Department of Computer Science
Series of Publications A
Report A-2020-7

Privacy-Aware Opportunistic Wi-Fi

Otto Waltari

Doctoral dissertation, to be presented for public examination with the permission of the Faculty of Science of the University of Helsinki, in Auditorium CK112, Exactum, Kumpula, Helsinki on the 8th of October, 2020 at 10 o'clock before noon.

University of Helsinki
Finland

Supervisor
Jussi Kangasharju, University of Helsinki, Finland

Pre-examiners
Jukka Manner, Aalto University, Finland
Edith Ngai, Uppsala University, Sweden

Opponent
Andrea Passarella, Institute for Informatics and Telematics (IIT), National Research Council (CNR), Pisa, Italy

Custos
Jussi Kangasharju, University of Helsinki, Finland

Contact information

Department of Computer Science
P.O. Box 68 (Pietari Kalmin katu 5)
FI-00014 University of Helsinki
Finland

Email address: info@cs.helsinki.fi
URL: http://cs.helsinki.fi/
Telephone: +358 2941 911

ISSN 1238-8645
ISBN 978-951-51-6621-0 (paperback)
ISBN 978-951-51-6622-7 (PDF)
Helsinki 2020
Unigrafia

Privacy-Aware Opportunistic Wi-Fi

Otto Waltari

Department of Computer Science
P.O. Box 68, FI-00014 University of Helsinki, Finland
Otto.Waltari@cs.helsinki.fi
https://www.cs.helsinki.fi/u/owaltari/

PhD Thesis, Series of Publications A, Report A-2020-7
Helsinki, September 2020, 51+44 pages
ISSN 1238-8645
ISBN 978-951-51-6621-0 (paperback)
ISBN 978-951-51-6622-7 (PDF)

Abstract

Over the past decade Internet connectivity has become an increasingly essential feature on modern mobile devices. Many use-cases representing the state of the art depend on connectivity. Smartphones, tablets, and other devices alike can even be seen as access devices to Internet services and applications. Getting a device connected requires either a data plan from a mobile network operator (MNO), or alternatively connecting over Wi-Fi wherever feasible. Data plans offered by MNO's vary in terms of price, quota size, and service quality based on regional causes. Expensive data, poor cell coverage, or a limited quota has driven many users to look for free Wi-Fis in hopes of finding a decent connection to satisfy the ever-growing transmission need of modern Internet applications.

The standard for wireless local area networks (WLAN, IEEE 802.11) specifies a network discovery protocol for wireless devices to find surrounding networks. The principle behind this discovery protocol dates back to the early days of wireless networking. However, the scale at which Wi-Fi is deployed and being utilized today is magnitudes larger than what it used to be. In more recent years it was realized that the primitive network discovery protocol combined with the large scale can be used for privacy violations. Device manufacturers have acknowledged this issue and developed mechanisms, such as MAC address randomization, for preventing e.g. user tracking based on Wi-Fi background traffic. These mechanisms have been proven to be inefficient.

The contributions of this thesis are two-fold. First, this thesis exposes problems related to the 802.11 network discovery protocol. It presents a highly efficient Wi-Fi traffic capturing system, through which we can show distinct characteristics in the way how different mobile devices from various brands and models scan for available networks. This thesis also looks at the potentially privacy-compromising elements in these queries, and provides a mechanism to quantify the information leak. Such collected information combined with public crowdsourced data can pinpoint locations of interest, such as home, workplace, or affiliation without user consent. Secondly, this thesis proposes a novel mechanism, *WiPush*, to deliver messages over Wi-Fi without association in order to avoid network discovery entirely. This mechanism leverages the existing, yet mostly inaccessible Wi-Fi infrastructure to serve a wider scope of users. Lastly, this thesis provides a communication system for privacy-preserving, opportunistic, and lightweight Wi-Fi communication without association. This system is built around an inexpensive companion device, which makes the concept adaptable for various opportunistic short-range communication systems, such as smart traffic and delay-tolerant networks.

Computing Reviews (2012) Categories and Subject Descriptors:

Networks → Wireless access networks
Networks → Network privacy and anonymity
Security and privacy → Privacy protections

General Terms:
Wi-Fi, Privacy, Tracking, Communication systems

Additional Key Words and Phrases:
Device fingerprinting, Traffic monitoring, Opportunistic communication

Acknowledgements

I am deeply grateful to my supervisor Professor Jussi Kangasharju for giving me the trust, support, and guidance in finishing this project. It has been an immeasurably valuable lesson with both its ups and downs. During this journey, I have been privileged to collaborate with distinguished researchers through various research visits. Foremost, I would like to thank Dr. Fahim Kawsar and Dr. Utku Günay Acer for hosting my research visit at Bell Labs. The three-month visit provided me invaluable experience in research and the industry in general. I would also like to thank other projects and institutions, namely BCDC, EIT ICT Labs, and HIIT, for offering numerous encounters and insightful discussions with researchers from different fields and backgrounds.

I will forever be thankful to the Department of Computer Science, and to all of its wonderful staff for contributing to a pleasant studying and working environment. Having completed all B.Sc., M.Sc., and Ph.D. degrees at the same department, I hereby consider myself to have received the full experience. I explicitly want to thank DoCS, and Dr. Pirjo Moen in particular, for all the support and guidance required to finish this journey. Last but not least, I want to thank my former colleagues in the Collaborative Networking research group: Ossi, Nitinder, Aleksandr, Pengyuan, Walter, Suzan, Julien, Liang, and Mikko.

After all, more important than *where* you work — is *who* you work with.

Thank you.

Helsinki, September 2020
Otto Waltari

Contents

List of Reprinted Publications

Research Paper I: Otto Waltari and Jussi Kangasharju. 2016. The Wireless Shark: Identifying WiFi Devices Based on Probe Fingerprints. In Proceedings of the First Workshop on Mobile Data (MobiData '16). ACM, New York, NY, USA, 1-6.

> *Contribution:* This publication was led by the author who formulated the problem, built the capturing system, implemented software modifications for using the system, and eventually performed the measurements. Prof. Jussi Kangasharju and several former members of the Collaborative Networking research group were involved in discussing and brainstorming the results.

Research Paper II: Waltari O., Kangasharju J. (2018) Quantifying the Information Leak in IEEE 802.11 Network Discovery. In: Chowdhury K., Di Felice M., Matta I., Sheng B. (eds) Wired/Wireless Internet Communications. WWIC 2018. Lecture Notes in Computer Science, vol 10866. Springer, Cham

> *Contribution:* The author was in charge of formulating and writing this publication. Ideas behind the paper were extensively discussed with Prof. Jussi Kangasharju. Data sets collected at public locations were carried out by M.Sc. student Kalle Lammenoja under Prof. Jussi Kangasharju's supervision. In addition to writing the paper, the lead author was responsible for SSID classification, formulating *uniqueness*, and carrying out *active vs. passive* measurements.

Research Paper III: Utku Günay Acer and Otto Waltari. 2017. WiPush: Opportunistic Notifications over WiFi without Association. In Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2017). ACM, New York, NY, USA, 353-362.

Contribution: The research behind this publication was led by Dr. Utku Günay Acer, who designed the communication model and the push notification system. The author participated in this research during his internship at Bell Labs, Alcatel-Lucent in 2015. The visit was hosted by IoT Research Activity led by Prof. Fahim Kawsar. The author was responsible for implementing the Wi-Fi communication protocol on an Android device. Along with providing the technical and low-level programming skill, the author participated extensively in discussing and designing the concept of unassociated communication over Wi-Fi.

Research Paper IV: Otto Waltari and Jussi Kangasharju. 2020. Prongle: Lightweight Communication over Unassociated Wi-Fi. In The 35th ACM/SIGAPP Symposium on Applied Computing (SAC '20), March 30-April 3, 2020, Brno, Czech Republic. ACM, New York, NY, USA, 8 pages.

Contribution: The author was in charge of planning and writing this publication. The Prongle system presented in this paper was designed, implemented and evaluated by the author. The application scenarios and use cases were discussed in depth within the Collaborative Networking research group supervised by Prof. Jussi Kangasharju. These discussions were of great importance while laying out the groundwork for this publication.

List of Abbreviations

ACK	Acknowledgement
AP	Access Point
BSS	Basic Service Set
DA	Destination Address
DTN	Delay-Tolerant Nework
ESS	Extended Service Set
ESSID	Extended Service Set Identifier
GAS	General Advertisement Service
ICMP	Internet Control Message Protocol
IE	Information Element
ISP	Internet Service Provider
IBSS	Independent Basic Service Set, often referred to as *"ad hoc"*
MAC	Medium Access Control
NIC	Network Interface Card
OUI	Organizationally Unique Identifier
PHY	Physical Layer
PII	Personally Identifiable Information
PNL	Preferred Networks List
RSSI	Receive Signal Strength Indicator
RTT	Round-Trip Time
SC	Sequence Control
SSID	Service Set Identifier
SA	Source Address
STA	Station, a logical Wi-Fi entity
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UUID	Universally Unique Identifier
WLAN	Wireless Local Area Network
WNIC	Wireless Network Interface Card
WPAN	Wireless Personal Area Network

Chapter 1

Introduction

Internet connectivity has become an invaluable feature on mobile devices. Many of the smart applications that have gained foothold in our everyday life depend on Internet connectivity. As a well-known example, various social media platforms have been adopted so profoundly that some even experience distress because of a *fear of missing out* (FOMO) shortly after ending up disconnected. Social connections, hobbies, and everyday tasks in general are nowadays quite dependent on e.g. instant messaging platforms and other online services. Even voice calls are slowly but steadily being shifted to VoIP-calls that operate over a packet switching network. Whether the use-case concerns social media, general Internet queries, news, audio or video content, usually the content requires a data connection to a cloud service hosted by data centers around the world.

There are generally two ways to get a personal smart device connected. One way is to have a data plan included in a mobile subscription. These data plans have differences in pricing, quota, and service quality depending on regional factors and available mobile network operators (NMOs). Finland being the top country in data consumed by mobile subscription[1] is a class example; nationwide coverage, modern infrastructure and MNO competition has led to subscriptions with *basically unlimited everything* for a flat rate. Unlike in many other countries, where a typical data plan includes a quota of few gigabytes at a fixed price, and for anything beyond that an extra charge incurs. For example, the average price per 1 GB of mobile data in the U.S. in 2019 was $12.37[2]. Data roaming while traveling abroad can also become quite expensive for someone not fully aware of the circumstance and not knowing how much applications actually consume

[1] http://www.oecd.org/sti/broadband/broadband-statistics

[2] https://www.cable.co.uk/mobiles/worldwide-data-pricing

data. This uncertainty in usage-based billing has led many to look for an alternative way to get connected.

