A Practical Characterization of 802.11 Access Points in Paris

Guillaume Valadon Florian Le Goff Christophe Berger
UPMC Univ Paris 06, UMR 7606 LIP6, Paris
Email: First-name.Last(-)name@lip6.fr

Abstract

Unlike other wireless technologies, the deployment of 802.11 networks is not limited to operators: access points can easily be installed by end-users for domestic use. This singular type of deployment is the reason why 802.11 networks are omnipresent in our urban landscapes. Indeed, in metropolitan areas, laptops frequently detect tens of 802.11 access points from the same location. In this work, we describe both simple and more complex data about access points obtained in two Paris districts during an extensive experiment from August to October 2007. We introduce a lightweight scanning platform that runs on common smartphones. Using the obtained data, we examine various parameters: (1) SSID, (2) manufacturers, (3) security modes, (4) density, (5) data rates, and (6) channels utilization. For example, we show that in the two districts that we mapped as few as 7% of the Wi-Fi networks are not secured. Similarly, we provide a practical evidence that 90% of detected access points where installed along with DSL Internet access.

1 Introduction

The 802.11b/g protocol, commonly referred to as Wi-Fi , changed the way people interact with networks. More and more consumer devices contain 802.11 chipsets. Laptops, phones, or even music players, printers and game consoles can now be connected to wireless networks. According to the Wi-Fi Alliance [4], more than 120 millions Wi-Fi chipsets were shipped in 2005. In addition, wireless networks are available in diverse locations such as cafés, parks, hotels or airplanes. This growth in usage raises several questions and issues concerning security, or automatic assignment of channels. A rigorous functional mapping of the Wi-Fi landscape will for instance provide data about access points using Wired Equivalent Privacy (WEP), and therefore help to protect users [21].

Collecting information about 802.11 access points is a challenging and time consuming task. The most common technique is called *wardriving* and implies to drive around a city in a car with a laptop, a GPS receiver and a PCMCIA wireless card in order to detect access points and record their geographical coordinates. However, there is no generic methodology to perform access points surveys. This raises questions concerning the observed characteristics and how they should be interpreted. The area where the data is obtained is for example of great importance to determine the density of access points. Likewise, the moving speed during surveys impacts statistics and their significance. It is therefore crucial to collect information about access points with an explicit methodology.

Characterization works already exist but only few of them analyze Wi-Fi access points in Europe [10, 11, 9]. As data is usually gathered by random wardriving, it is rather difficult to use these surveys to make firm assumptions about access points parameters. To our knowledge, our study is the first one to log rich information about access points by a systematic mapping of a European urban area. Moreover, our work is singular as it provides an analysis of data based on the French residential Internet access business, impacting the 802.11b/g landscape for the following reasons: (1) most DSL subscriptions come with a dedicated modem/router including a Wi-Fi access point; (2) some Internet Service Provider (ISP) deploy their own Wi-Fi access points on their customer modems.

In this work, we analyze various parameters of 802.11b/g access points that we obtained in two Paris districts from August to October 2007. The data was produced while walking in the streets with a mobile phone and a GPS receiver, as would a regular user. This approach does not discover every access points but has the advantage of reflecting the wireless environment a user encounters while wandering in the streets. Data analysis provides realistic estimates of Wi-Fi parameters such as access points security modes, density or channels utilization. Unlike common wardriving based studies that only log the SSID[1], the BSSID[2] and geographical coordinates of an access point [16] , we collect as much parameters as possible, ranging from channels to

[1] Service Set IDentifier; name of the network.

[2] Basic Service Set IDentifier; unique ID of the access point.

security modes. This work offers three important contributions: (1) a dedicated tool to log access points parameters on Nokia smartphones with external GPS receivers; (2) a data set containing parameters of 30904 access points discovered in two Paris districts, respectively of 7.15 km^2 and 2.54 km^2; (3) a pratical characterization of access points in the districts.

2 Wardriving

Wardriving became popular with softwares such as *Kismet* [1] or *Netstumbler* [2]. At first, their goal was to detect open networks and associate geographical coordinates to access points. Later, they were enhanced to perform attacks against protected networks by cracking WEP keys. The research community previously used concepts similar to wardriving. For instance, the PlaceLab project [17] showed that it is possible to perform efficient Wi-Fi based geolocation if geographic coordinates of access points are available. Likewise, characterization works and access points databases [22, 20] exists but their contents has several limitations: (1) there is no detailed information about the way that the data was gathered; (2) only simple information such as BSSID or SSID are usually captured; (3) they mainly cover USA.

