SCREAM 'N SHOUT

What if devices were constantly
screaming out our personal information?

ANDREA SILVANO
2020/2021

Introduction

I will be collecting wi-fi data from students at Politecnico di Milano. In particular the collection will focus on **probe requests**.

“Probe requests are sent by a station to elicit information about access points, in particular to determine if an access point is present or not in the nearby environment. Some devices (mostly smartphones and tablets) use these requests to determine if one of the networks they have previously been connected to is in range, leaking their preferred network list (PNL) and, therefore, your personal information.”

from ProbeQuest

Creation of dataset

Once ProbeOSX begins its collection, it then returns several information:

- DATE: timestamp of collection
- STRENGTH OF SIGNAL: is the device close to me?
- MAC ADDRESS: unique identification code for device
- SSID: Name of the network the device is looking for
- VENDOR: What brand is the device linked to?

I'll be mainly focusing on SSID and VENDOR information, which are probably the most sensitive information leaked.

If our devices could scream, what would they say?

Lots of automated processes are happening at the same time when we use our devices, and we don't even know they are happening. What if we could give voice to this processes, thus understanding more and more what our devices are constantly up to?

What if they could "Scream 'n shout and let it all out"? As the famous song from will.i.am ft. Britney Spears "Scream & Shout" says.

The image on the side is taken from a project I previously created on screaming and translation of human emotions in 2019, called URLOGRAFO

andreasilvano.github.io/scream-n-shout/

SCREAM 'N SHOUT

Our mobile devices, they just let it all out!

WHAT DOES THIS MEAN?

Mind Your probes!

The reason why exposing personal information as simple as wifi probes could be dangerous is explained in this research paper from *Università La Sapienza di Roma*.

“In this paper we present the idea of exploiting WiFi probe requests to de-anonymize the origin of participants in large events. We make use of several, publicly available datasets containing more than 11M of probe requests collected in scenarios that are of citywide, national (two political meetings), and international religion-related relevance. We show how, by exploiting the semantic information brought by the relative WiFi probes, we are able to discover with high accuracy the provenance of the crowds in each event.”

from Mind Your Probes

Mind Your Probes: De-Anonymization of Large Crowds Through Smartphone WiFi Probe Requests

Adriano Di Luzio*, Alessandro Mei, and Julinda Stefa
Department of Computer Science, Sapienza University of Rome, Italy.
Email: *diluzio.1487872@studenti.uniroma1.it, {mei, stefa}@di.uniroma1.it.

***Abstract*—Whenever our smartphones have their WiFi radio interface on, they periodically try to connect to known wireless APs (networks the user has connected to in the past). This is done through WiFi Probe requests—special wireless frames that contain the MAC address of the sending device and, in most of the cases, the human-readable name-string (SSID) of the known AP. This semantic information, inherent to the network protocol, is sent in the clear and, if sniffed, can help discover important information and phenomena of people and human nature that have nothing to do with technology.**

In this paper we present the idea of exploiting WiFi probe requests to de-anonymize the origin of participants in large events. We make use of several, publicly available datasets containing more than 11M of probe requests collected in scenarios that are of citywide, national (two political meetings), and international religion-related relevance. We show how, by exploiting the semantic information brought by the relative WiFi probes, we are able to discover with high accuracy the provenance of the crowds in each event. In particular, the de-anonymization outcome of the two political meetings held few days before the election days in Italy match surprisingly well the official voting results reported for the two respective parties.

***Index Terms*—Privacy, WiFi probe requests, social sciences computing, social networks.**

I. Introduction

Posting on social platforms, contacting friends and family, online banking, emails, entertainment, almost anything we could need is, nowadays, just a touch of our thumb away. All thanks to our smartphones. They have undoubtedly changed the way we interact with technology. Not only do they allow us to navigate anytime and everywhere, but they are built to do so in the most efficient way. Take the wireless interface for example: If on, it automatically connects to WiFi networks, even if we are already covered by a 3G network (known to be more expensive and less energy efficient than WiFi connectivity). This is enabled by a type of special wireless frame called *WiFi probe request* [1] that our devices periodically send in the clear to discover the availability of known WiFi networks in range. As we will further discuss in Section II-A, the probe requests contain the MAC address that uniquely identifies the sending device. The MAC address can be the device’s real universal address or a temporary address if MAC randomization is used—as it is done in the very latest versions of mobile operating systems. Probes can be of *broadcast* type—not specifically directed towards a particular WiFi network—or *directed*—specifying the SSID (string name identifier) of a particular WiFi network.

