Crowd Mobility Analysis using WiFi Sniffers

Anas Basalamah
Computer Engineering Department
Umm Al-Qura University
Makkah, Saudi Arabia 21955

***Abstract*—Wi-fi enabled devices such as today's smart-phones are regularly in-search for connectivity. They continuously send management frames called Probe Requests searching for previously accessed networks. These frames contain the sender's MAC addresses in clear text, which can be used as an identifier for that sender. Being able to sniff that MAC address at several locations allows us to understand the mobility behavior of that device. In this paper, we present a solar-powered, beagle-bone based standalone system that continuously sniffs the air for probes and extract their MAC addresses. We deployed the system in the world's largest gathering (The Hajj) and tested it at scale. Our objective was to build an infrastructure for non-invasive mass crowd analysis. Our deployment had a total of 8 sniffers covering a population of 185,000 people. We detected 37.5% of the population, analysed their arrival and departure behaviours, identified their smartphone manufacturers and extracted their transition patterns from one sub-location to another. By presenting valuable insights on the mobility of our target crowd, we validated the potential of our platform for crowd mobility analysis.**

Keywords*—WiFi Probes; Crowd Monitoring; Crowd Mobility Analysis

I. Introduction

Monitoring and understanding peoples behavior in public and private spaces like streets, parks, hospitals, malls is a big challenge for cities. Systems with information about mobility behavior of people are able to support the control and management process of pedestrians and vehicles and are able to reduce the management costs and travel times. Citizens can be informed about dense areas, traffic and travel times. Cities can better design public spaces to enrich the happiness of citizens.

Cameras and image processing technique have been used for years to detect crowd mobility [1][2]. However these techniques suffer from high implementation costs and accuracy limitations specially when tracking crowds across multiple feeds. Moreover, cameras raise major privacy concerns [3]. Therefore, cheaper, less invasive, more accurate techniques are encouraged to achieve crowd mobility monitoring.

The air around us is packed with WiFi packets transmitted from one device to another carrying massive amount of information. These packets originate from personal computers, laptop computers, ipads, TVs, and most importantly, the smartphones that we carry with us almost everywhere we go. The increasing usage of WiFi in smartphones offers new possibilities to monitor crowd mobility without the need for expensive additional hardware installations. This paper tries to leverage this unbreakable association between people and their smartphones to capture mobility patterns and understands people's mobility behaviors.

Wifi-enabled devices periodically broadcast management frames called probes in search for nearby access points. These probes are sent in clear text and carry useful data such as device MAC address, which acts as a unique identifier for this particular device. Reading these over-the-air packets using low cost sniffers enables us to identify the presence of people at the 50-100 meter range of the Wi-Fi sniffer. Placing several sniffers around the city (or monitored spaces) allows us to understand the mobility of crowds by the observing the density of probes, duel times, and mobility traces across multiple sniffers.

However, and to the best of our knowledge, a crowd mobility system based on Wi-Fi probes has not been realized in a scenario where hundred thousands of crowds mobilize in a controlled environment. We leveraged the worlds largest annual pilgrimage where 2-3 million people gather to perform the Hajj pilgrimage in the city of Makkah, Saudi Arabia. We deploy our system in a subset of a camp area called Arafat, where pilgrims occupy for only one day. The camp is infrastructure-less; no WiFi connectivity. This guarantees that there are no possible biases in the data from existing networks that could invalidate our findings. We installed 8 solar-powered WiFi sniffers at an isolated island inside Arafat where a population of 185,000 pilgrims gather, and collected data two days before and three days after the Arafat day.

Our results show that we were able to detect 69,467 devices leading to a detection rate of 37.5% of total population. 33.26% of which are iOS devices and remainder is Android and Windows phones. The data showed expected increase in arrival and decrease in departure as pilgrims arrived and departed the site. The data also showed expected mobility pattern between sniffer locations reflecting true mobility of pilgrims.