A study in Darthmouth [12] showed that wardriving and warwalking present some issues that must be taken into account when collecting and analyzing the data. First, the speed influences the number of access point discovered: the slower an observer goes, the highest number of access points are discovered. Then, it shows that access points inside large and tall buildings are not discovered. Whereas warwalking does not discover 100% of the access points, it provides practical estimates of the Wi-Fi landscape encountered while walking in the streets.

3 Methodology and data set

In this section, we describe 802.11 frames involved in the discovery of access points, as well as the various parameters that they contain. Then, we jointly discuss the methodology and the tool used to collect data in Paris and present the resulting data set.

3.1 Scanning 802.11b/g access points

The 802.11 specification defines two methods that a station[3] can use to look for surrounding 802.11b/g access points. The first one is based on *Beacon* frames which are periodically sent by access points to advertise their presence to neighboring stations. In the second method, the station broadcasts a *Probe Request* frame and waits for *Probe Response* frames unicasted by access points that received the request. While both methods retrieve the same information, the Beacon based one does not ensure that access points are reachable by the station because they emit frames with a lower power than access points. The Probe Request method therefore discovers fewer access points. Moreover, when geographical coordinates are recorded this method delivers results with a better accuracy as access points will be given a location closer to the real one. When a station looks for surrounding access points, it performs a scan: the station successively cycles from one channel to another to report access points available on all of the fourteen 802.11b/g channels.

Beacon and Probe Response frames do not only notify stations of the presence of an access point but also carry the following information:

1. **SSID** Name of the Wi-Fi network
2. **BSSID** MAC address of the access point
3. **Mode** Ad-hoc or Infrastructure
4. **Protection** Security scheme (open, WEP, WPA, WPA-PSK, or 802.1x)
5. **Supported Rates** Transmission rates supported by the access point
6. **Channel** Channel on which the access point operates
7. **Optional Information Elements** Additional data about the access point

Optional Information Elements are variable length fields that mainly contain vendor or country specific information. In some cases, they can be used to precisely identify the manufacturer and model of an access point and get details about the supported modulations.

The 802.11 standards define that an access point is a base station that possesses only one BSSID. Note that one SSID can correspond to several BSSID, but that one BSSID can't correspond to different SSID. Some recent access points can however manage several virtual access points but they have the same properties as a regular access point. In the following discussions, we consider that one BSSID equals one access point.

3.2 The data set

Our objective is to characterize the Wi-Fi landscape from a regular user point of view. For that reason we selected smartphones to perform our warwalking campaigns. For this experiment, we used Nokia N95 and N80 phones with external GPS receivers, and a dedicated wardriving software

[3] for example a laptop or a smartphone.

developed in Python [19]. This configuration has an important battery lifetime: we did continuous scanning up to five hours long. On these smartphones, using a specific Nokia API [15], it is possible to perform Probe Request based scans and only retrieve essential frame level information presented in the previous section. While logging the full frame would obviously provide more detailed results, our approach is a compromise between disk usage and valuable data contained in the frame. The interval between each scan was set to 3 seconds: we experimentally evaluated that it is an accurate trade-off to avoid duplicates in the log files and missing few access points.

The data was gathered during a rigorous warwalking campaign conducted in Paris 5th and 13th districts from August to October 2007. Altogether, these two scans took around 44 continuous hours to complete. The districts were divided into areas in which the authors scanned the streets by walking along them. In order not to alter measures, we tried to avoid going to the same streets more than once. The districts are approximately 10 km^2. However we estimate than only 4.2 km^2 in Paris 13th and 1.9 km^2 in Paris 5th were scanned since we do not have access to areas such as hospitals or private properties. In Paris 5th and Paris 13th, 9307 and 21597 access points were respectively discovered.

For each access point, both Paris 5th and 13th data sets contains geographical coordinates, positioning accuracy, reception level, security and connection modes, capability field[4], SSID and BSSID. The Paris 5th data set also includes operating channel, supported rates and optional information elements. The anonymized data sets and the source code of the software are provided for free use at `http://content.lip6.fr/warwalking/`.

4 Analysis

In this section, we first explain why the French ISP market significantly influences the number of access points that we discovered. Then, we describe parameters starting from direct observations, such as SSID, to more detailed ones. Except for channels and data rates that are only available in the Paris 5th set, results for both districts are jointly discussed.