The other way to get connected is through Wi-Fi. Something that likely started as a nice addition to conference venues and hotel services has since become an important asset for anyone traveling. Complimentary Wi-Fi, also commonly known as *free Wi-Fi*, is nowadays available at practically any hotel and airport. A study conducted in 2017 [56] shows that for up to 75% of survey respondents free Wi-Fi is a deciding factor when choosing a hotel. For roughly 50% of the respondents free Wi-Fi affects the choice of airlines and restaurants. The popularity and extent of how much free Wi-Fis are sought after varies per country. As an example, since Finland has been a forerunner in deploying and providing cellular data, there has not emerged a solid demand for public Wi-Fis. Sparse population has also alleviated congestion-related problems caused by the so-called *mobile data explosion* during the last decade. Some commercial hotspot providing services have been available[3], but these are more enterprise-oriented solutions and deployed primarily at congress centers and business hubs.

However, on a global scale technologies like Wi-Fi offloading [7, 48] are a relevant topic. According to Cisco [2] up to 59% of all mobile data will be offloaded to Wi-Fi by 2022. Exponential growth in mobile data causes congestion problems for mobile network operators, and deploying Wi-Fi hotspots for customers at strategically chosen locations can alleviate these problems. Many operators complement their mobile subscriptions by offering unlimited data over a network of Wi-Fi hotspots they provide. User device association to such hotspot networks is often pre-programmed by mobile subscription retailers. Typically, Wi-Fi access is provided in exchange for purchased goods or services, but it may also be public and require no premises for joining. Such hotspots, typically located at malls, restaurants, and cafés, can also be seen as attractions to potential customers. As an example, a major U.S.-based coffee shop franchise has teamed up with Google, and have since been known to provide decent quality free Wi-Fi at each location of their business. This has become common knowledge, and almost anyone used to traveling knows which coffee shop to look for when in need of Internet and coffee.

Even if a Wi-Fi is advertised to be free of charge, it may come at a price. A study conducted by Norton in 2017 [56] revealed that 92% of Americans admit having taken risks in form of accessing e.g. online banking services over public Wi-Fi. Merely 27% use VPN when using public hotspots, while up to 41% are not able to distinguish a secure Wi-Fi from an insecure one.

[3] e.g. Sonera Homerun, DNA WLAN

There are various kinds of risks involved in using free Wi-Fis. To begin with, associating to a free Wi-Fi implies trust between the user and the network provider. Explicit claims of a Wi-Fi being secure can not always be trusted since a fraudulent access point would make the same claims. As an example, *evil twin* and *rogue* access points [15] are a common way execute Man-in-the-Middle (MitM) style phishing attacks wherever there is free Wi-Fi available [10,69]. Such attacks may be attempted by anyone because Wi-Fi operates over an unlicensed band, and therefore eavesdropping and transmitting fraudulent data frames is merely a matter of programming and requires no special hardware. This is generally not considered to be a problem as long as data channels are encrypted and secured accordingly. However, various applications and even personal demographic information can be identified by analyzing the characteristics of data flows [11, 12]. Disregarding such risks – or users simply not being aware of them – free Wi-Fis are widely used and sought after.

The aforementioned offloading culture combined with the continuous user-driven search for free networks at disposal has led to a situation where Wi-Fi is always kept enabled. As a result, keeping Wi-Fi enabled causes a device to intermittently query for networks to associate with. These queries are eavesdroppable and may violate user privacy unknowingly and without consent. This intentional device behavior enables various suspicious activities, such as user tracking and profiling. These problems are widely acknowledged and have since been addressed with mechanisms such as MAC address randomization [31, 67]. Unfortunately practical software implementations from device manufacturers have been shown to be ineffective one after another [50, 52, 73].

Contributions of this thesis are two-fold. First, this thesis demonstrates privacy problems caused by a combination of two habits; *i*) keeping Wi-Fi enabled at all times, and *ii*) using random available free networks. We start by presenting a multichannel Wi-Fi monitoring system, which reveals a new identifying characteristic in the way devices query intermittently for available and previously known networks. We then use this system to collect data at various locations, and show that a privacy-violating network discovery mechanism is still used by roughly one third of seen devices. We show that exposing SSID names of previously associated networks can invalidate the effects of any MAC address randomization mechanism. We then introduce a metric to quantify the information leak caused by this mechanism, and ultimately evaluate an alternative discovery mechanism, which does not violate privacy or allow tracking. Secondly, this thesis proposes novel ways of using Wi-Fi without association. The rationale behind

this is to avoid two things; *i*) privacy-compromising network discoveries, and *ii*) the need for associating to random free Wi-Fis. We first present an opportunistic way to deliver push notifications over Wi-Fi without association. This system leverages the high density of already deployed Wi-Fi access points and proposes a network-centric mechanism to deliver contextual notifications. We then present an infrastructure-less, association-free, and opportunistic Wi-Fi communication system for various novel use cases, such as smart traffic and delay-tolerant networks.

1.1 Background and Motivation

Wi-Fi is a commonly used trademark name for the *Wireless Local Area Networking* (WLAN, IEEE 802.11, [1]) standard belonging to the IEEE 802 family of standards. Wi-Fi was designed to work seamlessly together with Ethernet and effectively replace the last hop with a wireless link. This would then provide mobility to devices like laptop computers. Both Wi-Fi and Ethernet implement the bottom two layers of the OSI networking stack. Due to a different medium, their approach on the *physical layer* differs, while on the *data link layer* they share many similarities for the sake of seamless interoperability. However, extending e.g. a company intranet over Wi-Fi requires much stronger access control, since potential intruders do not need physical access to network wires. Hence, authentication has been a core feature in Wi-Fi from the beginning.

In order for a client device to interact with a Wi-Fi network it has to discover the network first. The standard specifies a network discovery protocol which involves sending out *probe requests* from the client-side, and the access point (AP) replying with a *probe response*. This is often referred to as *active network discovery*, although not specified as such by the standard. Active network discovery was strictly necessary for discovering so-called hidden networks. For a long time there was a myth about hidden APs being more secure and less prone to intrusion attacks than the ones periodically advertising their presence. These myths have been busted in literature and it is been generally acknowledged that hidden networks provided merely a false impression of security [64]. Despite hidden networks are deprecated, the network discovery protocol still used today reminds us of their existence. There is a dedicated field in probe requests for a list of network names, i.e. service set identifiers (SSIDs), the client is looking for. Probe requests looking for specific networks are known as *directed* probes.

As of today, most APs transmit *beacons* at regular intervals to advertise the SSID for the network they offer, and thus directed probes are not

needed. However, for unjustified reasons directed probes are still used in vain, but more importantly, they introduce a potential privacy violation. The extent at which Wi-Fi is deployed and used today has led to a situation where anyone with a laptop can start collecting lists of network SSIDs that surrounding Wi-Fi enabled devices have been associated to in the past. Connecting the dots in these lists with the help of external crowdsourced access point mapping (a.k.a. wardriving) services can reveal a surprising amount of *personally identifiable information* (PII) without the target user knowing anything about the information leak.

The scale and widespread use of Wi-Fi was probably not expected back when active network discovery was on the drawing board. The Wi-Fi standard has since been revised and amended regularly, but plaintext SSID names in background traffic still persist. The Wi-Fi standard does support an alternative network discovery method referred to as *passive discovery*. However, many device manufacturers and especially their operating system departments still put out products which employ active network discovery. The mentality appears to be such that since active network discovery still works, it does not need fixing.

Various novel networking paradigms, such as opportunistic and delay-tolerant networking (DTN) [27], require nimble communication mechanisms to operate. Since communication encounters may be short and sparse, overhead in establishing the communication channel should be minimized. The conventional way of communicating over Wi-Fi implies first discovery of an appropriate service set, followed by authentication and association. This whole procedure can take several seconds, which could be long enough to defeat the purpose of establishing a connection in the first place. Several Wi-Fi variations [26, 71, 79] have been proposed, but they often do not gain foothold due to complex deployment and low-level modifications to devices. There is even an IEEE standard for Wi-Fi mesh networks [37], but devices supporting it off-the-shelf are uncommon.

The motivation for this thesis goes back to creating mobility models for opportunistic and delay-tolerant networks. The user traces for these models were to be based on background Wi-Fi traffic collected from random users at public locations. The first field experiment was conducted in downtown Helsinki in late summer of 2014. A summary of this experiment is presented in Appendix A of this thesis. The outcome of this experiment was interesting per se, but more interesting was the alarming amount of seemingly private network names embedded in the collected data. After realizing the fact that devices using MAC address randomization would appear as several different devices, we started working on

tracing *random* MAC addresses back to their original entities. In order to get a holistic view of surrounding background traffic we designed the *Wireless Shark*, which we present in Paper I. While using this system to collect more data we found out that seemingly random devices were broadcasting identical long lists of previously associated private network names. This discovery led us to further investigate the information leak in Wi-Fi network discovery, which we present in Paper II. The second half of this thesis was motivated by association-free and ubiquitous Wi-Fi communication for various novel use-cases. In Paper III we propose an opportunistic notification delivery mechanism called WiPush. The proposed system provides a network-centric top-down message delivery mechanism which leverages the density of deployed Wi-Fi access points for contextual awareness. In Paper IV we propose a system for bidirectional and association-free Wi-Fi communication between mobile peers.

1.2 Problem Statement

Research questions we explore in this thesis can be divided into two areas which reflect the title of this thesis; *Privacy-Aware Opportunistic Wi-Fi*. In this section we present our research questions and the rationale behind them. The first set of questions is privacy-oriented:

RQ1: What kind of device and/or user related information is deducible from eavesdropped Wi-Fi background traffic?

RQ2: How effective are MAC address randomization techniques introduced by various manufacturers in preserving user privacy?

RQ3: How can we prevent private information from leaking through the network discovery protocol defined by the Wi-Fi standard?

The problem behind RQ1 was discovered while collecting data for mobility traces. Since Wi-Fi traffic can be listened by practically anyone, suspicious activities such as user fingerprinting can be performed without user consent. We want to find out how much personal or otherwise user identifying information is exposed through background traffic. Device manufacturers and software providers have acknowledged the issue regarding personal information leaking and tracking being done based on background. The general solution has been MAC address randomization. With RQ2 we want to find out are these mechanisms effective – but more importantly – have they fixed the problem? Both RQ1 and RQ2 focus on a

problem caused primarily by active network discovery, which is performed intermittently by Wi-Fi capable devices in a stand-by state. With RQ3 we want to raise the idea of avoiding active network discovery in its entirety.

The second set of research questions relate to opportunistic networking:

RQ4: Can we leverage the transmission range of Wi-Fi clients and use it as a location-centric addressing mechanism?

RQ5: Can we utilize the existing Wi-Fi infrastructure of restricted access points and make it useful for a broader scope of clients?

RQ6: How could experimental Wi-Fi communication systems be piloted with minimal deployment effort and overhead?

Wi-Fi has a typical transmission range from a few ten meters up to a hundred meters, or even more depending on the circumstances. While transmission range is often considered to be a restricting factor, we ask with RQ4 whether range could be used as a location-defining property for e.g. context-aware notifications. The density of deployed access points is so high that urban areas are fully covered with Wi-Fi. However, in practice it is merely a small fraction of them that are accessible or otherwise useful to an average user. With RQ5 we ask whether we can leverage the high density of access points to serve a larger audience. Many novel communication protocols and networking systems require low-level changes to user devices. On modern heterogeneous smart devices such changes can be complicated, warranty-voiding, or even impossible to implement. However, novel networking systems, such as the ones sought after in RQ4 and RQ5, require opt-in users for testing and piloting. In RQ6 we ask what would be an effortless and attractive way to engage opt-in users in such experimental systems.

