimatteo, P. Hui, B. Han, and V. Li, "Cellular traffic offloading through wifi networks," in *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, 2011, pp. 192–201.
[4] E. Bulut and B. K. Szymanski, "Wifi access point deployment for efficient mobile data offloading," in *Proceedings of the First ACM International Workshop on Practical Issues and Applications in Next Generation Wireless Networks*, ser. PINGEN '12, New York, NY, USA, 2012, pp. 45–50.