4.1 Economical background

The knowledge of the French ISP market gives insightful information on the number of Wi-Fi access points used in France. In 2003, an ISP named Free started to provide a new triple play box[5]: the *Freebox*. Its immediate commercial success prompted Free's competitors to roll out their own boxes At first all these boxes were regular NAT routers but in 2005, they were upgraded with an embedded Wi-Fi access point. Since a box is provided along with a DSL subscription, each French DSL customer potentially owns a Wi-Fi access point.

	Nationwide	Paris 5th	Paris 13th
Households	25689000	32749	87196
Boxes	7800000	9944	26475

Table 1. Estimated number of Internet Boxes

As of March 2007, 13.7 millions French households were connected to the Internet using a DSL line [13]. 57% of these subscribers[6] are using VoIP services provided with their Internet access. As these services are only available to customers using boxes, we use this figure as an estimate of boxes in France. Using the official number of households and areas sizes, we evaluate the lower bound of potential Wi-Fi access points in Paris 5th and Paris 13th as shown in Table 1. As a matter of fact, these numbers are underestimates as a household in Paris has a higher probability of owning a box than everywhere else in France but we were unable to determine in which proportions.

4.2 Identifying the boxes

The *Freebox* that *Free* provides is difficult to identify as it acts as a virtual access point: it adverts four SSID using four different but consecutive BSSID. Nevertheless, we found out a method to count all BSSID belonging to *Freeboxes* that performs the following steps: (1) it looks for all BSSID with SSID *freephonie* and security WPA (5352, 17.3%); (2) using these results, it generates a list of the three BSSID that precede results of the first step; (3) finally it compares this list against the whole data set and returns BSSID that correspond to *Freeboxes* in our data set (10013, 32.4%). Other boxes are easier to identify. In our data set, we easily distinguished that 17976 access points (57.7%) using well-known SSID that correspond to ISP[7]. Consequently, this classification suggests that ISP provide 90.1% of access points present in our data set.

4.3 SSID & Connection modes

In general, the analysis of unique network names does not give much information by itself. However, it is an indicator for detecting access points using their default factory settings such as *hpsetup*: the factory SSID of HP printers and laptops. In our data set, we manually identify that

[4] bit field indicating capabilities of the access point.

[5] NAT router and DSL/cable modem providing Internet access, VoIP and digital TV.

[6] 7.8 millions French households.

[7] such as *Wanadoo**, *Livebox-**, *ALICE-*:*, *TECOM-**, and *THOMSON*.

Rank	%	Name
1	38.2	Unknown
2	11.7	Hon Hai Precision Ind.
3	11.5	USI
4	7.6	TECOM Co., Ltd.
5	7	neuf cegetel
6	3.7	Freebox SA
7	3.3	Cisco Systems
8	2.4	Netgear, Inc.
9	2	D-Link Systems, Inc.
10	1.9	ASKEY COMPUTER CORP.

Table 2. TOP 10 manufacturers

2.9% of access points are in this situation. Among the most frequent SSID, three of them correspond to access points shipped by ISP: *freephonie*, *N9UF_TEL9COM* and *THOMSON*. The following three SSID are factory defaults of popular manufacturers: *NETGEAR*, *linksys*, and *dlink*. The SSID is therefore a criterion that can easily help identifying the manufacturer of an access point.

Access points configured in Infrastructure mode are predominant. This is an expected result as this mode delivers better performance and is the typical deployment of Wi-Fi networks. In both districts, there is around 1% of access points, exactly 310, configured in Ad-Hoc mode. Among them, 76 have *hpsetup* as SSID, and 2 belong to an ongoing OLSR[8] experiment in the districts, and to personal wireless music centers from Philips. All other SSID are unique and cannot be classified.

4.4 Manufacturers

The classification of access points by manufacturers shown in Table 2 was produced by comparing BSSID to the Organizational Unit Identifier (OUI) database [7]. As expected, manufacturers of boxes *Hon Hai Precision*, *USI*, *TECOM*, *Freebox SA* and *neuf cegetel* are ranked before well known network vendors. It is interesting to note that 38.2% of access points are labeled as *Unknown*; their corresponding BSSID are not in the OUI database. However, using the method described at the end of Section 4.1, we found out that 33% of these access points are in fact associated with boxes supplied by *Free*.