1.3 Thesis Contributions

Contributions of this thesis are two-fold. The first half, i.e. Papers I and II, have a focus on background traffic that is leaking from user devices and how much of a privacy issue it is. The second half, Papers III and IV, focuses on alternative, association-free and opportunistic ways of using Wi-Fi for various novel use cases. Table 1.1 shows a mapping between Papers I through IV reprinted in this thesis and the Research Questions presented in Section 1.2.

		Research Question					
Research Paper		**1**	**2**	**3**	**4**	**5**	**6**
1.	The Wireless Shark: Identifying WiFi Devices Based on Probe Fingerprints	X	X				
2.	Quantifying the Information Leak in IEEE 802.11 Network Discovery	X	X	X			
3.	WiPush: Opportunistic Notifications over WiFi without Association				X	X	
4.	Prongle: Lightweight Communication over Unassociated Wi-Fi			X	X		X

Table 1.1: A table indicating how Papers I through IV [3, 75–77] address the Research Questions 1 through 6 presented in Section 1.2.

In Paper I [75] we present a multichannel Wi-Fi capturing system we call the *Wireless Shark*. We demonstrate its effectiveness and use it to collect background data from several devices in a controlled environment. We expose network discovery, i.e. probing behavior of these devices and classify different kinds of behavior. We also expose what a single network discovery attempt looks like when listening to all channels simultaneously. To the best of our knowledge, this is the only published research that exposes channel sweeping characteristics and differences of network discovery implementations on smart devices.

In Paper II [76] we further inspect data that can be collected with a Wi-Fi monitoring system. We classify different types of SSID names and provide a mechanism to quantify the occurring information leak. We introduce a metric, *uniqueness*, which indicates how unique an entity is in a crowd. We apply all known MAC address de-randomization techniques [51, 52, 73] to our six data sets, and show that MAC address randomization does not have a dramatic impact on the uniqueness distribution in a crowd. We also evaluate an alternative network discovery mechanism, *passive discovery*, which does not leak private information.

Paper III proposes a mobile push notification system called *WiPush* [3]. It is an opportunistic and context-aware message delivery system that operates over conventional Wi-Fi without association. The system leverages existing Wi-Fi infrastructure and has the capability of targeting user groups with a granularity level defined by the transmission range of an access point. In addition to close range notification, WiPush has the capability to forward cloud- and cell-based notification to end-users as well. We implemented WiPush on an Android smartphone and an OpenWRT-based

access point. We evaluate it in terms of energy consumption, delivery rate, latency, and impact on other network traffic.

An important lesson learned from WiPush is that implementing low-level changes on off-the-shelf hardware can be a complicated and tedious process – lucky if even possible with devices at disposal. In Paper IV we propose the Prongle system [77]. It is a lightweight communication system for various use-cases requiring opportunistic communication, such as smart traffic, delay-tolerant networks, and push notification systems, such as WiPush. Prongle devices communicate over conventional Wi-Fi hardware in an unassociated manner. The system is implemented on a separate device, and hence requires no modifications on smartphones. A Prongle device is paired over Bluetooth to an Android smartphone, from where interaction happens through an app. A Prongle device acts as a proxy between opportunistic communication and a user device. This results in an interface protecting user privacy while still being able to engage in opportunistic and novel networks.

Contributions of this thesis are covered by this manuscript as follows. Chapter 2 presents privacy-related problems originating from the current Wi-Fi network discovery protocol. These problems were originally presented and discussed in Papers I and II. Chapter 3 covers two proposals of opportunistic communication systems that are not affected by privacy problems presented in Chapter 2. These two systems were originally presented in Papers III and IV respectively.

Chapter 2

Exposing the Problem

For an average user privacy may not be of as great importance as other more visible and pragmatic features on a smartphone. An all-too-common mentality is that a privacy violation can not occur if a person has nothing to hide. This thinking boils down to the false premise of privacy being all about hiding something that is wrong or illegal [68], hence privacy is often overlooked. However, if and when a violation is revealed and demonstrated to affected subjects, privacy instantly becomes a highly appreciated quality. After the violation incident has occurred there may not be any courses of action to correct whatever harm was done. The scale and potential impact of privacy violations often exceeds common assumptions, which was witnessed in 2018 with Facebook and Cambridge Analytica [42].

We argue that demonstrating privacy-related problems to an audience is an effective wake-up-call for users to self-reflect their habits and ways of operation. In this chapter we discuss issues related to Wi-Fi background traffic and present a multichannel capturing system for more efficient traffic monitoring. We also discuss privacy problems caused by the widely used active network discovery protocol and provide a way to quantify how much it leaks personally identifiable information (PII).

2.1 Background Traffic

Since wireless transmission is a broadcast medium and Wi-Fi operates on the unlicensed ISM-band[1], all traffic is observable by any receiver within transmission range. Even if an access point (AP) uses encryption to protect data packets sent over the air, third parties are able to eavesdrop an ongoing Wi-Fi packet exchange. The IEEE 802.11 [1] standard defines three

[1]Industrial, Scientific and Medical radio band defined by the ITU Radio Regulations

categories of frames: data, control, and management frames. Data frames tend to be encrypted, but control and management frames are exchanged prior to any encryption keys, which means the intent behind these frames is visible to anyone. The primary reason for anyone to observe background traffic is to gather information about the surrounding network. This information can be used for both good and evil purposes. As an example, passive device fingerprinting [43] is often used by malicious parties in order to find specific networked devices or protocols with known vulnerabilities that can be compromised or hijacked. Other malicious activities requiring network monitoring are various denial of service attacks [14]. Channel switch and quiet attacks [45] as well as deauthentication and disassociation [20] attacks require state information, i.e. a counterfeit identity, correct timing and valid sequence numbers, in order to succeed.

Traffic monitoring can also be used for good intentions, such as detecting and reacting to aforementioned threats [5, 6, 8, 15, 32, 33, 41, 69], as well as debugging interference and other misbehavior in wireless networks [58]. Various novel proposals even use background traffic (commonly referred to as *noise*) as input signals in their system [4, 38, 66, 72, 80]. Regardless of the intentions wireless monitoring is used for, a more effective monitoring system provides a more comprehensive understanding of surrounding network activity. In this section we present a multichannel monitoring system, the Wireless Shark, originally presented in Paper I [75].

2.1.1 Methodology

Wi-Fi operates commonly on the 2.4 and 5.0 GHz radio bands. These bands are further divided into channels, which can be used to alleviate congestion caused by simultaneous transmissions. For a monitoring entity activity of interest may be ongoing on any of the channels. However, conventional Wi-Fi chips on consumer and professional-grade devices are technically limited to operate – either transmit or receive – on only one channel at a time. Some amendments[2] of the 802.11 standard support MIMO (*multiple input, multiple output*), which allows simultaneous transmission links over multiple antennas, i.e. channels, in order to achieve spatial multiplexing. Even if devices supporting MIMO are capable of receiving up to 4 simultaneous streams, that is only a fraction of the total amount of available channels. Multichannel monitoring is often implemented through channel hopping, which allocates one input stream to different channels turn by turn. This reduces dwell time per channel linearly depending on how many channels

[2] 802.11n, 802.11ac, 802.11ax

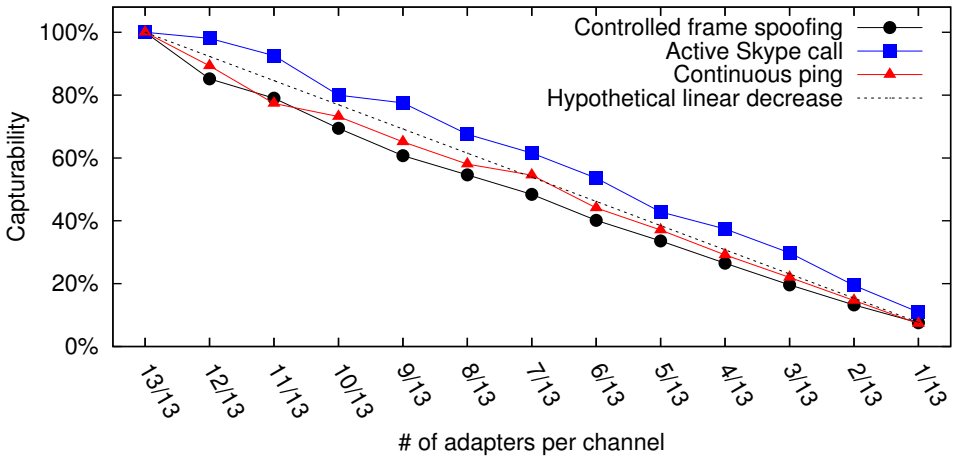


Figure 2.1: Capturability. Figure was originally presented in Paper I.

are being monitored in total. The effectiveness of capturing, i.e. capturability, can be optimized through e.g. allocating more time to channels that are more active, or reducing the amount of channels to be monitored.

Despite the amount of activity regarding wireless traffic monitoring there are few papers or literature about capturing systems themselves. Work by Meng et al. [53] explains very thoroughly how a wireless capturing tool is built. However, their work also implies channel hopping for multichannel monitoring. Various distributed monitoring systems have also been proposed [9, 55]. Our motivation for multichannel monitoring with a non-distributed single host system is to achieve microsecond time resolution between captured frames on different channels. This would then allow us to get insight on how devices perform channel sweeps when scanning for networks. We argue that true multichannel monitoring is achievable only through dedicating Wi-Fi adapters for individual channels. In Paper I we build such a system and compare it to various adapters-per-channel configurations utilizing channel hopping. Figure 2.1 shows the linear decrease of traffic captured as the amount of network adapters. Our monitoring approach has a premise to be as fundamental as possible in capturing all surrounding traffic.

2.1.2 Data Collection Considerations

User consent is a topic that must not be omitted when collecting seemingly private data. The problem with collecting data from a network is that consent can be tricky to ask since the person responsible for the data remains

unknown. There may be no other trace of the person other than the MAC address of the device. Device-specific MAC addresses on the other hand are not bound in any way to the person carrying the device, and since MAC address randomization became more common the idea of coupling a MAC address back to a person is even more challenging. Nevertheless, MAC addresses have been classified as PII. The European Data Protection Supervisor (EDPS) working party 29 (WP29) outlined in their statement 13/2011 that a MAC address combined with location information is personal data. Since we know the locations and the times our data sets were collected, we can safely say that our data shall be treated accordingly.

A MAC address is a 48-bit long identifier, which is usually represented as six octets. The first half of the identifier is the so-called *organizationally unique identifier* (OUI) governed by IEEE[3]. This part identifies a device and/or chipset manufacturer, and it is often the same throughout a range of devices of the same brand. The second half of a MAC address can be assigned by manufacturers as they wish, but ideally with respect to each address being unique. The data sets we have collected for publications reprinted in this thesis have been anonymized. In order to retain manufacturer information and whether it is a universally (UAA) or locally (LAA) administered address[4], we merely one-way hashed the latter half of each MAC address.