The contribution of this paper is of three folds:

- The paper presents an system architecture for wifi sniffing that is used to log all WiFi probe packets for mobility analysis. The system was built on Beaglebone and is powered by solar energy. Data was sent to servers using 3G in real-time.
- We deployed our system in a real-world massive scenario where more than 185,000 people mobilize. The deployment was in the worlds largest pilgrim event (The Hajj).
- By analyzing our collected dataset we demonstrate a set of valuable insights on the mobility of our target crowd validating the potential of our platform for crowd mobility analysis.

The system presented in this paper can be generalized to any public, private, indoor or outdoor space, and allows cities and businesses properly manage crowds and optimize their mobility.

II. Related Works

We have seen many explorations in the literature that address crowd mobility using passive sniffing of WiFi packets. In [4], Musa et al. deploy a real-world test bed for WiFi probe sniffing for traffic management in the city of Chicago. They present an approach for detecting wifi probes and estimating the spatio-temporal trajectories of smart phones. In [5], by only looking at WiFi probes, Barbera et al. was able to infer significant information on the social structure of a large crowd and its socioeconomic status. In [6], Chon et al. use smart phones as sniffers to demystify the potential and the threat that exist in sniffing wifi probes. They were able to showcase the effectiveness of estimating urban mobility via only a small number of participants. They also analyzed many issues related to sniffing WiFi probes such as feasibility, coverage, scalability, and threats to privacy. In [7], Bonn et al. deploy a WiFi sniffing experiment at a university campus and at a music festival. Their intention was to study the density estimates of crowds at those locations.

The limitation of previous works lies in the missing ground truth for real-world data and the lack of scalability of their deployments. Also not enough analysis was provided to demonstrate the validity of the WiFi probes for mass crowd analysis.

III. Methodology

In this section we explain the concept of sniffing WiFi probes and present the platform we built specifically for this purpose.

A. WiFi Probes

Smartphone manufacturers are keen in enabling users to connect to the internet as seamless as possible. As such, smartphones are designed to aggressively search for nearby Access Points (AP) and automatically re-connect to previously accessed networks that are saved on the smartphone. To achieve seamless auto-connectivity, a low-latency service discovery is required. A smartphone can discover nearby WiFi APs using a passive or an active discovery process [8]. In the passive mode, AP broadcast periodic beacon messages to advertise its presence within its coverage range. If the mobile device happens to be listening at that moment on that exact channel, it will receive this broadcast, otherwise it will be missed. After some time, the frame will be received and mobile devices will respond accordingly. Unfortunately, this approach leads to high discovery delays. In contrast, the active discovery mode requires mobile devices to continuously broadcast Probe Request messages at every channel consecutively to discover the nearby APs. This method is much faster in service discovery, and is the only way to find hidden networks. As such, mobile phones implement active mode and constantly send probe request messages. Empirical results with a variety of mobile devices show that an active scan is performed at least once within two minutes [7].

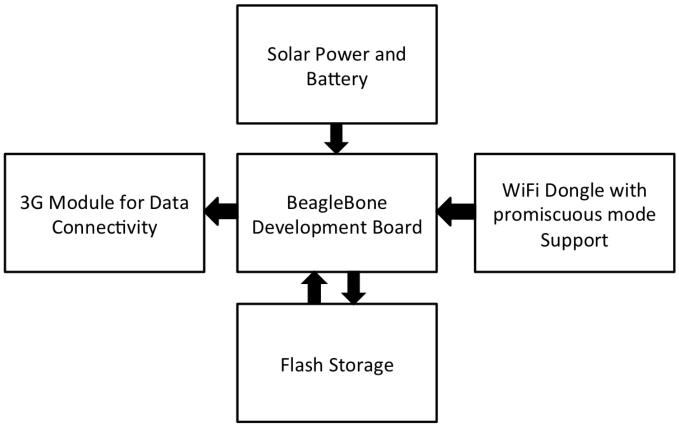
Solar Power and Battery
3G Module for Data Connectivity
BeagleBone Development Board
WiFi Dongle with promiscuous mode Support
Flash Storage

Fig. 1. Architecture for the sniffing platform

Probe requests frames carry the senders MAC address in clear text, which can be used as an identifier for the specific mobile phone. As a result, the more mobile devices with active WiFi are present in one area, the more corresponding WiFi probe frames are present, each with a unique identifier. Being able to capture those frames gives an indication of crowd density in that area. The ability to redetect the IDs at other locations can enable the measure of mobility between areas.