4.5 Security

Five years ago, a non-exhaustive wardriving survey conducted by one of the authors in five Paris districts revealed that most access points were open (i.e. encryption-free). Today, only 11.3% are, see Table 3. However, the associated SSID reveal that fourteen ISP use open access points as part of their business, and force their users to authenticate on a captive portal. Not considering these ISP, only 6.6% access points are open. Based on their SSID, we estimate that almost all of them are still using their default factory setup. WEP is still used by 40.4% of access points whereas better modes exist, even though WEP has proved to be unsafe around 2001. The percentage of access points using WPA-PSK is higher than expected, because some French ISP are shipping access points configured by default with WPA-PSK. On the other hand, WPA (also known as WPA-Enterprise) represents 30.2% of access points. This is surprising as it is generally used to secure private company networks. In fact most associated SSID correspond to *Free* that uses its customers' access points to extend its own Voice over Wi-Fi network. Once removed, WPA only accounts for 0.3% of access points.

Mode	%	Mode	%
Open	11.3	WPA	18.1
WEP	41.4	WPA-PSK	30.2
Open; Not ISP	6.6	WPA; Not ISP	0.3

Table 3. Security modes

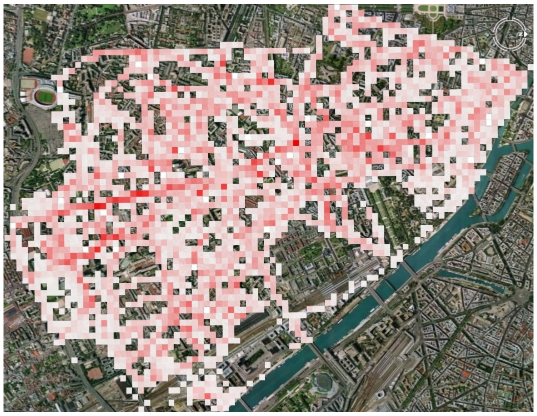

Figure 1. Density of access points in Paris 5th and 13th (North is on the right side)

4.6 Density

In January 2006, there were around 1900 access points per km^2 in Manhattan [8]. As of early this year (2007), there are close to 3000 access points per km^2 in Tokyo urban areas[9]. In the two districts that we consider, there is between 3107 and 5090 access points per km^2. In the first result, the

[8] a routing protocol for mesh networks [6].

[9] private conversation with a PlaceEngine [18] Team member.

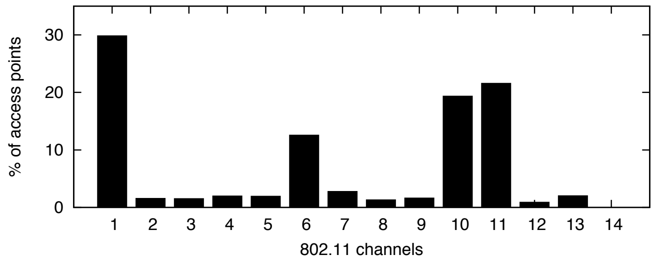

Figure 2. Channels used in Paris 5th

total number of observed access points was divided by the exact area of the districts. In the second one, it was divided by the area of the scanned zone. However as described earlier, some access points are counted more than once since one ISP uses its customers access points to broadcast their own SSID (*freephonie*). After the removal of these SSID, densities are 2626 and 4301 access points per km^2. When multiple geographical coordinates were associated to a single access point in the data set, we applied a centroid based algorithm [5] to aggregate them and obtain a unique coordinate. We now study the correlation between access points and population density; as shown in Figure 1, we observe that the $50m * 50m$ squares with the highest density (in red) are located around tall buildings or avenues where there is a higher concentration of apartments.

4.7 Channels

Prior to this analysis, we were expecting that access points would be using French default channels (10 or 11) as a similar pattern was previously shown in the USA [3]: 42% of the discovered access points were using the default channel (6). Therefore, usages of channels shown in Table 2 surprised us as the three non-overlapping channels (1, 6, and 11) represent 64% of access points in Paris 5th. Whereas we did not come up with a firm explanation of this behavior, we believe that it is not a coincidence as some boxes shipped by ISP are able to select the least perturbed channel at boot time.