2.2 Privacy Problems

For the sake of clarity in terminology, let us define the meaning of three key concepts in the scope of this thesis; *privacy*, *anonymity*, and *uniqueness*:

Privacy is the capability of keeping information private. In Wi-Fi tracking context, such information typically concerns home location, workplace, affiliation, travel destinations, and so on.

Anonymity is the ability to perform tasks without revealing identity. The task may be observed by others, but it shall not reveal sensitive information. Such tasks can be e.g. a network discovery query.

Uniqueness is the concept we use to describe how much an entity stands out in a crowd. The more unique a user is, the less likely there is another one that appears and acts the same.

[3] IEEE Registration Authority

[4] UAA or LAA is indicated by the second least significant bit of the first octet.

With these terms defined, we can claim that privacy starts to deteriorate when data points from the same anonymous entity are aggregated. The situation could get even worse through exposing information about the user, which we will demonstrate a practical scenario about in Section 4.2. In this section we present two privacy problems, i.e. fingerprinting and PNL, related to Wi-Fi background traffic, and finally introduce *user uniqueness* as a metric to quantify how unique a device is in a crowd.

A prominent source of Wi-Fi background traffic is the active network discovery protocol specified by the Wi-Fi standard. Tracking is one way in which background traffic has been exploited for e.g. targeted advertising on public displays on recycling bins in London back in 2013. Harnessing a network of Wi-Fi scanners inside trashcans and collecting information regarding where a particular user is and profiling that user for advertisements was a privacy violation big enough to make the news. However, since passive monitoring can not be detected, it is hard to say whether similar systems are still active.

However, tracking users is not inherently malicious behavior. Various kinds of novel systems benefit from e.g. mobility models generated from user traces. Appendix A in this thesis explains early work [74] by the author which covers the basic concept of generating user traces based on real people movement. There are several proposed systems in this area that differ in both scale and e.g. other technologies they augment [60, 62, 65].

2.2.1 Fingerprinting

Device fingerprinting [57] has shown that privacy-preserving techniques involving pseudonyms and MAC address randomization are ineffective. Wireless driver implementations and low-level networking components of operating systems have distinct characteristics and patterns in how traffic and frames are generated. Active fingerprinting involves querying devices in a specific way and monitoring the response to those queries [17]. On the contrary, passive fingerprinting requires no interaction with a target device, which makes the process completely unobtrusive. Typically, passive techniques exploit recognizable patterns in frame headers including flags and fields used in them [54], such as information elements encapsulated in probe requests [73], or the content of *preferred networks list* (PNL), which we discuss closer in Section 2.2.2. Statistical methods have also proven to be effective, which perform device profiling based on e.g. duration values wireless devices tend to choose [18] or the timing between consecutive dispatched frames [52]. For comprehensive device fingerprinting it is desirable to use as many individualizing parameters as possible.

In this thesis we present yet another fingerprinting parameter. Since Wi-Fi networks may operate on different channels in order to avoid problems caused by RF congestion, devices look for networks on several channels. With the multichannel monitoring *Wireless Shark* monitoring system we presented in Section 2.1.1 we are able to inspect the interchannel behavior of wireless devices. Our measurements show that different devices and operating systems discover networks differently. A network discovery attempt consists of several *probe request* frames transmitted in a so-called *burst*. The amount of probe frames and the duration of one burst varies. The channel sweeping pattern and the time spent on each channel varies as well. Figure 2.2 illustrates two different network discovery attempts. Additional burst characteristics are presented in Paper I [75].

2.2.2 Preferred Networks List

When a device is initially associated to a Wi-Fi network, various information elements are stored for future associations. This so-called *Preferred Networks List* (PNL) stores wireless network identifiers, i.e. SSIDs, as well as authentication-related security details. The user may choose to deliberately *forget* a particular network, but on many devices' inclusion of a network to the PNL is the default behavior. An SSID is a cleartext handle through which networks are recognized by users and devices. In order for devices to conveniently join familiar networks the SSID and relevant authentication information must be stored on the device. Hence, the purpose of a construct like PNL is justified. However, broadcasting SSID names outside the device is not necessary[5], nor justified. Despite being unnecessary, exposing the names of previously associated networks could potentially compromise privacy.

Collecting leaked PNLs from surrounding background traffic is trivial. PNL entries, i.e. user-requested SSID names, are encapsulated as cleartext in probe requests. These frames are of management type [1], which are by design exchanged prior to any key exchange, and therefore not encrypted in any way. A generic *undirected* probe request is a broadcast question asking whether there are any networks around. On the other hand, a *directed* probe asks around for one or several specific networks. In a common case the latter is not required, since access points (AP) advertise themselves through beacon frames periodically. Despite active network discovery is not necessary, it is still widely employed. Conducted research from recent years indicates that active probing is still used [13, 25, 28, 38, 76]. The data

[5]Hidden networks require active probing, but they are strongly deprecated [64].

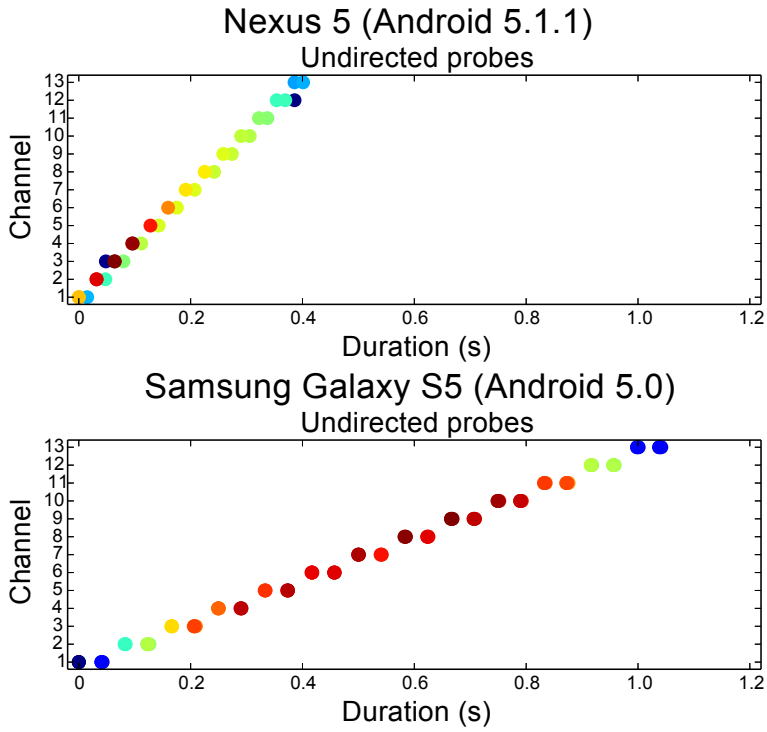


Figure 2.2: Illustration of two different network discovery attempts. On the Nexus 5 one channel sweeping burst of probe requests takes roughly 400 ms, while on the Galaxy S5 it takes over 1000 ms. The amount of frames per burst also varies.

Table 2.1: Data set described in numbers. Table was originally presented in Paper II [76].

Data set	Probe count	Directed probes	Unique MACs	Total entities	Leaked PNLs	MAC addr. randomizers
Eurosys 2017	101.1 k	41.8%	3558	2077	55.1%	608 (29.3%)
Pop concert	129.4 k	33.0%	5225	2280	28.8%	543 (23.8%)
Workers day	96.9 k	34.4%	10363	5541	25.3%	1376 (24.8%)
Movie	108.6 k	28.7%	5869	2540	29.9%	678 (26.7%)
Mall	98.4 k	33.0%	7787	5567	30.8%	1030 (18.5%)
Campus	205.5 k	43.0%	6824	2606	39.1%	652 (25.0%)

sets we collected show that on average roughly 35% of wireless entities were leaking out PNL information. Further details regarding the data sets can be found in Table 2.1 and Paper II [76].

2.2.3 User Uniqueness

Attempts of improving user privacy in Wi-Fi has been seen in the past. Disposable MAC addresses [31, 67], through which wireless devices can act as "random" entities, has been proposed to eliminate traceability. It has, however, been shown that using this so-called MAC address randomization is not sufficient to eliminate tracking [51, 52]. Several studies have shown that hiding behind pseudonyms is not enough because there are many other parameters that can be used for identifying, i.e. fingerprinting, Wi-Fi clients [24, 25, 57]. The key idea behind using random pseudonyms is to have an alternative identity that seemingly blends into a crowd. A pseudonym should also be disposable, since if one gets compromised it is easy to introduce a new one. Conceptually this can be categorized as MAC address spoofing, which to many networking oriented people has a malicious connotation.

Even if an entity manages to conceal the true identity behind disposable identifiers, actions and behavior can reveal the identity behind several identifiers. One way to connect fake identifiers is through device fingerprinting [57]. Another way for anonymity chasing entities to reveal their identity is through exposing parts of their preferred networks list (PNL). Unarguably the best way to stay unnoticed and untraceable through Wi-Fi is to not transmit anything. However, since users tend to leave Wi-Fi enabled and devices have an urge to get connected, there often is background traffic that allows e.g. tracking. The second best way to stay anonymous is to not transmit anything that can be connected to earlier appearances, or

Figure 2.3: Distribution of SSID significance values. Popular SSIDs have high significance values. The heavy tail indicates that most witnessed SSIDs are unique. Figure was originally presented in Paper II [76].

that is otherwise identifiable. According to our collected data (Table 2.1) on average 35% of devices transmit PNL information, which compromises anonymity. In Paper II [76] we present a metric to quantify how unique a single user is in a crowd. We use *uniqueness* to describe how well a wireless entity stands out, i.e. how unique it is, in a crowd based on the background traffic we can passively collect. In order to calculate uniqueness we first need background data with PNL information. We then define uniqueness as follows.

Let entity e have a PNL with k distinct SSID names (2.1) and rank of n be the number of entities that have network n in their PNL (2.2):

$$PNL_e = \{n_1, n_2, ..., n_k\} \tag{2.1}$$

$$rank_{n_i} = |n_i| \tag{2.2}$$

First we calculate a significance value S for each SSID in e's PNL:

$$S_i = min\left\{\frac{|n_i|^{1+\frac{1}{k}}}{T}, 1\right\},$$

The significance of a single SSID depends on how common that SSID in the context it appears in. As a practical example, an SSID related to a mobile network operator is common in the area where that MNO operates, but can be unique in another country. Figure 2.3 shows the distribution of significance values in the data sets we collected. A low significance value contributes more to the uniqueness of an entity. The heavy tail of the

distribution indicates that most SSIDs make users broadcasting them more unique. Further details and SSID classification can be found in Paper II.

Finally, we calculate the uniqueness value for a given entity e with the following formula:

$$uniqueness_e = 1 - \Big(S_1 \cdot S_2 \cdot ... \cdot S_k\Big).$$

Uniqueness values are normalized values between 0 and 1. A high uniqueness value indicates how well a user stands out from a crowd by looking at the PNL content that is exposed. Anonymous users have a uniqueness value of 0 by definition.