Directed probes grant a highly efficient, reliable, and automatic network discovery. Most of the current mobile OSs make use of the device PNL (Preferred Network List) and adopt directed probes to request the availability of WiFi networks to which the user has connected before. In this way our devices are able to automatically, in just a few seconds, switch to our “Home WiFi” as soon as we cross the doorstep. But there is much more than that: The SSIDs of the WiFi networks contained in these frames, inherent to the technological side of the connection protocol, is full of semantic information. This is what makes directed probes very valuable from a sociological point of view: They can help discover aspects of human nature that have nothing to do with technology. Indeed, by just *listening* to what smart-phones are *shouting* through their probes it is possible to draw a detailed picture of the people surrounding us. As we will also discuss in Section V, many insightful works have shown how wireless probes can be used to infer the relationships among people [2], predict who they will meet [3], [4], or even discover the welfare of large crowds [5].

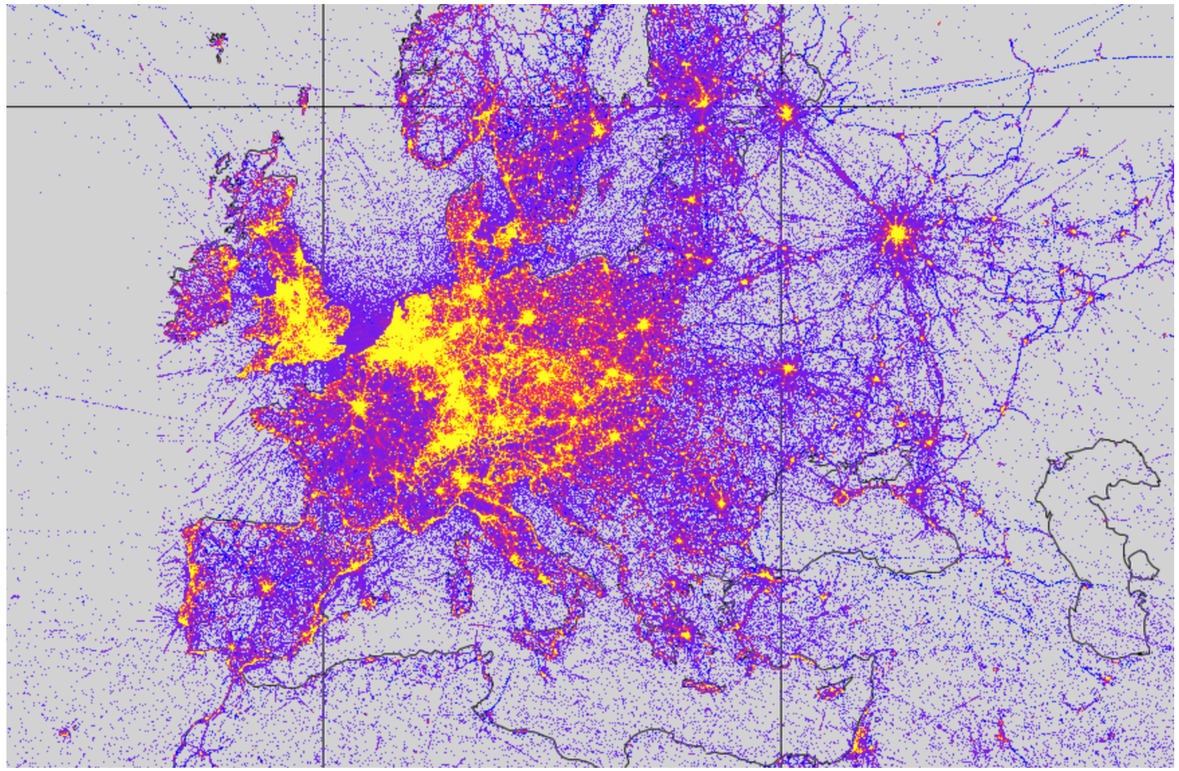
In this paper we take a step forward towards the understanding of human nature through probe requests. Our goal is to uncover, with high accuracy, the geographical provenance of people in large gatherings. Our de-anonymization process, described in detail in Section III, is based on the probe request frames released by their mobile devices just by default; i.e., not requiring any intervention neither by the device owners, nor by any other. Upon the tiny pieces of information included in these frames we build an automatic methodology to de-anonymize the provenance of tens of thousands of people participating in gatherings of citywide, national (political meetings of two parties held around election days), and international (religion related) events, lasting just a few hours each. Finally, we test our de-anonymization methodology by comparing its outcome with ground truth data—the official election results for the two parties whose meetings, held around election days, we target in this paper. The comparison shows that the result of our de-anonymization match surprisingly well the official general election results of the two parties (Section IV).

To the best of our knowledge, this is the first work that shows how to achieve this amount of detailed knowledge on large crowds of people based solely on their probe requests. Knowledge whose impact and applications, as we further discuss in Section VII, can span from advertising of commercial

Our networks on a map.
The WiGLE website

WiGLE (or Wireless Geographic Logging Engine) is a website for collecting information about the different wireless hotspots around the world. Users can register on the website and upload hotspot data like GPS coordinates, SSID, MAC address and the encryption type used on the hotspots discovered. In addition, cell tower data is uploaded and displayed.

from Wikipedia



SCREAM 'N SHOUT

What if devices were constantly
screaming out our personal information?