4.8 802.11b/g

We classified access points using the *supported rates* information based on the fact that a 802.11b access point only supports 1 Mbps, 2 Mbps, 5.5 Mbps, and 11 Mbps, while a 802.11g access point supports higher rates. First of all, we found only one access point providing 802.11g, and 57% of access points only supporting 802.11b. This last result is also a consequence of the ISP technical choices, as 64% of these 802.11b/g access points are boxes.

5 Conclusion

In this paper, we presented a new characterization effort of 802.11 access points. We meticulously scanned two Paris districts using a specific Wi-Fi scanner running on Nokia smartphones. The resulting data set contains essential frame-level parameters about each access point such as the channel, the security or information elements. We released both the scanner and the data set to the community so as to conduct similar campaigns and compare the outcomes. The analysis of the data set confirmed that the French market of access points is dominated by boxes shipped by ISP. Results are consequently different from similar surveys: density values are for instance 150% higher in our data set than in Tokyo.

Our future work will investigate different techniques to identify and classify boxes using multiple parameters conjointly. We will focus on the content of the capability field, and information elements to determine if they can be used to discover the model of an access point. Moreover, we would like to study the evolution of the Wi-Fi landscape overtime by performing other rounds of measurements in the same areas. Specifically, we want to find out if the repartition of the security schemes is stable or not. Finally, we would like to use the data set as a basis for a geolocation framework; this would allow us to observe realistic mobility patterns of a group of users by giving them smartphones with a software continuously logging the surrounding Wi-Fi environment.

During the scanning campaigns, it was sometimes frustrating not to have a low-level access to the whole scanning system. For instance, we found out that in practice, it is more efficient to perform multiple quick scans than to wait longer for Probe Response frames on each channel. We are currently investigating the best way to implement a fine-grained scanner that injects its own 802.11 Probe Request frames, and logs Beacons frames along with Probe Response ones. The main objective of this tool is to produce PCAP [14] warwalking traces in order to achieve more detailed post-campaigns analysis.

Acknowledgment

The Python modules were developed as a part of the MOSOMUSO project with funding from the French Minister of Industry RIAM grant. The authors would like to acknowledge Laurent Bernaille for sharing his SQL wisdom, and Clémence Magnien and Matthieu Latapy for comments and valuable discussions.

References

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

[2] http://www.netstumbler.com.

[3] A. Akella, G. Judd, S. Seshan, and P. Steenkiste. Self Management in Chaotic Wireless Deployments. *Wireless Networks Journal (WINET), Special Issue on Selected Papers from MobiCom 2005*, September 2005.

[4] W.-F. Alliance. http://www.wi-fi.com.

[5] Y. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm. Accuracy characterization for metropolitan-scale wi-fi localization. 2005.

[6] T. Clausen and P. Jacquet. Optimized Link State Routing Protocol (OLSR). RFC 3626 (Experimental), Oct. 2003.

[7] IEEE. OUI assignments. http://standards.ieee.org/regauth/oui/index.shtml.

[8] K. Jones and L. Liu. What Where Wi: An Analysis of Millions of Wi-Fi Access Points. In *Proceedings of 2007 IEEE Portable: International Conference on Portable Information Devices*, May 2007.

[9] Kaspersky Lab. Wardriving in England. http://www.viruslist.com/, May 2006.

[10] Kaspersky Lab. Wardriving in Paris. http://www.viruslist.com/, December 2006.

[11] Kaspersky Lab. Wardriving in London. http://www.viruslist.com/, May 2007.

[12] M. Kim, J. Fielding, and D. Kotz. Risks of using AP locations discovered through war driving. May 2006.

[13] La lettre de l'autorité de régulation des communications électroniques et des postes. http://www.arcep.fr/uploads/tx_gspublication/lettre57.pdf, September 2007.

[14] libpcap and tcpdump. http://www.tcpdump.org.

[15] Nokia. http://wiki.forum.nokia.com/index.php/SDK_API_Plugin.

[16] Ozone, http://www.ozoneparis.net. Wardriving Ozone de Noël 2005. Technical report, December 2005.

[17] Place Lab. http://www.placelab.org.

[18] PlaceEngine. http://www.placeengine.com.

[19] Python for S60 devices. http://pys60.sf.net.

[20] Seattle WiFi Map Project. http://depts.washington.edu/wifimap.

[21] E. Tews, R.-P. Weinmann, and A. Pyshkin. Breaking 104 bit wep in less than 60 seconds. Cryptology ePrint Archive, Report 2007/120, April 2007.

[22] WiGLE. http://www.wigle.net.