2.3 Active vs. Passive Network Discovery

What could we do to correct the privacy threats introduced in this section? One effective way to reduce the amount of background traffic is to use passive instead of active network discovery. In Paper II [76] we compared active versus passive network discovery, and based on the evaluation we can conclude that in most cases the extra time it takes for passive discovery to find a network is negligible. With typical beaconing intervals around 100 ms from the AP the discovery time is 0.6 seconds longer. Figure 2.4 shows a comparison of the two. Another motivation to reduce background traffic is for the common good. It has been shown that aggressive network discovery deteriorates throughput and increases energy consumption [39]. Techniques to detect the causes of unnecessary network scanning have been proposed [29], which could help firmware developers create more sophisticated network discovery strategies. Some manufacturers have introduced devices with location-aware active network discoveries.

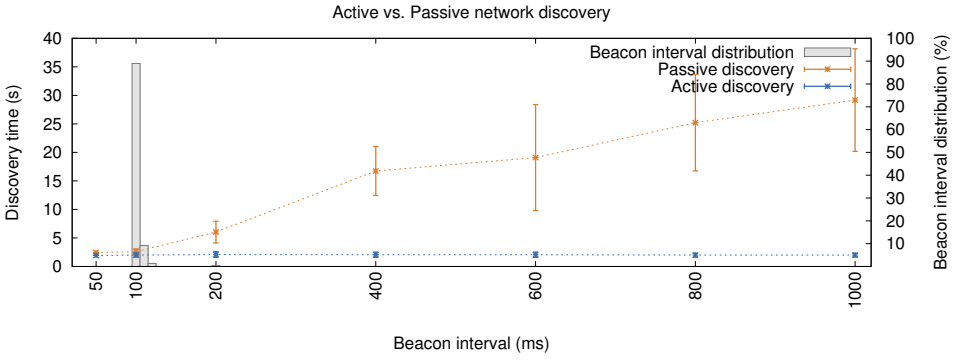


Figure 2.4: Active vs. passive scanning.

2.4 Summary

In this section we pointed out problems caused by background Wi-Fi traffic primary belonging to active network discoveries. We implemented a multichannel Wi-Fi monitoring system, and demonstrated yet another way to fingerprint devices based on distinct channel sweeping patterns employed by different devices during network discovery. We used the monitoring system to collect data sets which contain potentially sensitive information regarding networks a user device has associated to in the past. We introduced a metric to quantify how unique a user is in a crowd if a list of previously associated network names is exposed. We also compared active and passive network discovery protocols, and argued that in the vast majority of cases the increase in discovery time is negligible.

Chapter 3

Opportunistic Wi-Fi

All the privacy threatening phenomenons presented in this thesis are related to network discovery, and the habit of carelessly associating to any free Wi-Fi. These are widely recognized problems, but the strong need for Internet connectivity often drives users to take risks [56]. Protocols like Hotspot 2.0 have been proposed [78] to alleviate these risks and the inconvenience of typing in login credentials and passphrases while joining a Wi-Fi. In 2012 Cisco listed [21] "login process" and "hotspot selection" as *user frustrating usability problems* with public hotspots back then. Eight years later we can safely say that these usability problems are still around to frustrate users.

Because of the constantly increasing amount of mobile users and rapid growth of data being consumed by them [2], the so-called *mobile data explosion* puts a lot of pressure on networking technologies. While mobile network operators (MNO) struggle to meet the ever-increasing demand of data, offloading technologies using alternative transmission links have gained interest [7, 48, 63]. According to Cisco [2] up to 59% of mobile data will be offloaded over Wi-Fi by 2022. How MNOs and networking equipment and device manufacturers will achieve this remains to be seen. The idea of a metropolitan-scale free and open Wi-Fi is what many cities would surely like to offer, but eventual gains would not cover deployment and maintenance costs. Especially since Internet connectivity can be monetized by MNOs. The economic viability of providing public Wi-Fi connectivity was questioned already back in 2002 [36]. The aforementioned Hotspot 2.0 has been proposed as an enabling technique for handling associations to offloading networks automatically [81]. As of today, Hotspot 2.0 is a subscription service that operates through roaming, which has an impact on e.g. handover performance due to the overhead introduced by ANQP and credential checking [47].

Opportunistic networks have been proposed as alternative transmission links [35, 40, 59] for mobile data offloading. Many proposals exploit human mobility and social behavior in order to improve communication in various ways [16, 34, 61]. One big obstacle for opportunistic networks is how to establish communication links between endpoints. Several proposals rely solely on Wi-Fi in different configurations, including Wi-Fi Direct [22, 30], ad hoc [49], and infrastructure [26, 70].

Another novel idea for accessing offloading capabilities is through Wi-Fi without association. In such a scenario any available Wi-Fi could satisfy the need for communication with no authentication and association required. As a remark, it is crucial to note that "association-free Wi-Fi" is not the same as "free Wi-Fi", which has been mentioned earlier in this thesis. This so-called ubiquitous Wi-Fi was visualized as early as in 2002 [36] when wireless networking started to become a widespread commodity. It has since persisted as a research vision, but in practice repeatedly outmaneuvered by developments in cellular data [44]. The high density of access points at metropolitan areas has coverage for a city-wide offloading Wi-Fi, but the vast majority of networks require authentication, which renders them useless for an average user. Other open questions regarding ubiquitous Wi-Fi are e.g. who provides the service, and whether networks can be trusted. Security-wise it is a positive and current trend that security is migrating more and more to the application layer.

Implementations for association-free Wi-Fi exist [79], but deploying such typically require low-level changes to software on devices, which in turn effectively discourages potential user bases to form. In this section we present two systems representing opportunistic and association-free communication over Wi-Fi.

3.1 Push Notifications over Wi-Fi

Push notifications are small messages delivered from cloud services to user devices intended to notify the user of e.g. an incoming message or another event. Major mobile operating systems run their own notification services; Google Cloud Messaging (GCM) and Apple Push Notification (APN). Such services enable third-party app developers to push notification messages to app users. The notification service – knowing how to reach the user – will then take care of delivering the notification through some available data transport channel.

In Paper III we propose a system called WiPush. The system is an opportunistic notification delivery system which leverages the dense de-

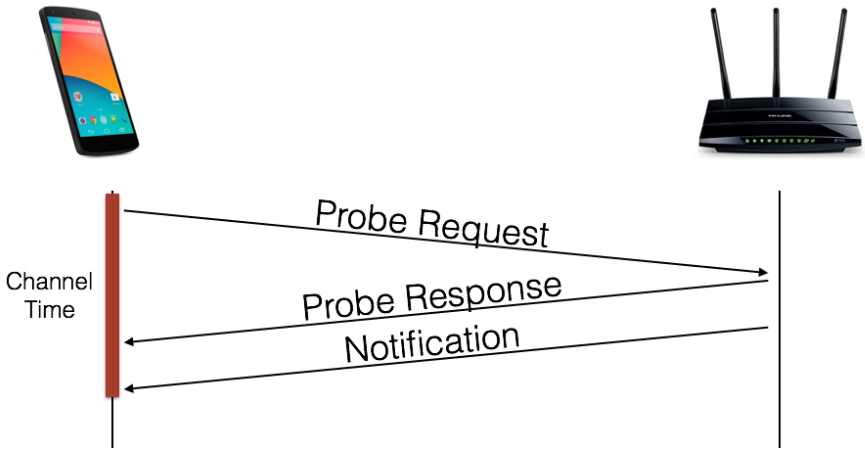


Figure 3.1: WiPush delivery mechanism.

ployment of Wi-Fi access points (AP). WiPush is a best-effort messaging layer which operates over Wi-Fi without association. The transmission range of APs provide an intrinsic spatio-temporal addressing mechanism for the system. Contextual notifications, such as information regarding surrounding services, can thus be disseminated from specific APs instead of first resolving and then addressing all relevant clients within an area. Hereby any services initiating a notification delivery do not need to know the locations of target users.

Since WiPush is opportunistic and association-free, we exploit incoming network discovery protocol queries from the client-side to deliver messages when a device is listening. When a device dispatches probe requests in order to discover networks, it has to wait for a brief moment after each query for incoming probe responses. After listening for a specified time the device then switches channel and transmits probe requests on that channel. This channel sweeping behavior during network discovery is illustrated in Figure 2.2. WiPush leverages this so-called *channel time* window, and delivers the notification to a device during it. Figure 3.1 illustrates the delivery mechanism.

WiPush was designed with three design challenges; **DC1:** Compliance with the existing Wi-Fi specification. Since WiPush uses public action frames to deliver notifications, it does not conflict or violate the Wi-Fi standard in any way. Contextual notification protocols similar to WiPush have been proposed, but often proximity in them is complemented by some other technology, such as Bluetooth [46,71]. Entirely Wi-Fi based solutions exist, but e.g. Beacon stuffing [19] can be considered to abuse the standard.

DC2: Directed notification messages. An essential property for push notifications is the ability to target them to specific users. WiPush uses MAC addresses exposed by probe requests to address individual devices. Probe responses and notification encapsulating action frames are sent to the same recipient successively. How an AP is able to validate a user and prevent hijacking of push notifications through MAC address spoofing was left for future work.

DC3: Minimal energy expenditure. Battery life is an important and highly valued asset on modern smart devices. Hence, we wanted to minimize energy expenditure. WiPush exploits the channel time listening window initiated by network discovery. This way WiPush does not cause extra channel switching, frame transmissions, or other hardware activity on the client-side in order to operate.

WiPush can ideally be implemented on existing commodity hardware, which reduces deployment costs. Our pilot deployment of the system was implemented on an OpenWRT based access point and a Google Nexus 5 android-based smartphone. A system description, implementation details, and system evaluation regarding performance and energy expenditure can be found in Paper III [3].

3.2 Novel Applications

Many novel communication protocols require low-level changes to wireless drivers or operating system components [79]. With ordinary consumer devices such modifications can be complicated to carry out. Many manufacturers make it deliberately hard or practically impossible to implement modifications. This does not help with piloting experimental systems and attracting new users. In Paper IV we propose a system for lightweight communication over unassociated Wi-Fi. We labeled it the Prongle system. The system uses so-called *prongle devices* to create a communication layer. Prongle devices are personal companion devices that act as gateways to various kinds of novel and opportunistic networks. A separate communication device provides more flexibility and control in using Wi-Fi to communicate. The Prongle system also provides a privacy-protecting interface between personal devices and public activity. This interface is illustrated in Figure 3.2. Since all communication goes through a prongle device, only the prongle is visible to the public, allowing end-user devices to remain in the background. Communication gateways are known as proxies, and the device itself has the form of a dongle. Hence the name *Prongle*.