B. Sniffing WiFi Frames

There are several ways to sniff WiFi traffic and filter probe requests. There are some specialized products available for network professionals [9], and packages for mobile devices [10]. However, since we built our platform for a more stable setup, we decided to capture probes can using a type of wireless network interface cards (NIC) that operate on monitors mode, along with a linux-based packet capturing software (Wireshark[11]). Once monitor mode is enabled, the NIC is fixed on one frequency out of the Eleven frequencies defined in the IEEE802.11 standard. Since a probe is usually sent by each mobile device once on every channel, fixing one channel will increase the probability of capturing the probe frame.

C. WiFi Sniffing Platform

Fig. 1 shows the architecture of the platform we developed to capture WiFi probe requests and the components needed to operate it. The platform uses the beaglebone development board to run our linux based sniffing scripts, manage the data handling and communicate with the 3G module. We use a solar panel to charge a battery to run the platform 24h a day. A 3G dongle is used to communicate the data using a 3G network with unlimited data plan to the database. Flash storage was used to log the probe data temporarily before sending it to the 3G module. Fig. 2 shows the view of an installed system.

D. Deployment Design

Hajj is a unique annual event that happens only in Makkah, Saudi Arabia. Two to Three Million Muslims from all over the world gather to perform a pilgrimage referred to as Hajj. Hajj is a sequence of rituals performed within specific times and places. The Spatio-Temporal restrictions make the management of Hajj a very challenging task. It is only

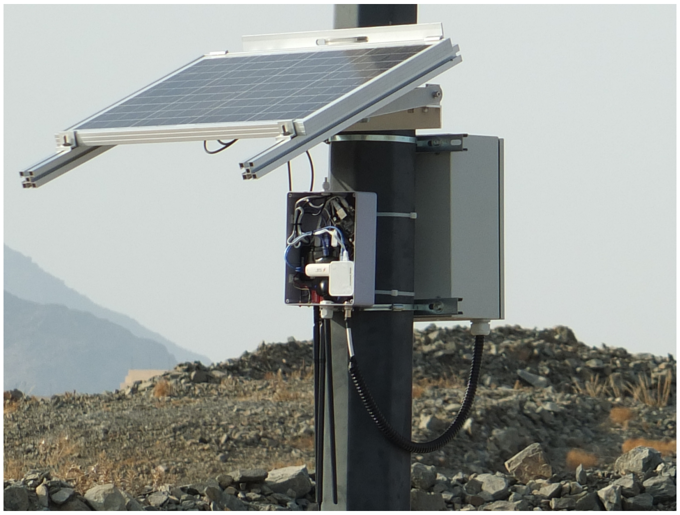

Fig. 2. Deployed Sniffing Platform

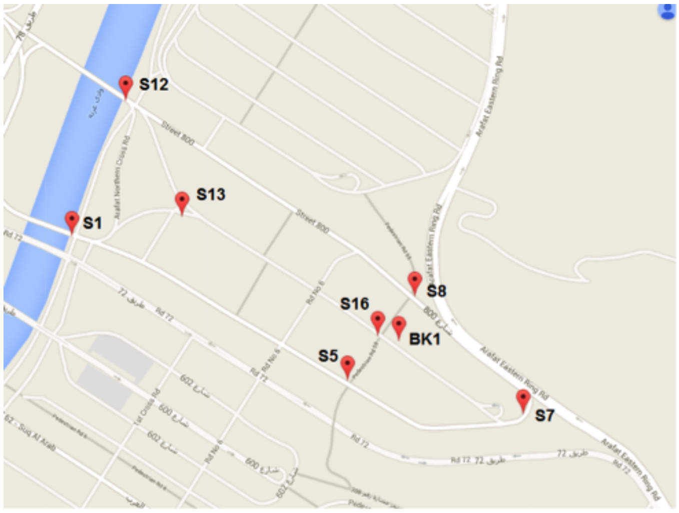

Fig. 3. Selected Location in Arafat

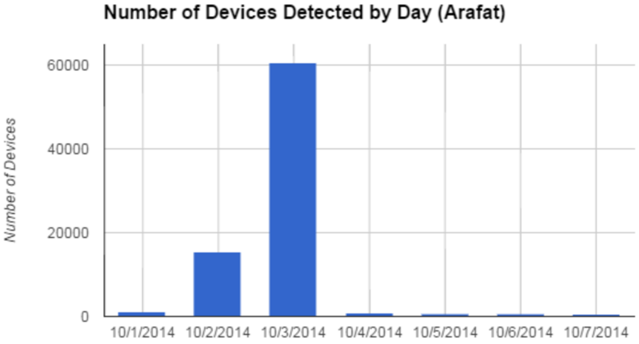

Fig. 4. Total unique Devices per day

by understanding how pilgrims behave, their patterns, their interactions, their needs and demands, that we can reach to a satisfying level of providing services and experiences. Considering the importance and the unique nature of the Hajj event, and enormous amount of data that can be collected for understanding crowd mobility, we decided to implement our system on a sub-location of Hajj, in a site called Arafat.

Arafat is located 12km southeast the Holy City of Makkah. It covers an area of 13m2 of desert camp. No infrastructure exists as people stay in temporary tents. The camp is deserted throughout the year, except for the day of Arafat where all pilgrims spend their day before mobilizing to Muzdalifa (a neighboring camp) on that evening. Few management staff arrive few days before and leave few days after. Also some pilgrims arrive a day before to avoid traffic congestion.

We have chosen a contained island inside Arafat allocated for pilgrims coming from Turkey, Europe and the Americas. For management purposes, this area is sieged such that no other pilgrims can come in. The estimated size of population from those countries is estimated to be 185,000 pilgrims from other data sources. We placed 8 sniffing platforms in Arafat as shown in Fig. 3. The locations have been chosen carefully to capture all possible traffic in and out of the camp. Arafat day was on Oct 3rd which would exhibit the maximum number of detections in our data-set.

E. Data Analysis

In this section, we show the data analysis of the data collected on this deployment.

1) Basic Summery Statistics: The data we collected started Oct 1st and ended Oct 6th. The number of Wi-Fi probe records captured thought-out these days was 4,517,687. These probes resulted from 69,467 unique devices. Since we know the estimated ground truth for the number of pilgims in this subarafat area is 185,000 pilgrims, we were able to detect 37.5% of the population. This number is very significant. It tells us that 37.5% of pilgims in that location carry smartphones with Wi-Fi enabled and can be tracked. While it is obvious that we cannot generalize this number with all pilgrims from all nationalities due to economical differences, it does indicate the potential of our system in capturing detailed traces of pilgrims non-invasively.

It is also worth nothing that the average number of probes per device was 65.03.

2) Unique Detections per day: Fig. 4 shows the total number of detections per day considering all the sites combined. It is obvious that most of the detections happen on Arafat day (3rd of October) reaching up to 60,000 unique devices. The data also shows a large number of devices detected on the 2nd of October. This number is due to the arrivals of many pilgrims the day before Arafat day to avoid traffic. Although ground truth data for 2nd of October is not available to compare with this number, we found this result very insightful and would benefit the Hajj administration.

Fig. 5 shows the average of Unique number of devices shown per site per day. From the figure we notice that inner located sites such as 5, 8 and 16 show high detections compared to outer located sites such as 1 and 12. Generally, the reason is that outer sites are located at entry and exit points we pilgrims pass through quickly, leading to less detection opportunity compared to inner located sites were pilgrims spend most of their day increasing the likelyhood of detection.

3) Unique Detections per Hour: Fig. 6 shows the number of total unique detections per hour on Arafat day. The figure

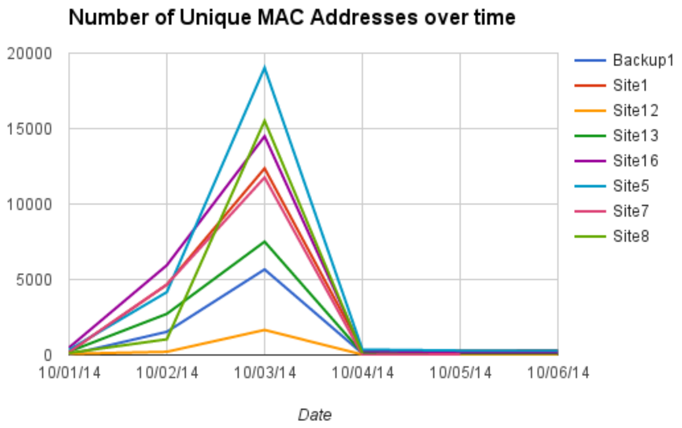

Fig. 5. Unique devices per pay per site

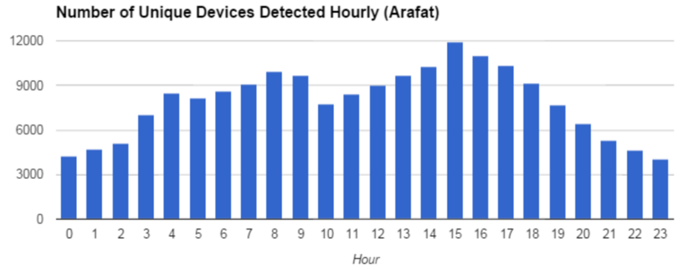

Fig. 6. Unique devices per Hour on Arafat Day

shows a peak at 6-9AM and another peak at 3-5PM. Those peaks successfully detect largest mobility activities of the day. In the morning peak, people come to Arafat, in the Afternoon peak, people leave to Muzdalifa. The pattens successfully show the overall activity at Arafat.

4) Device Manufacturers: The MAC addresses captured in the data-set allows us to map them to their manufacturers. Table I shows the list of mobile phone manufacturers. The numbers shown are really interesting because they can indicate the economic distribution of this population. 33.36% are Apple users compared to 26.98% Samsung users. 12.7 device manufacturers were un-known due to lack of information. Only around 13.79% of devices were from low cost unfamiliar manufacturers.

5) Time between detections: The time between detections is presented in Table II. This data shows that the majority of detections occur in sequence of less than a second (2,634,117 detections). This indicates the frequency of probe detections and the efficiency of using it as a tool for crowd analysis.

6) Transition from one sub-location to another: Table III shows the percentages of mobile devices that are detected at one sniffer then detected at another sniffer. As an example, 85.78% devices being detected at 13 are detected next at 16. The opposite is only 50.35% which indicate that the majority of people mobilize from coverage location 13 to 16.

TABLE I. THE TYPES OF DEVICE MANUFACTURERS

Manufacturer	Number of Records	Percentage
Apple	23104	33.26
Samsung Electronics Co.	18742	26.98
Unknown	8823	12.7
Murata Manufactuaring Co.	6723	9.68
Nokia	2494	3.59
Total for Top 5	59886	86.21

TABLE II. TIME BETWEEN DETECTIONS

Time between detections(sec)	Number of Detections
0	2,634,117
1	181,820
2	56,810
3	39,738
4	41,967
5	32,977
6	31,875
7	23,582
8	19,280
9	21,535
10	27,698
11-60	676,542
61-300	418,268
301-600	84,945
600-3600	102,286
3600-7200	19,903
7200+	34,877

TABLE III. TRANSITION MATRIX

	BK1	S1	S12	S13	S16	S5	S7	S8
BK1		4.24	2.25	1.82	33.72	1.92	1.07	25.18
S1	1.69		69.23	8.37	3.91	3.15	1.54	15.97
S12	0.38	2.08		0.96	1.01	0.43	0.57	8.23
S13	1.86	25.56	8.17		50.35	0.67	0.39	4.38
S16	86.19	25.2	9.59	85.78		3.91	2.63	22.72
S5	4.86	31.71	5.21	2.1	4.6		89.32	13.64
S7	2.16	7.15	1.66	0.55	2.17	86.21		9.48
S8	2.79	4.02	3.79	0.4	1.86	3.66	4.48	

IV. CONCLUSION

This paper presented a system for crowd behaviour analysis using non-invasive WiFi probes. We presented a system architecture for wifi sniffing that is used to log all WiFi probe packets for mobility analysis. The system was built on Beaglebone and is powered by solar energy. Data was sent to servers using 3G in real-time. We deploy our system in a real-world massive scenario where more than 185,000 people mobilize. The deployment was in the worlds largest pilgrim event (The Hajj). By analyzing our collected dataset we demonstrate a set of valuable insights on the mobility of our target crowd validating the potential of our platform for crowd mobility analysis. The system presented in this paper can be generalized to any public, private, indoor or outdoor space, and allows cities and businesses properly manage crowds and optimize their mobility.

REFERENCES

[1] A. Marana, S. Velastin, L. Costa, and R. Lotufo, “Automatic estimation of crowd density using texture,” *Safety Science*, vol. 28, no. 3, pp. 165 – 175, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925753597000817

[2] H. Rahmalan, M. S. Nixon, and J. N. Carter, “On crowd density estimation for surveillance,” in *2006 IET Conference on Crime and Security*, June 2006, pp. 540–545.

[3] D. Gavrila, “The visual analysis of human movement,” *Comput. Vis. Image Underst.*, vol. 73, no. 1, pp. 82–98, Jan. 1999. [Online]. Available: http://dx.doi.org/10.1006/cviu.1998.0716

[4] A. B. M. Musa and J. Eriksson, “Tracking unmodified smartphones using wi-fi monitors.” in *SenSys*, M. R. Eskicioglu, A. Campbell, and K. Langendoen, Eds. ACM, 2012, pp. 281–294.

[5] M. V. Barbera, A. Epasto, A. Mei, V. C. Perta, and J. Stefa, "Signals from the crowd: Uncovering social relationships through smartphone probes," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13. New York, NY, USA: ACM, 2013, pp. 265–276. [Online]. Available: http://doi.acm.org/10.1145/2504730.2504742

[6] Y. Chon, S. Kim, S. Lee, D. Kim, Y. Kim, and H. Cha, "Sensing wifi packets in the air: Practicality and implications in urban mobility monitoring," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14. New York, NY, USA: ACM, 2014, pp. 189–200. [Online]. Available: http://doi.acm.org/10.1145/2632048.2636066

[7] B. Bonn, A. Barzan, P. Quax, and W. Lamotte, "Wifipi: Involuntary tracking of visitors at mass events," in *2013 IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, June 2013, pp. 1–6.

[8] S. Seneviratne, A. Seneviratne, P. Mohapatra, and P.-U. Tournoux, "Characterizing wifi connection and its impact on mobile users: Practical insights," in *Proceedings of the 8th ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation &#38; Characterization*, ser. WiNTECH '13. New York, NY, USA: ACM, 2013, pp. 81–88. [Online]. Available: http://doi.acm.org/10.1145/2505469.2505480

[9] "WIFI PINEAPPLE MARK V STANDARD (Last Accessed November 23, 2015)," https://hakshop.myshopify.com/products/wifi-pineapple.

[10] "Android PCAP Capture - Utility for capturing raw 802.11 frames. (Last Accessed November 23, 2015)," https://www.kismetwireless.net/android-pcap/.

[11] "WLAN (IEEE 802.11) Capture Setup using WireShark (Last Accessed November 23, 2015)," https://wiki.wireshark.org/CaptureSetup/WLAN.