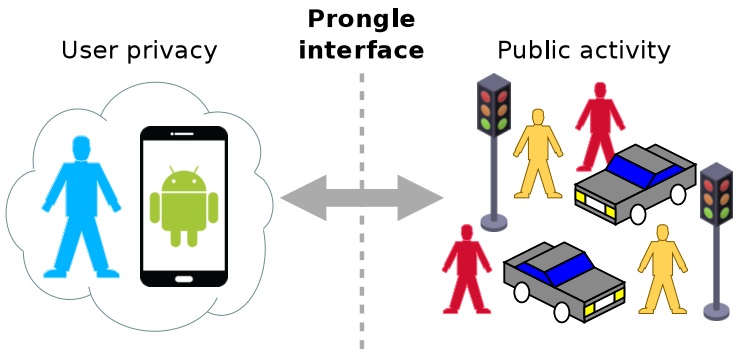


Figure 3.2: Prongle system creates an interface between user privacy and public activity.

The Prongle system communicates on top of a layer of prongle devices, which in turn communicate with each other in an unassociated and opportunistic way over conventional Wi-Fi hardware. End-users interact with the system through smart devices, such as smartphones. Each smartphone is paired to a prongle device over Bluetooth, and all communication to the Prongle systems goes via the prongle device. An illustration of the communication path can be seen in Figure 3.3. From the smartphone point-of-view, accessing opportunistic networks through a Bluetooth accessory device leaves other Internet connection links, i.e. cellular and Wi-Fi, untouched on the device. Since opportunistic networks may be able to provide only delay-tolerant communication, it is justified to reserve cellular data and integrated Wi-Fi capabilities on smartphones for real-time connections. This separation of opportunistic communication to an external device also implies that no modifications are required on user devices, which makes piloting novel systems easier as users can use any device they prefer. One of the key design principles was to have a system which is effortless for new users to opt-in.

We propose four use-cases for our Prongle system; **Smart traffic.** In the current state-of-art pedestrian and cyclist detection relies solely on sensors on vehicles and object detection through on-board cameras. Vehicles with smart electronics can utilize digital communication and protocols like vehicle-to-vehicle (V2V) to announce their presence in a traffic scenario. We propose that our Prongle system could be used for communication between light-traffic users and vehicles. A prongle device would announce its presence by periodically transmitting beacons, which could then be noted by other surrounding smart traffic users. We like to think of this as a wireless reflector — without the need for line-of-sight to be spotted.

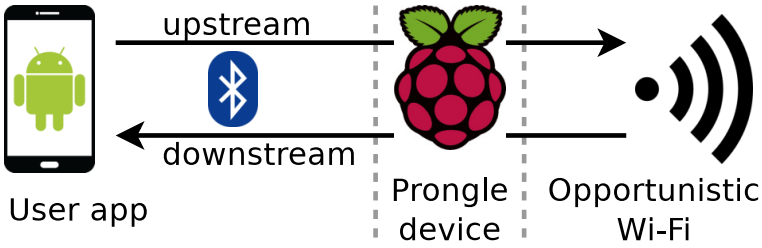


Figure 3.3: Illustration of the communication path between an Android-based user device and opportunistic Wi-Fi through a Prongle device.

Push notifications. Similarly to WiPush we presented in Paper III and Section 3.1, the Prongle system can be used for contextual opportunistic push messages. An important lesson learned while developing WiPush was that system piloting and deployment should be made as effortless as possible. With the Prongle system opportunistic and unassociated Wi-Fi communication requires no modifications, rooting or implementing changes on a heterogeneous set of opt-in users devices.

Audience response systems. Various public events are augmented by including responses from the audience. One way to achieve this is to provide a hotspot through which the audience can access inputs of the response system. Participation in such situations tends to be somewhat reluctant. With Prongle system communication users would not have to associate with the hotspot provided, and anonymity could be preserved.

Delay-tolerant networks. Ad hoc communication has a key role in enabling opportunistic and delay-tolerant networks (DTN). Establishing communication links and routing in a *mobile ad hoc network* [23] is a widely researched topic. Our Prongle system provides a flexible platform to implement opportunistic and DTN strategies on top of.

A prongle device consists of a Raspberry Pi single-board computer and a battery pack to power it. An Android app is used to interact with the prongle device. Implementation for both the prongle device[1] and the app[2] are publicly available. Details regarding the implementation and performance evaluation can be found in Paper IV [77].

[1]https://github.com/owaltari/btprongle_server
[2]https://github.com/owaltari/btprongle_app

3.3 Summary

One way to alleviate the issues caused by background traffic is to reduce the need to resort to incidental free Wi-Fis. With the so-called mobile data explosion around the corner, this can be a tricky task. Opportunistic networks have emerged as a complementing communication paradigm for data offloading. Such networks typically operate on layers established on mobile devices, which intrinsically distributes cost to all participating users. In this section we presented two systems (Paper III [3], Paper IV [77]) that employ opportunistic communication and leverage existing Wi-Fi hardware in order to be cost-effective, as well as effortless to adopt.

Chapter 4

Discussion

In this chapter we revisit the research questions presented in Section 1.2. We also present public attention our work has been exposed to. Finally, we conclude this thesis with some final remarks.

4.1 Research Questions Revisited

RQ1: What kind of device and/or user related information is deducible from eavesdropped Wi-Fi background traffic?

Device fingerprinting is used to profile devices in a crowd. In Paper I we present a multichannel monitoring system which is able to inspect the channel sweeping pattern different devices use when querying for networks. This information can be used as yet another parameter to individualize devices. Fingerprinting can be used to trace disposable MAC addresses back to the original device. After collecting plenty of background traffic and applying mechanisms presented in Paper II to the data, user profiles can be deduced. Section 4.2 presents a practical scenario demonstrating a user profile pulled from background traffic.

RQ2: How effective are MAC address randomization techniques introduced by various manufacturers in preserving user privacy?

Earlier research has shown that MAC address randomization techniques are not sufficient because of various reasons. The first one relates to the poor implementation of the randomizing technique itself. Secondly, even if devices use pseudonyms instead of their physical MAC address, fingerprinting provides a way for a monitoring party to connect seemingly random devices to the same entity. Paper II provides a metric to quantify how unique a particular device is in a crowd.

After analyzing the uniqueness distribution of users in six different data set both before and after applying MAC address randomization reversing techniques, we argue that the effects of address randomization are not significant. Even if MAC address randomization was implemented properly, mistakes like exposing a PNL deteriorates user anonymity.

RQ3: How can we prevent private information from leaking through the network discovery protocol defined by the Wi-Fi standard?

A prominent cause for private information leaking are the directed probes employed by active network discovery. An alternative discovery mechanism, passive network discovery, does not expose names of previously associated networks. In Paper II we evaluate performance implications between active and passive network discovery.

RQ4: Can we leverage the transmission range of Wi-Fi clients and use it as a location-centric addressing mechanism?

In Papers III and IV we present two different opportunistic communication systems that leverage the transmission range of Wi-Fi. WiPush [3] uses transmission range as an intrinsic addressing mechanism to deliver spatio-temporal push messages. Use-cases of the Prongle system [77] imply communication with nearby nodes within a typical range of Wi-Fi hardware. Our evaluation shows that we are able to get a 95% transmission success rate at a 50-meter distance.

RQ5: Can we utilize the existing Wi-Fi infrastructure of restricted access points and make it useful for a broader scope of clients?

In Paper III we introduce an opportunistic push notification system the leverages the high density of access points in metropolitan areas. The system can ideally be deployed on consumer-grade hardware, which could reduce deployment costs. The system we propose exploits active network discovery and operates coordinated with it in order to minimize energy expenditure.

RQ6: How could experimental Wi-Fi communication systems be piloted with minimal deployment effort and overhead?

Novel networking systems often require low-level modifications on participating devices. In Paper IV we propose the Prongle system, which introduces a companion device that provides a more flexible and controllable platform to develop novel communication systems

on top of. Interaction with the Prongle system happens through a smartphone app, which involves no changes to opt-in users' devices.

4.2 Publicity and Impact

As stated earlier in Chapter 2, privacy may not be of great importance to an average user — until it gets violated. As long as something as ordinary as Wi-Fi delivers the promised connectivity, users tend to easily think that everything is in order regarding the protocol. Breaking this illusion can be ultimately hard, especially since vendors seem to have overlooked the privacy issues discussed in this thesis. An effective wake-up call to anyone is to witness a violation personally. In the spring of 2016 a perfect opportunity occurred for us to demonstrate early work presented in this thesis at the 9th Science Slam[1] held in Helsinki. It is a science popularizing event which welcomes researchers from all fields to present interesting things in an entertaining style. The presentation we held was called "Turn Off Your WiFi!" and it was given by Professor Jussi Kangasharju.

During the presentation a setup of the Wireless Shark (Paper I, [75]) was collecting PNLs from the audience. The presentation explained in popular terms how network discoveries reveal information about other networks stored on the device. The presentation also demonstrated SSID pinpointing, which can be done with the help of external crowd-sourced services like WiGLE[2]. The presentation demonstrated an example seen in Figure 4.1. Red circles in the figure show the information WiGLE has from an SSID called "honeypot" in downtown Helsinki, and the red X indicates the location we know this access point used to be.

Additionally, the presentation demonstrated a brief analysis of a previously witnessed user whose PNL had been exposed. From this users PNL we publicly deduced the following:

- Device manufacturer (Apple, based on OUI)
- Three non-public networks at company Z (Workplace?)
- Restaurant nearby the campus (Lunchplace?)
- Three airports (A, B, and C, where B is a hub between A and C)
- Airline in-flight Wi-Fi (Airline flies from A to C)
- Conference Wi-Fi (Conference site less than 50km from C)
- Neighboring state university Wi-Fi
- Helsinki and Aalto universities (Academic affiliations?)

[1] https://scienceslam.fi/
[2] https://wigle.net/

Figure 4.1: Circles show the information WiGLE has from a SSID called "honeypot". The red X indicates the actual location of this access point.

Even if we do not have a name for the person in question, this list tells a story about someone. Neither can we tell the time when the user was visiting these networks. However, this information combined with some other information sources (conference proceedings, company Z and university staff, etc.) could get us on track regarding a real identity. Gladly this person's device did not reveal a unique home SSID, since that typically narrows the search down significantly. Our presentation ended with displaying all the PNL entries that had been collected during the evening from the crowd. Nothing was deduced from that data, and it was only to demonstrate how such information can be collected without consent. The winner at Science Slam events is voted by the audience. Our presentation apparently made a point because it won first place.

The impact of such public awareness events remains to be seen. Making an actual change is eventually up to the developers of networking components and drivers — and not to forget vendor initiative. I sincerely hope that related publications get as much visibility as possible at networking conferences and task forces, since those are the most prominent venues for constructive discussion to emerge regarding the current situation.

4.3 Conclusion

In this thesis we built a multichannel monitoring system and used it to collect and investigate background traffic that is generated by Wi-Fi capable smart devices. We found out yet another way to fingerprint devices and make anonymization attempts through MAC address randomization even less effective. We introduced a metric to quantify the uniqueness of a wireless device, and showed that the effects of MAC address randomization are not that significant. We exposed the problem with active network discovery and that exposing the so-called *preferred networks list* (PNL) through directed probe requests can tell a whole lot about the user possessing the device. The habit of resorting to any incidental "Free Wi-Fi" escalates the problem. On the other hand, the rapid increase in cellular data and price plans encourages users to offload traffic through alternative networks.

Opportunistic networks have been proposed in various flavors to serve as alternative ways to move data and complement the struggling cellular networks. Such networks typically operate on layers established on mobile devices, which intrinsically distributes cost to all participating users. In this thesis we presented two opportunistic networking systems that operate over conventional Wi-Fi hardware. WiPush is an opportunistic push notification system which leverages the existing Wi-Fi infrastructure. The system provides spatio-temporal push notifications over Wi-Fi without association. Finally, we presented a system for opportunistic, lightweight communication over unassociated Wi-Fi for various use cases, such as smart traffic and delay-tolerant networks.

The contributions of this thesis were two-fold. For the latter part, only time will tell what kind of communication paradigms will cope in our ever-evolving connected environment. As for the privacy-related contributions, the most important take-away is:

Turn off your Wi-Fi — unless you're actually using it.

References

[1] IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)*, 2016.

[2] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 White Paper, 2017.

[3] U. G. Acer and O. Waltari. WiPush: Opportunistic Notifications over WiFi Without Association. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MobiQuitous 2017, pages 353–362, New York, NY, USA, 2017. ACM.

[4] V. Acuna, A. Kumbhar, E. Vattapparamban, F. Rajabli, and I. Guvenc. Localization of WiFi Devices Using Probe Requests Captured at Unmanned Aerial Vehicles. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, March 2017.

[5] M. Agarwal, S. Biswas, and S. Nandi. Detection of de-authentication denial of service attack in 802.11 networks. In *2013 Annual IEEE India Conference (INDICON)*, pages 1–6, Dec 2013.

[6] M. Agarwal, S. Biswas, and S. Nandi. An Efficient Scheme to Detect Evil Twin Rogue Access Point Attack in 802.11 Wi-Fi Networks. *International Journal of Wireless Information Networks*, 25(2):130–145, Jun 2018.

[7] A. Aijaz, H. Aghvami, and M. Amani. A survey on mobile data offloading: technical and business perspectives. *IEEE Wireless Communications*, 20(2):104–112, April 2013.

[8] A. Arora. Preventing wireless deauthentication attacks over 802.11 networks. *CoRR*, abs/1901.07301, 2019.

[9] P. Arora, N. Xia, and R. Zheng. A Gibbs Sampler Approach for Optimal Distributed Monitoring of Multi-Channel Wireless Networks. In *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, pages 1–6, Dec 2011.

[10] N. Asokan, V. Niemi, and K. Nyberg. Man-in-the-middle in tunnelled authentication protocols. In B. Christianson, B. Crispo, J. A. Malcolm, and M. Roe, editors, *Security Protocols*, pages 28–41, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[11] J. Atkinson, M. Rio, J. Mitchell, and G. Matich. Your WiFi Is Leaking: Ignoring Encryption, Using Histograms to Remotely Detect Skype Traffic. In *Military Communications Conference (MILCOM), 2014 IEEE*, pages 40–45, Oct 2014.

[12] J. S. Atkinson, J. E. Mitchell, M. Rio, and G. Matich. Your WiFi is leaking: What do your mobile apps gossip about you? *Future Generation Computer Systems*, 80:546 – 557, 2018.

[13] 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*, IMC '13, pages 265–276, New York, NY, USA, 2013. ACM.

[14] J. Bellardo and S. Savage. 802.11 denial-of-service attacks: Real vulnerabilities and practical solutions. In *USENIX security symposium*, volume 12, pages 2–2. Washington DC, 2003.

[15] R. Beyah and A. Venkataraman. Rogue-Access-Point Detection: Challenges, Solutions, and Future Directions. *IEEE Security Privacy*, 9(5):56–61, Sep. 2011.

[16] C. Boldrini, M. Conti, and A. Passarella. Exploiting users' social relations to forward data in opportunistic networks: The hibop solution. *Pervasive and Mobile Computing*, 4(5):633–657, 2008.

[17] S. Bratus, C. Cornelius, D. Kotz, and D. Peebles. Active behavioral fingerprinting of wireless devices. In *Proceedings of the First ACM Conference on Wireless Network Security*, WiSec '08, pages 56–61, New York, NY, USA, 2008. Association for Computing Machinery.

[18] J. Cache. Fingerprinting 802.11 implementations via statistical analysis of the duration field. *Uninformed.org*, 5, 2006.

[19] R. Chandra, J. Padhye, L. Ravindranath, and A. Wolman. Beacon-Stuffing: Wi-Fi without Associations. In *Eighth IEEE Workshop on Mobile Computing Systems and Applications*, pages 53–57, March 2007.

[20] R. Cheema, D. Bansal, and S. Sofat. Deauthentication/disassociation attack: Implementation and security in wireless mesh networks. *International Journal of Computer Applications*, 23(7):7–15.

[21] Cisco. The Future of Hotspots: Making Wi-Fi as Secure and Easy to Use as Cellular, 2012.

[22] M. Conti, F. Delmastro, G. Minutiello, and R. Paris. Experimenting opportunistic networks with wifi direct. In *2013 IFIP Wireless Days (WD)*, pages 1–6. IEEE, 2013.

[23] S. Corson and J. Macker. RFC2501: Mobile Ad Hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations, 1999.

[24] M. Cunche, M. Kaafar, and R. Boreli. I know who you will meet this evening! Linking wireless devices using Wi-Fi probe requests. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012 IEEE International Symposium on a*, pages 1–9, June 2012.

[25] A. Di Luzio, A. Mei, and J. Stefa. Mind your probes: De-anonymization of large crowds through smartphone WiFi probe requests. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[26] D. J. Dubois, Y. Bando, K. Watanabe, and H. Holtzman. Lightweight Self-organizing Reconfiguration of Opportunistic Infrastructure-mode WiFi Networks. In *2013 IEEE 7th International Conference on Self-Adaptive and Self-Organizing Systems*, pages 247–256, Sep. 2013.

[27] K. Fall. A Delay-tolerant Network Architecture for Challenged Internets. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '03, pages 27–34, New York, NY, USA, 2003. ACM.

[28] J. Freudiger. How Talkative is Your Mobile Device?: An Experimental Study of Wi-Fi Probe Requests. In *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, WiSec '15, pages 8:1–8:6, New York, NY, USA, 2015. ACM.

[29] H. Fulara, G. Singh, D. Jaisinghani, M. Maity, T. Chakraborty, and V. Naik. Use of Machine Learning to Detect Causes of Unnecessary Active Scanning in WiFi Networks. In *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pages 1–9, June 2019.

[30] A. Garcia-Saavedra and P. Serrano. Device-to-device communications with wifi direct: Overview and experimentation. *IEEE Wireless Communications*, page 97, 2013.

[31] M. Gruteser and D. Grunwald. Enhancing location privacy in wireless lan through disposable interface identifiers: A quantitative analysis. In *Proceedings of the 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, WMASH '03, pages 46–55, New York, NY, USA, 2003. ACM.

[32] F. Guo and T.-c. Chiueh. Sequence Number-based MAC Address Spoof Detection. In *Proceedings of the 8th International Conference on Recent Advances in Intrusion Detection*, RAID'05, pages 309–329, Berlin, Heidelberg, 2006. Springer-Verlag.

[33] M. Hafiz and F. Ali. Profiling and mitigating brute force attack in home wireless LAN. In *Computational Science and Technology (ICCST), 2014 International Conference on*, pages 1–6, Aug 2014.

[34] B. Han, P. Hui, V. S. A. Kumar, M. V. Marathe, J. Shao, and A. Srinivasan. Mobile Data Offloading through Opportunistic Communications and Social Participation. *IEEE Transactions on Mobile Computing*, 11(5):821–834, May 2012.

[35] B. Han, P. Hui, and A. Srinivasan. Mobile Data Offloading in Metropolitan Area Networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 14(4):28–30, Nov. 2010.

[36] P. S. Henry and Hui Luo. WiFi: what's next? *IEEE Communications Magazine*, 40(12):66–72, Dec 2002.

[37] G. R. Hiertz, D. Denteneer, S. Max, R. Taori, J. Cardona, L. Berlemann, and B. Walke. IEEE 802.11s: The WLAN Mesh Standard. *IEEE Wireless Communications*, 17(1):104–111, February 2010.

[38] H. Hong, G. D. De Silva, and M. C. Chan. CrowdProbe: Non-Invasive Crowd Monitoring with Wi-Fi Probe. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3), Sept. 2018.

[39] X. Hu, L. Song, D. Van Bruggen, and A. Striegel. Is There WiFi Yet? How Aggressive Probe Requests Deteriorate Energy and Throughput. In *Proceedings of the 2015 Internet Measurement Conference*, IMC '15, pages 317–323, New York, NY, USA, 2015. Association for Computing Machinery.

[40] C.-M. Huang, K.-c. Lan, and C.-Z. Tsai. A survey of opportunistic networks. In *22nd International Conference on Advanced Information Networking and Applications-Workshops (aina workshops 2008)*, pages 1672–1677. IEEE, 2008.

[41] N. Husted and S. Myers. Mobile location tracking in metro areas: Malnets and others. In *Proceedings of the 17th ACM Conference on Computer and Communications Security*, CCS '10, pages 85–96, New York, NY, USA, 2010. ACM.

[42] J. Isaak and M. J. Hanna. User data privacy: Facebook, cambridge analytica, and privacy protection. *Computer*, 51(8):56–59, August 2018.

[43] Ke Gao, C. Corbett, and R. Beyah. A passive approach to wireless device fingerprinting. In *2010 IEEE/IFIP International Conference on Dependable Systems Networks (DSN)*, pages 383–392, June 2010.

[44] J. A. Kilpatrick, R. J. Cyr, E. L. Org, and G. Dawe. New sdr architecture enables ubiquitous data connectivity. *RF DESIGN*, 29(1):32, 2006.

[45] B. Könings, F. Schaub, F. Kargl, and S. Dietzel. Channel switch and quiet attack: New dos attacks exploiting the 802.11 standard. In *2009 IEEE 34th Conference on Local Computer Networks*, pages 14–21, Oct 2009.

[46] M. Koühne and J. Sieck. Location-based services with ibeacon technology. In *2014 2nd International Conference on Artificial Intelligence, Modelling and Simulation*, pages 315–321, 2014.

[47] V. Lavrukhin. An overhead analysis of access network query protocol (anqp) in hotspot 2.0 wi-fi networks. In *2013 13th International Conference on ITS Telecommunications (ITST)*, pages 266–271, Nov 2013.

[48] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong. Mobile Data Offloading: How Much Can WiFi Deliver? In *Proceedings of the 6th International COnference*, Co-NEXT '10, pages 26:1–26:12, New York, NY, USA, 2010. ACM.

[49] S. Lu, S. Shere, Y. Liu, and Y. Liu. Device discovery and connection establishment approach using ad-hoc wi-fi for opportunistic networks. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 461–466. IEEE, 2011.

[50] J. Martin, T. Mayberry, C. Donahue, L. Foppe, L. Brown, C. Riggins, E. C. Rye, and D. Brown. A study of mac address randomization in mobile devices and when it fails. *arXiv preprint arXiv:1703.02874*, 2017.

[51] J. Martin, T. Mayberry, C. Donahue, L. Foppe, L. Brown, C. Riggins, E. C. Rye, and D. Brown. A study of MAC address randomization in mobile devices and when it fails. *CoRR*, abs/1703.02874, 2017.

[52] C. Matte, M. Cunche, F. Rousseau, and M. Vanhoef. Defeating MAC Address Randomization Through Timing Attacks. In *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, WiSec '16, pages 15–20, New York, NY, USA, 2016. ACM.

[53] K. Meng, Y. Xiao, and S. V. Vrbsky. Building a wireless capturing tool for WiFi. *Security and Communication Networks*, 2(6):654–668, 2009.

[54] C. Neumann, O. Heen, and S. Onno. An empirical study of passive 802.11 device fingerprinting. In *2012 32nd International Conference on Distributed Computing Systems Workshops*, pages 593–602, June 2012.

[55] H. Nguyen, G. Scalosub, and R. Zheng. On quality of monitoring for multichannel wireless infrastructure networks. *IEEE Transactions on Mobile Computing*, 13(3):664–677, March 2014.

[56] Norton by Symantec. Norton Wi-Fi Risk Report, 2017.

[57] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall. 802.11 user fingerprinting. In *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking*, MobiCom '07, pages 99–110, New York, NY, USA, 2007. ACM.

[58] U. Paul, A. Kashyap, R. Maheshwari, and S. R. Das. Passive measurement of interference in wifi networks with application in misbehavior detection. *IEEE Transactions on Mobile Computing*, 12(3):434–446, March 2013.

[59] L. Pelusi, A. Passarella, and M. Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *IEEE communications Magazine*, 44(11):134–141, 2006.

[60] A.-C. Petre, C. Chilipirea, M. Baratchi, C. Dobre, and M. van Steen. Chapter 14 - WiFi Tracking of Pedestrian Behavior. In F. Xhafa, F.-Y. Leu, and L.-L. Hung, editors, *Smart Sensors Networks*, Intelligent Data-Centric Systems, pages 309 – 337. Academic Press, 2017.

[61] A.-K. Pietilänen and C. Diot. Dissemination in opportunistic social networks: the role of temporal communities. In *Proceedings of the thirteenth ACM international symposium on Mobile Ad Hoc Networking and Computing*, pages 165–174. ACM, 2012.

[62] G. Pipelidis, N. Tsiamitros, M. Kessner, and C. Prehofer. HuMAn: Human Movement Analytics via WiFi Probes. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 370–372, March 2019.

[63] M. H. Qutqut, F. M. Al-Turjman, and H. S. Hassanein. MFW: Mobile femtocells utilizing WiFi: A data offloading framework for cellular networks using mobile femtocells. In *2013 IEEE International Conference on Communications (ICC)*, pages 6427–6431, June 2013.

[64] S. Riley. Myth vs. reality: Wireless SSIDs, October 2007.

[65] P. Sapiezynski, A. Stopczynski, R. Gatej, and S. Lehmann. Tracking Human Mobility Using WiFi Signals. *PLOS ONE*, 10(7):1–11, 07 2015.

[66] J. Scheuner, G. Mazlami, D. Schöni, S. Stephan, A. De Carli, T. Bocek, and B. Stiller. Probr - a generic and passive wifi tracking system. In *2016 IEEE 41st Conference on Local Computer Networks (LCN)*, pages 495–502, Nov 2016.

[67] D. Singelée and B. Preneel. Location privacy in wireless personal area networks. In *Proceedings of the 5th ACM Workshop on Wireless Security*, WiSe '06, pages 11–18, New York, NY, USA, 2006. ACM.

[68] D. J. Solove. "I've got nothing to hide" and other misunderstandings of privacy. *San Diego L. Rev.*, 44:745, 2007.

[69] Y. Song, C. Yang, and G. Gu. Who is peeping at your passwords at Starbucks? — To catch an evil twin access point. In *2010 IEEE/IFIP International Conference on Dependable Systems Networks (DSN)*, pages 323–332, June 2010.

[70] S. Trifunovic, B. Distl, D. Schatzmann, and F. Legendre. Wifi-opp: Ad-hoc-less opportunistic networking. In *Proceedings of the 6th ACM Workshop on Challenged Networks*, CHANTS '11, pages 37–42, New York, NY, USA, 2011. Association for Computing Machinery.

[71] O. Turkes, H. Scholten, and P. J. Havinga. Blessed with opportunistic beacons: A lightweight data dissemination model for smart mobile ad-hoc networks. In *Proceedings of the 10th ACM MobiCom Workshop on Challenged Networks*, CHANTS '15, pages 25–30, New York, NY, USA, 2015. ACM.

[72] G. Vanderhulst, A. Mashhadi, M. Dashti, and F. Kawsar. Detecting human encounters from wifi radio signals. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, MUM '15, pages 97–108, New York, NY, USA, 2015. ACM.

[73] M. Vanhoef, C. Matte, M. Cunche, L. S. Cardoso, and F. Piessens. Why MAC Address Randomization is Not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, ASIA CCS '16, pages 413–424, New York, NY, USA, 2016. ACM.

[74] O. Waltari. Offloading delay tolerant data through opportunistic networks. In *Proceedings of the 2015 on MobiSys PhD Forum*, PhDForum '15, pages 23–24, New York, NY, USA, 2015. ACM.

[75] O. Waltari and J. Kangasharju. The Wireless Shark: Identifying WiFi Devices Based on Probe Fingerprints. In *Proceedings of the First Workshop on Mobile Data*, MobiData '16, pages 1–6, New York, NY, USA, 2016. ACM.

[76] O. Waltari and J. Kangasharju. Quantifying the Information Leak in IEEE 802.11 Network Discovery. In K. R. Chowdhury, M. Di Felice, I. Matta, and B. Sheng, editors, *Wired/Wireless Internet Communications*, pages 207–218, Cham, 2018. Springer International Publishing.

[77] O. Waltari and J. Kangasharju. Prongle: Lightweight Communication over Unassociated Wi-Fi. In *Proceedings of the 35th ACM/SIGAPP*

Symposium on Applied Computing, SAC '20, New York, NY, USA, 2020. ACM.

[78] Wi-Fi Alliance. Launch of Wi-Fi CERTIFIED Passpoint™ Enables a New Era in Service Provider Wi-Fi®. *AUSTIN, TX, June*, 26, 2012.

[79] H. Wirtz, T. Zimmermann, M. Ceriotti, and K. Wehrle. CA-Fi: Ubiquitous mobile wireless networking without 802.11 overhead and restrictions. In *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, pages 1–9, June 2014.

[80] W. Xi, J. Zhao, X. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang. Electronic frog eye: Counting crowd using WiFi. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 361–369, April 2014.

[81] W. Xu, H. Zhou, Y. Bi, N. Cheng, X. Shen, L. Thanayankizil, and F. Bai. Exploiting hotspot-2.0 for traffic offloading in mobile networks. *IEEE Network*, 32(5):131–137, Sep. 2018.

Appendices

Appendix A

Mobility Modelling Through Wi-Fi Eavesdropping

This is a poster based on early work presented in [74]:

O. Waltari. "Offloading delay tolerant data through opportunistic networks". In *Proceedings of the 2015 on MobiSys PhD Forum, PhDForum '15, pages 23-24, New York, NY, USA, 2015. ACM.*

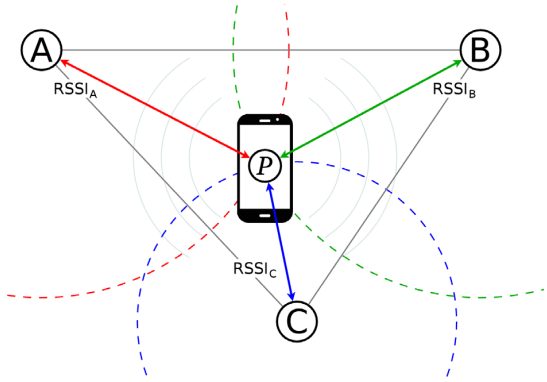
This poster has been presented at several events at the Department of Computer Science at University of Helsinki, but has never been peer reviewed.

MOBILITY MODELLING THROUGH Wi-Fi EAVESDROPPING

Otto Waltari
otto.waltari@helsinki.fi
University of Helsinki, Department of Computer Science

INTRODUCTION

Wireless network traffic is trivial to monitor with inexpensive hardware. Traffic is present almost anywhere at any time – thanks to various mobile apps which generate traffic on a regular basis. Even unassociated Wi-Fi enabled devices send probe requests to discover present WLAN access points.



METHODOLOGY

- All wireless network packets carry the MAC address of its sender
- RSSI level gives a rough estimate of the distance to a user
- Multiple fixed observation points (A, B and C) allows us to estimate the position *P* of a user through triangle geometry

→After capturing data over a period of time we can aggregate mobility traces of users!

EXPERIMENT

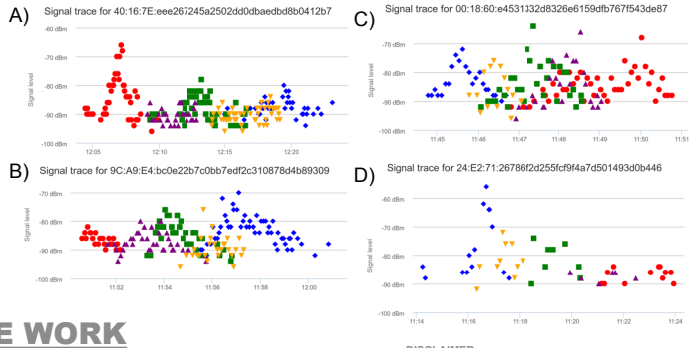
- At Esplanadi park in down town Helsinki
- Traffic monitored with Raspberry Pis and generic WLAN adapters
- Data was collected between 11:00 – 13:00 on a normal working day
- Over 40 000 network frames from over 4700 different devices

Locations of the observation points are illustrated on the left.

PRELIMINARY RESULTS

From the collected data we can distinguish users who traverse the Esplanadi park. Four traces of random users are illustrated on the right.

- Users A and B were moving from west to east
- C and D were going the opposite direction
- The amount of data points shows how active the device was in transmitting data
- Horizontal axis gives an idea of how fast the user was moving



MOTIVATION AND FUTURE WORK

- Mobility models are crucial in simulating networks with mobile entities
- Real-life derived models are more accurate than synthetic models

→Accuracy of chosen model reflects to simulation results!

Currently we are working on profiling traffic generation of mobile devices in different power states (idle, suspend, deep sleep, etc.). This far we have seen significant variance in probe requests sent by the same operating system on different devices. A better understanding of what to expect will provide our research a better foundation.

DISCLAIMER:
Due to privacy concerns we do not store actual MAC addresses. The first half is original, but the second half is one-way scrambled.

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE