Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider - Biskra
Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département d'Informatique

THESE

En vue de l'obtention du diplôme

DOCTORAT LMD

Spécialité : Informatique

Towards Greater Respect for the Users Privacy in communication technologies.

Réalisée par : Nour El Houda SENOUSSI

Soutenue le 04 / 02 /2021, devant le jury :

Foudil CHERIF: Professeur, Université de Biskra, Algérie	Président
Yacine CHALLAL: Professeur, École Supérieure d'Informatique, Algérie	Examinateur
Salim BITAM: Professeur, Université de Biskra, Algérie	Examinateur
Abdelmalik BACHIR: Professeur, Université de Biskra, Algérie	Directeur
Abdelmadjid BOUABDALLAH: Professeur, Université de Compiègne, France	Co-Directeur

Année universitaire

2020/2021

Acknowledgement

This thesis would not have been accomplished without the support and guidance of several individuals to whom I would like to express my gratitude in this acknowledgment.

First and foremost, my deepest and sincerest gratitude goes to my supervisor, Professor Abdelmalik Bachir, who guided me through the entire journey. His patience, valued insight and great support underpinned this PhD. I also appreciate the encouragement he gave me to fulfill my PhD.

I would like to extend my gratitude to Professor Abdelmadjid Bouabdallah from University of Compiégne, Heudiasyc Laboratory, France, for his support, guidance and assistance.

I would also like to thank Doctor Mohamed Lamine Kerdoudi from LESIA Lab, for his helpful comments, corrections and suggestions.

I would also like to thank all the members of jury, Pr.Foudil Cherif, Pr.Yacine Challal, Pr.Salim Bitam for agreeing to judge my thesis.

The most eminent thanks go to my dearest parents who have always been there for me, for the sacrifices made towards me, for their support and all the efforts made for my education and training. To my husband Bouzaher M.Zakaria for his support and help. To my sister Soumia and brothers Chems eddine and M.Younes, to my lovely little niece Nourane. To all my family members.

I would never forget to express all my gratitude to all my uncles: Dr.Sadok

Senoussi and his wife Nicole, Dr.Rachid Senoussi, and Madani, and my aunt Dr.Salima Senoussi, for their welcome during my intership in France.

My special thanks the entire research team of Heudiasyc Laboratory, CNRS UMR 7253, for their welcome, team spirit and in particular, Pr. Abdelmadjid Bouabdallah, and Pr.Salim Bouzebda for their welcome, without forgetting Mohamed Mohammedi, A.B.Meriem, N.Boutheina.

Finally, I address my thanks to all the members of computer science department of University of Biskra, the entire research team of LESIA Laboratory, and all my friends and colleagues who have encouraged me, D.Wafa, G.Meriem, R.Afaf, A.Haizia, O.Sara, S.Asma.

To all those who helped and supported me, directly or indirectly, in the accomplishment of this thesis, thank you.

I would like to dedicate this thesis

To my noble parents

To my husband

To my dear sister, my brothers, and my niece

To all those nice people that I love ...

SENOUSSI Nour El-Houda

Publications

International Journal

Nour El Houda Senoussi, Abdelmalik Bachir, Abdelmadjid Bouabdallah. *On QoS-aware location privacy in mobile networks.* Accepted in the International Journal of Information and Computer Security (IJICS 2018).

International Conference

Nour El Houda Senoussi, Mohamed Lamine Kerdoudi, Abdelmalik Bachir, Abdelmadjid Bouabdallah. *On Enhancing Location Privacy and QoS for Video Streaming Over Wireless Networks.* In Proceedings of the IEEE Global Communications Conference (GLOBECOM). IEEE, 2018.

Abstract

Wireless Local Area Networks (WLANs) are an emerging technology that have become increasingly popular in several locations such as businesses, educational institutions, Internet cafes, airports, etc. Since the first time WLANs appeared, users were able to communicate and exchange information and data from anywhere, access cloud services anytime, regardless of their geographic position or time. Due to their importance, WLANs which have been popular by the WiFi technology, are becoming more and more vulnerable to attacks and threats from other parties, where attackers can retrieve important data by capturing transmitted packets through access points, in which they can extract the MAC address of the transmitter, and thus be able to record the movements of the transmitter over time even when the system uses encryption as layer 2 headers, particularly the MAC address which uniquely identifies a mobile station is not encrypted. Existing solutions to this privacy breach aim at using pseudonyms to avoid using the same MAC address and also introduce silent periods to make it difficult for an attacker to track users.

Due to the downside of the existing solutions, in this thesis we propose two contributions, the first contribution consists of a mathematical model to quantify the privacy and a decentralized algorithm that use silent periods (SPs) to change the MAC address, and allow users to attain their desired levels of privacy while lowering its effect on the QoS perceived by them. The introduction of silent periods may have negative effects on some applications particularly for those with

constraints on bandwidth and delay such as video streaming. Therefore, in our second contribution, we propose an enhanced silent-period-based solution that allows to maximize the perceived QoS for a preset privacy. Our solutions offer users the best trade-off between privacy and QoS. We experimented our proposal with a set of numerical simulations. The obtained results demonstrated the efficiency of the proposed solution.

Keywords: Wireless Local Networks (WLAN), WiFi, Medium Access Control (MAC), Location Privacy, Decentralized Algorithm, Entropy, Quality of Service (QoS), Video Streaming.

Résumé

Les réseaux sans fil (WLAN) sont une technologie émergente qui est devenue de plus en plus populaire dans plusieurs endroits, tels que les entreprises, les établissements d'enseignement, les cybercafés, les aéroports, etc. Depuis la première apparition, les utilisateurs des WLANs ont été en mesure de communiquer et d'échanger des informations et des données depuis n'importe où, d'accéder aux services du cloud computing à tout moment, indépendamment de leurs positions géographiques ou de leurs horaires. En raison de son importance, les WLAN redus populaires à travers la technologie WiFi, sont devenus de plus en plus vulnérables aux attaques et aux menaces d'autres parties, où les attaquants peuvent récupérer des données importantes en capturant les paquets transmis via des points d'accès, desquels ils peuvent extraire l'adresse MAC de l'émetteur, et ainsi pouvoir enregistrer les mouvements de l'émetteur dans le temps, même lorsque le système utilise le chiffrement comme l'en-tête de couche 2, en particulier l'adresse MAC qui identifie de manière unique une station mobile n'est pas chiffrée. Les solutions existantes à cette atteinte à la vie privée visent à utiliser des pseudonymes pour éviter d'utiliser la même adresse MAC et introduisent également des périodes de silence pour rendre difficile le suivi des utilisateurs par un attaquant.

En raison des inconvénients des solutions existantes, nous proposons dans cette thèse deux contributions. La première contribution consiste en un modèle mathématique permettant de quantifier la vie privée, et un algorithme décentralisé

qui utilise des périodes de silence pour changer l'adresse MAC, et permet aux utilisateurs d'atteindre les niveaux de vie privée souhaités tout en diminuant son effet sur la qualité de service. L'introduction des périodes de silence peut avoir des effets négatifs sur certaines applications, en particulier pour celles qui ont des contraintes de bande passante et de délai, telles que le streaming vidéo. Dans notre deuxième contribution, nous proposons une solution améliorée basée sur la *période de silence*, qui permet de maximiser la qualité de service perçue pour atteindre une vie privée prédéfinie. Nos solutions offrent aux utilisateurs le meilleur compromis entre la vie privée et la qualité de service. Nous avons expérimenté nos propositions avec un ensemble de simulations numériques. Les résultats obtenus démontrent l'efficacité des solutions proposées.

Mots clés : Réseaux sans fil, WiFi, MAC, Confidentialité de l'emplacement, Algorithme décentralisés, Entropie, Périodes de silence, Qualité de service, Streaming vidéo.

ملخص

الشبكات اللاسلكية WLAN هي تقنية ناشئة أصبحت ذات شعبية متزايدة ومهمة في عصر الحوسبة. منذ ظهور WLAN لأول مرة، تمكن المستخدمون من التواصل وتبادل المعلومات والبيانات من أي مكان، والوصول إلى الخدمات السحابية في أي وقت، بغض النظر عن موقعهم الجغرافي أو وقتهم. نظرًا لأهميتها، أصبحت الشبكة اللاسلكية أكثر عرضة للهجمات والتهديدات من الأطراف الأخرى، حيث يمكن للمهاجمين استرداد البيانات المهمة عن طريق التقاط الحزم المرسلة من خلال نقاط الوصول، والتي يمكنهم من خلالها استخراج عنوان MAC الخاص بالمرسل، وبالتالي يصبحون قادرين على تسجيل حركات المرسل بمرور الوقت حتى عندما يستخدم النظام التشفير كرؤوس الطبقة الثانية، وخاصة عنوان MAC الذي يحدد بشكل فريد محطة متنقلة ليس مشفر. تهدف الحلول الحالية لخرق الخصوصية إلى استخدام أسماء مستعارة لتجنب استخدام نفس عنوان MAC وكذلك تقديم فترات صامتة ليصعب على المهاجم تتبع المستخدمين. نظرًا للجانب السلبي للحلول الحالية، في هذه الأطروحة نقترح مساهمتين، تتكون المساهمة الأولى من نموذج رياضي لتقدير الخصوصية و خوارزمية لامركزية تستخدم فترات صامتة لتغيير عنوان MAC ، وتتيح للمستخدمين تحقيق مستويات الخصوصية المطلوبة في حين تخفض تأثيرها على جودة

الخدمة. قد يكون لإدخال الفترات الصامتة آثار سلبية على بعض التطبيقات وخاصة تلك التي لها قيود على عرض النطاق الترددي والتأخير في تدفق الفيديو. لذلك، في مساهمتنا الثانية، نقترح حلاً صامتًا محسنًا قائمًا على فترة تسمح بزيادة جودة الخدمة المصورة إلى أقصى حد لخصوصية محددة مسبقًا. توفر حلولنا للمستخدمين أفضل مفاضلة بين الخصوصية وجودة الخدمة. لقد جربنا اقتراحنا بمجموعة من عمليات المحاكاة العددية. أظهرت النتائج التي تم الحصول عليها كفاءة الحل المقترح.

الكلمات المفتاحية: الشبكة المحلية اللاسلكية (WLAN) ، WiFi ، MAC ، خصوصية الموقع، الخوارزميات اللامركزية ، Entropy ، فترة الصمت، جودة الخدمة (QoS) ، تدفق الفيديو.

Contents

List of Figures

List of Tables

General Introduction

Context

Nowadays, mobile devices equipped with WiFi technology have become ubiquitous and important part of our daily lives, especially with the rise of the internet which made people addicted to being online, keeping WiFi turned on all the time to benefit from the variety of services the internet provides, such as browsing websites, checking emails, or using social media.

In WiFi networks, every wireless interface is uniquely identified by a Medium Access Control (MAC) address that is used by the Access Point (AP) to identify mobile stations and grant them access to the wireless network through the association process. To seek association with APs and join networks, mobile stations constantly run a service discovery process which includes operations of authentication, in case of secure networks, and association.

Service discovery can be done either passively or actively. In passive service discovery, also known as passive scanning, a mobile station can just stay silent listening for announcement from APs providing services announced through periodic beacon transmissions by APs. Upon the reception of a beacon, the mobile station can choose to associate with a chosen AP and starts the association procedure, which is typically preceded by an authentication procedure in case of secure networks. However, in active service discovery, also known as active scanning, a mobile station

actively seeks association with a preferred set of APs, typically those the mobile stations associated with in the past. Active scanning is realized by the transmission of special management frames, called probe requests.

While it may seem that either active or passive scanning can be used interchangeable and it is a question of a choice the mobile station makes, technical considerations make active scanning a preferred method by mobile stations. One of the major drawbacks of passive scanning is that beacons announced by APs are broadcast at a rate of one beacon every 100ms to make a trade off between bandwidth usage of rapidity of service discovery. However, in some situation, particularly when users are using applications that require QoS guarantee such as VoIP, the discovery and the association with another AP in case of a hand-off between APs, passive scanning may lead to extra delay that would not meet users' experience requirements. Therefore, most of mobile stations use active scanning to accelerate the association process with APs, particularly in the case of hand-offs.

The use of active scanning for service discovery implies that mobile stations constantly broadcast probe request frames. Those frames are not encrypted and carry sensitive information on the user (MAC address), and also the list of their preferred networks (the names of those networks). Some network names may be revealing more information on the social behavior of the user.

Even when WiFi networks are securely protected, the current WiFi standard leaves the MAC header of encrypted data packets without encryption, which makes the MAC addresses available to location privacy attackers, such as the WiFi sniffers devices installed by attackers to keep records of the transmitted MAC frames and the identities (MAC addresses) of those transmitting them.

With the democratization and accessibility of low cost WiFi packet monitors, and the widespread of WiFi and its large-scale use in personal devices such as laptops, smart phones and tablets, it has become easy and appealing to identify

and track users by monitoring the MAC address used by their personal devices. Typical application aim at gathering statistics for enabling the prediction of user movements and occupancy ratios of premises. As an example, a tracking system that counts the number of users insides parts of a building has been developed in [149] and [156]. Also, the data extracted from WiFi management frames has been used in order to estimate trajectories [101], social relationships [16], waiting times in human queues [155], and in order to calculate density estimations [20]. In [115], the authors estimate the density of people in the real world environment such as the number of customer visits to a coffee shop. In [123], the authors estimate the number of people in Railway Station. In [102], the authors estimate the number of passengers on public transport.

Problem Statement

With the rise of applications making use of information leaking from WiFi equipped device and its use in a way that may be considered as privacy beaching, many concerns have been raised particularly because WiFi monitoring applications have been able to perform the following actions [42].

- *Identifying users*: in the current IEEE 802.11 specification the source address is included in the MAC-header, hence, the uniqueness of the MAC address can be used or combined with some probe request field to determine the owner of the device, as well as enables an adversary to track users by recording the location of the device.

- *Profiling users*: the exchanged probe request frames between WiFi devices and APs reveal important information about the device and its owner, such as the history of visited networks, which can be used to determine the habits of the owner, and the location the device visits.

- *Linking users*: the attacker can use the collected data to link different devices. As in [31], the devices with corresponding data may include a relationship between devices owners. Also, it can mean that the devices belong to the same owner.

To cope with these problems, a lot of research has been carried out to make it possible for users to avoid being continuously tracked while allowing them to continue profiting wireless services. Most of existing research efforts rely on the use of temporal identifiers, commonly called pseudonyms that change over time, instead of using a permanent MAC address. It has also been shown that the use of pseudonyms should be carefully designed, because otherwise attackers can make use of side information to establish correlations between old and new pseudonyms, and thus break the location privacy protection of the system. A common technique that is used is to avoid changing pseudonyms instantaneously but introduce a delay, usually called a silent period [80], when a pseudonym is about to cease being in use and before the introduction of a new pseudonym. These silent periods also need to be carefully chosen so that the entire privacy of the system is maximized.

The privacy of a system is usually quantified by the entropy metric [137]. In the case of privacy preserving solutions against location tracking, maximizing the entropy may require the need for using long and random silent periods along with pseudonyms. This approach may affect the quality of service perceived by users typically those requiring long sessions and low latency such as VoIP sessions. At best, the use of silent periods decreases the throughput available to the users. At worst, it makes applications completely nonfunctional. Therefore, to cope with these conflicting objectives, namely quality of service and privacy, users need to make good decisions that lead to the best trade-off for them. Traditional solutions have looked at the overall system and provided centralized solutions [50].

Existing solutions have not considered the problem of the Quality of Service

(QoS) that can be caused due to the use of Silent Period in some real application such as the video streaming. In this context, one key challenge is understanding the trade-offs in preserving the quality of services utility whilst ensuring user privacy. To achieve this end, two fundamental questions need to be considered.

- How to accurately model the relationship between the degree of location privacy protection and the perceived QoS associated with it?
- How can we offer to WiFi users the best trade-off between privacy and QoS, preferably with a completely decentralized solution?

Contributions

Throughout the course of this dissertation we provide solutions to the above mentioned problems. Our contribution has been oriented along two axes. The first axis is about modeling the location privacy solutions based on silent periods and pseudonyms along with their effect on QoS of applications, and the second is on providing users with a trade-off between location privacy QoS in typical media streaming applications. The summary of these contributions is as follows.

- QoS-Aware Location Privacy in Mobile Networks: [104]

  The first contribution aims at answering the first research question. It provides a comprehensive mathematical modeling and analysis that provides a clear understanding of the relation between privacy entropy, the use of pseudonyms, silent periods, and side information expressed as the mobility pattern.

- Location Privacy and QoS for Video Streaming over Wireless Networks: [132]

  The second contribution aims at answering the second research question. It offers an algorithm for QoS demanding applications. We take as an example

YouTube. We aim at allowing users preserve their privacy without losing the QoS required which we defined in this example as the length of interruptions that may caused during the privacy protection process. We particularly proposed an enhanced *silent period*-based solution that allows to maximize the perceived QoS for a preset privacy. Our solution is elaborated from our mathematical model that allows to: (i) quantify the desired user location privacy and QoS, (ii) compute the silent and active periods that cope with these two objectives, and (iii) offer users the best trade-off between privacy and QoS for a typical video streaming application. We experimented our proposal with a set of numerical simulations and the obtained results demonstrated the efficiency of the proposed solution.

Thesis Outline

The remainder of this dissertation is structured as follows. Chapter 1 presents an introduction to existing wireless networks technologies focusing on the problems related to location privacy in them. Chapter 2 presents in detail the IEEE 802.11 technology. In Chapter 3, we then present an overview of location privacy major source of attacks, followed by a survey on the state-of-the-art solutions designed to cope with vulnerabilities of the IEEE 802.11 standard used by WiFi networks. In Chapter 4, we present the description of our model that takes into account location privacy degree with the QoS which we use in Chapter 5, to devise our proposed decentralized solution to offer the optimum trade-off between location privacy and QoS in video streaming such as YouTube. Finally, the manuscript ends with general conclusion synthesizing our contributions, as well as possible research prospects, which we wish accomplish in the near future.

Chapter 1

Wireless Networks and Related Location Privacy Issues

1.1 Introduction

Technological evolution has allowed the advent of wireless networks whose development is booming because of the comfort of connection they provide. Depending on the range of the network and the bit rates used, different wireless networking technologies are now being used: from a simple connection between a device or a computer with a close-by access point via a Wireless Local Area Network (WLAN) to larger distance connections through Wireless Wide Area Networks (WWANs).

In this chapter, we provide an overview of the main technologies used for wireless networking from short range personal and local networks to global wide area wireless networks. We present the main application scenarios of these different classes of networks, and focus on privacy problems and how they are addressed in existing literature.

1.2 Wireless Wide Area Networks (WWANs)

The first generation (1G) of Wireless Wide Area Networks has been in place since the early 1980s with the goal of enabling voice communication. These kinds of networks use analog signals based on AMPS (Advance Mobile Phone Service) offering up 2.8Kbps of bandwidth [46]. Since the early 1990s the second generation (2G) was introduced to improve voice communication and cope with the many disadvantages of 1G such as the very limited capacity [46]. The subsequent generations which have been evolving over the decades vary from each other in several aspects such as spectrum of frequencies, rate of data transfer, as well as techniques of radio access, and the level of security and privacy solutions used.

1.2.1 Evolution of WWANs

WWANs evolved significantly from the second generation (2G) to the current fifth generation (5G) with ongoing research targeting a next generation towards (6G). The following is a brief description of these various generations.

1.2.1.1 Second Generation (2G)

Commonly known as Global System for Mobile communication (GSM), 2G was introduced in 1990s with a main goal of providing mobile telephony. It offered services that allowed end-to-end mobile station communication across the network (communication between two mobile stations or between a mobile station and a fixed station). It was based on digital signals for voice communications and provided speeds ranging from 64kbps for the initial 2G to 200kbps for the advanced 2.5G General Packet Radio Service (GPRS) and Enhanced Data Rate for GSM Evolution (EDGE) [5, 46, 90].

1.2.1.2 Third Generation (3G)

Commonly known as Universal Mobile Telecommunication System (UMTS), 3G was introduced in 2000s and provided a technology that efficiently enabled data packet switching. It increased data transfer bandwidth providing speeds ranging from 144kbps up to 2Mbps. The advantages of 3G are numerous including fixed and mobile Wireless Internet Access, Video calls, and Mobile TV, and security of communication, etc. [5, 36].

1.2.1.3 Forth Generation (4G)

Commonly known as Long Term Evolution (LTE), 4G was introduced in 2010. It provided similar features as 3G with additional services stemming from higher data speeds such as high-definition mobile TV, video conferencing, and multimedia messaging services. The data transfer rates of 4G is much faster than that provided by 3G with the 4G-LTE-Advanced achieving theoretical speeds of 300Mbps in downlink and 150Mbps in uplink [111].

1.2.1.4 Fifth Generation (5G)

5G aims at providing even better performance with higher bandwidth (up to 10 Gbit/s) [5] and lower latency (1 millisecond) [138]. 5G is expected to accelerate the development of the Internet of Things, and keep providing the features and services already provided by 4G.

1.2.2 WWAN Architecture

The applications over WWANs, also commonly known as Cellular Networks, are various and include: operator applications, consumer applications such as mobile television, VoIP, video conferencing, data telemetry and automotive applications,

mobile web services such as music and video streaming [118]. These applications are being made possible thanks to the following architecture (see Figure 1.1).

The architecture of WWANs can be viewed as a hierarchic system composed of four main different network components, which are: Mobile Stations, Base Stations, Base Station Controllers, Mobile Switching Centers.

- **Mobile Stations (MS):** A MS, also called User Equipment (UE), is a mobile phone, with a Subscriber Identification Module (SIM). Each MS has a unique physical identifier called the International Mobile Equipment Identity (IMEI). The identification of a MS by the network is realized with the help of SIM which contains the International Mobile Subscriber Identity (IMSI). IMSI is unique digit serial number concatenated of: the Mobile Country Code (MCC), the Mobile Network Code (MNC), and the Mobile Subscribe Identification Number (MSIN).

- **Base Station (BS):** A BS is the responsible for the wireless connections between a MS and a Mobile Switching Center (MSC). It is also called Base Transceiver Station (BTS).

- **Base Station Controller (BSC):** A BSC is a mobile network component with the role of controlling one or more base station (BS).

- **Mobile Switching Center (MSC):** A MSC, also called Mobility Management Entity (MME), is responsible for the authentication, routing, handoffs over different Base Station Controllers, etc. It maintains four databases: Home Location Register (HLR) to store personal information of the subscriber (IMSI, phone number, etc.), Visitor Location Register (VLR) to store dynamic information of the subscriber, Authentication Center (AuC) which holds the access data of the subscriber (secret key of SIM), and the Equipment

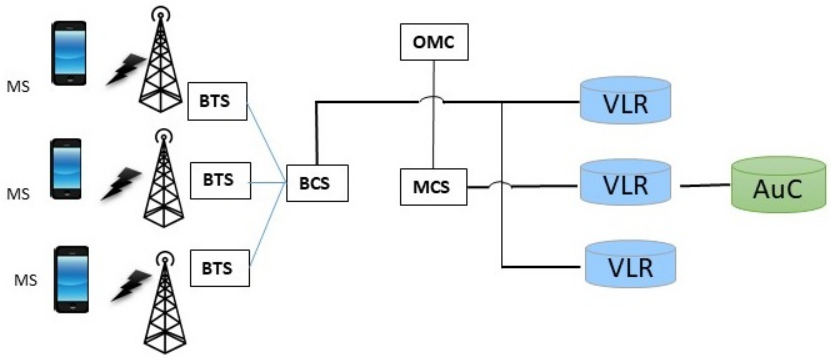

Figure 1.1: WWAN Architecture Components.

Identity Register (EIR) which stores the IMEI.

1.2.3 Location Privacy Issues and Solutions for WWANs

The introduction of cellular networks and their widespread use have opened doors to many problems related to the privacy of users ranging from eavesdropping voice and data contents to tracking user physical movements. Due to the large spectrum of privacy problems, we focus on this thesis on describing the main location privacy issues and the proposed solutions.

1.2.3.1 Issues

As described earlier, 2G/3G technologies rely on a permanent identifier (IMSI) that is linked to the SIM card. As IMSI is linked to the SIM card in a permanent way which could be a direct personally identifying information, a temporal pseudonym identifier called Temporary Mobile Subscriber Identity (TMSI) is used in GSM. Version of TMSI such as P-TMSI, and M-TMSI are used for GPRS and 4G respectively [44]. Even with a use of a temporary pseudonym, attacks on location privacy can be carried out by passive listeners. These attackers can detect the presence of a given user, and collaborate together to trace the movement of the user.

Several attacking schemes have been published in the literature [26, 59, 60, 61, 77, 134]. Further attacks to even reveal the IMSI can also be carried out through the IMSI-catcher technique which consists in installing rogue "towers" acting as a Man-In-The-Middle between the user and the operator's network. IMSI-catcher attacks can also be carried out in 3G and LTE networks by downgrading them to non LTE networks which do not require mutual authentication [134]. Other attacks such as the one described in [61] could also be carried out by exploiting 4G/5G paging protocol weaknesses to enable an attacker that knows a victim's phone number to verify the victim's presence in a particular cellular area.

1.2.3.2 Solutions

In order to avoid some of the aforementioned problems, works in [32, 40, 48, 70, 77, 85, 107, 146, 147] have addressed the IMSI catching problem, and proposed solutions to stop IMSI Catchers. For instance, in [146], the authors introduced the Pseudo Mobile Subscriber Identifier (PMSI) to defend against the IMSI catching attack. They replaced the IMSI with changing pseudonyms based on SIM information where the SIM uses the new PMSI the next time it is requested to reveal its IMSI. PMSI is encrypted in a semantically secure way to keep it confidential between the SIM and authentication server.

Besides, to prevent cellular devices from being illegitimately tracked, and prevent an attacker to correlate two paging messages sent to the same UE due to the infrequent update of TMSIs, authors in [139] analyzed prior attacks, and proposed new ephemeral UE identifier called "P-TMSI" (Pseudo-TMSI), a defense mechanism to protect the 4G and 5G cellular paging, P-TMSI will be used instead of TMSI, and will be refreshed by time-synchronized UE and core network simultaneously.

In addition, several contributions have been proposed to deal with location privacy attacks in WWAN. For example, in [59], the authors proposed GUTI (which

stands for Globally Unique Temporary Identifier in 4G networks) reallocation as a mechanism to update the temporary identifier in current LTE (4G) implementations. In [26], the authors proposed a schema using a dynamic mobile subscriber identity (DMSI) instead of the IMSI to protect the permanent of a user in LTE. In [71], the authors proposed a novel authentication approach for 3G (UMTS) and 4G (LTE) mobile systems that does not affect existing Service Networks (SN) and mobile phones while avoiding IMSI-catchers. In their solution, the IMSI is never sent across a communication channel. Instead a changing pseudo-IMSI is used. The pseudo-IMSIs appear as new subscribers to the service network (SN), and are unlinkable. In [35], the authors proposed a solution that is able to route calls to users who move through cellular networks without violating their location privacy through two new entities called pseudonym provider and location provider. In [142], the authors proposed a scheme that uses of multiple IMSIs for a single SIM.

1.3 Wireless Personal Area Networks (WPANs)

A wireless personal area network is mainly characterized by its small communication range which does not exceed a few tens of meters. This type of networks is generally used to connect peripherals (printer, mobile phone, etc.) to a computer without wired connection. There are several technologies used for WPANs such as the Bluetooth and ZigBee.

1.3.1 Bluetooth

Bluetooth is a short range wireless communication technology operating in the 2.4 GHz frequency. It was invented in 1994 by Ericson [13] with the goal of creating an ad hoc wireless network that allowed devices to connect with one another. It is based on the standard IEEE 802.15.1 for wireless connections.

There are three classes of devices offering three connectivity ranges [89]. Class 1 devices transmit at 100mW and offer a range of 100 meters, Class 2 devices transmit at 2.5mW and 10 meters, and Class 3 devices which transmit at 1mW and have a range of about 1 meter.

Bluetooth provides both point-to-point and point-to-multipoint wireless connection. It allows fixed and mobile devices to easily transmit both voice data at relatively high speeds over a short distance. The connection between two or more Bluetooth devices form a *Piconet* [127] (see Fig 1.2). As an example, a Piconet can be formed by a connection between a cell phone (master) and wireless headset (slave). The master is the one that initiates the Piconet by first searching for devices within its range.

Every Bluetooth device has a unique physical address (48 bits) that is assigned by the manufacturer. While necessary for the functioning of Bluetooth protocols, this address creates vulnerabilities which can be exploited to carry location privacy attacks and track users movements.

Figure 1.2: Bluetooth Example.

1.3.1.1 Bluetooth Architecture

Bluetooth technology is nowadays widespread and adopted in a variety of portable devices, such as laptops, phones, headsets, speakers, printers, keyboards, automobiles and medical devices etc. In addition, Bluetooth is being used in automating smart homes to provide monitors and access control for IoT devices such as lights, thermostats, door locks, security system, medical devices.

In order to establish a connection between two devices and exchange data in the form of packet, Bluetooth devices need to be authenticated to form a trusted relationship through three main steps in Bluetooth security procedures, which are: authentication, authorization and confidentiality [28].

In Bluetooth a trusted relationship "authentication or pairing" between two devices are formed by exchanging shared secret codes referred to as PINs in order to encrypt the communication. A "master" device has the option of pairing with up to seven "slave" devices establishing a network called a piconet, and the connectivity in piconet is in an ad-hoc manner [28]. The network topology of a group of wireless devices that connect to another via Bluetooth technology can be changed due to the movement of the devices inside the piconet. Moreover, every Bluetooth device has a unique Bluetooth device address. Using this address a certain Bluetooth device can be identified, tracked and monitored, which may be potentially dangerous to the privacy of user's personal information.

1.3.1.2 Privacy Issues and Solutions for Bluetooth

Bluetooth devices are exposed to several security threats. Most of them are due to the process of paring one device to another. There are a number of security threats for Bluetooth which can be found in [57]. In this thesis, we are most interested in location privacy attacks.

Nowadays, the majority of smart devices (smartphones, laptops, tablets) are

equipped with both Wi-Fi and Bluetooth wireless communications. Both network interfaces are identified by a unique 48-bits MAC address, assigned during the manufacturing process and unique worldwide. This unique address is included in every frame sent by the device, can be easily collected through packet sniffing and later used to perform higher level analysis task or tracking [65, 128].

In [58, 89, 97], the authors presented different types of attacks named Blue-Printing, Blue-Stumbling, Blue-Tracking, with the goal of extracting information about the victim's device by monitoring it closely and using the collected information for future attacks.

Other tools have been used for packet sniffing [163], such as Ubertooth One [106], which is an eavesdropping tool that is used to monitor the Bluetooth traffic [8].

Bluetooth tracking has also been used in applications that aim at estimating the occupancy of spaces such as in [88] where the authors proposed a system based on capturing both WiFi probe requests and Bluetooth management frames. In [87], the authors proposed an algorithm for pairing WiFi and Bluetooth MAC addresses to improve the precision of indoor localization and crowd density estimation systems.

Regarding solutions to location privacy in Bluetooth, the authors in [7, 8] proposed BlueEar, a practical Bluetooth traffic sniffer, and proposed practical counter measure that can reduce the packet capture rate of the sniffer. However, as a general rule, it is recommended to turn off Bluetooth radio completely when it is not in use [57].

1.3.2 ZigBee

The ZigBee [17, 75] is a wireless technology standard that defines a set of communication protocols for short range communications. It has been particularly designed for home appliances control over a sensor network, on the top of the IEEE 802.15.4 standard for wireless personal area networks (WPANs) [3]. The ZigBee

protocol stack is specified and maintained by the ZigBee alliance which handles the software part by defining the network, security and application layers. However, the IEEE 802.15.4 handles the hardware part by defining the physical and media access control layers for low range wireless personal area network (LR-WPAN) [75]. ZigBee is a standard that addresses the need of very low cost implementation for low power devices with low data rate for short range wireless communications. It operates in the 868MHz (Europe), 916MHz (North America and Australia) and 2.4GHz (available worldwide) ISM band, with up to 20kbps, 40kbps and 250kbps data rate respectively [119, 154], and a transmission radius up to 400 meters [159]. ZigBee can support a large number of nodes (up to 65000) [162].

1.3.2.1 ZigBee Architecture

ZigBee is generally useful for applications that need low data rate. It is the most used communication technology in home automation and smart lighting [120, 159]. Figure 1.3 illustrates a ZigBee example in home automation. The ZigBee is used for example in following areas: Health (e.g., patient surveillance, equipment monitoring, health and fitness monitors) [103], Home Automation (e.g., automated meter reading, temperature and humidity controls, intrusion detection systems) [39, 82], Industrial Application (e.g., environment control (HVAC), vehicle auto-diagnosis, warehouse stock location), Smart Agriculture Application (e.g., temperature, humidity, soil moisture and light intensity), etc.

From architectural point of view, ZigBee has mainly three topologies: Star Topology, Cluster Tree Topology and Mesh Topology. ZigBee also defines three types of nodes, Coordinator, Router, and Device. These form a ZigBee network as shown in the example of Figure 1.3, and are described as follows:

- Coordinator: a Coordinator is the root of the network and is one coordinator

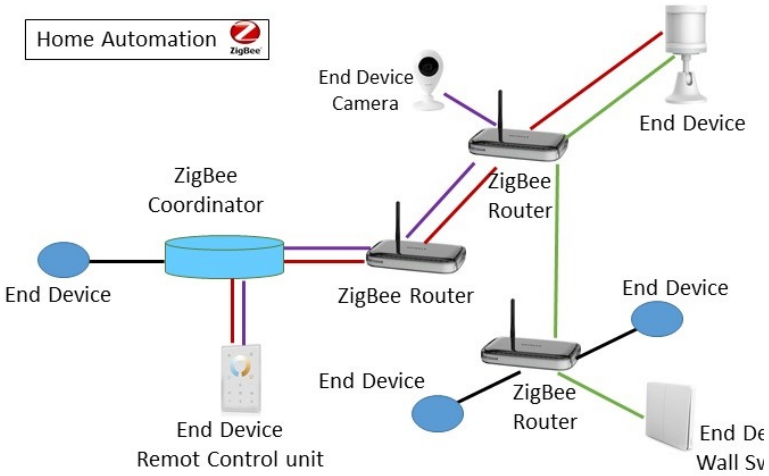

Figure 1.3: Example of a ZigBee Network.

per network. It is responsible for establishing the network, defining mode of operation, allowing and associating other nodes to the network.

- Router: a Router is an intermediate node which is only present in tree and mesh topology networks. It can associate end-devices to the network and relay data from other devices.

- End Device: an End Device is a simple node. It is may be a low-power battery-operated device. It collects various information from sensors and switches, and forwards it to a router or to the coordinator, but cannot transmit data from other devices.

Note that physical devices associated with IEEE 802.15.4 standard are of two types: Full Function Device (FFD) and Reduced Function Device (RFD). Full Function Devices (FFD) can perform all available operations within standard, including routing mechanism, coordination and sensing tasks. An FFD plays can play the roles of coordinator, router, or end devices, and can communicate with another FFD and a RFD. However, a RFD has a limited function, and can only

communicate with an FFD, typically a router or the coordinator [152]. RFDs are typically sensors which collect data and forward it to other FFDs [76].

1.3.2.2 Privacy Issues and Solutions for ZigBee

Similar to other wireless networking technologies, ZigBee networks are prone to a large spectrum of security attacks with a multitude of goals [72, 144, 150] and a lot of research has been carried out to propose solutions as a countermeasure to these attacks. Regarding privacy, and particularly attacks against location privacy, there have not been significant amount of work in this area as most components of a ZigBee network are rather immobile or with very limited mobility. Therefore, most of the attacks have focused on aiming at eavesdropping packets with the aim of getting access to the content of the data included in the packets themselves or establishing behaviors of profiles which can be built by learning from traffic patterns (e.g., duration and frequency of messages, location of the home user, etc.) [6, 105]. Another form of attack prevention has been based on exploiting physical characteristics of the received signals such as the one presented in [69] where the authors proposed to analyze the Received Signal Strength (RSS) values of received frames to prevent identity theft or data spoofing.

1.4 Wireless Local Area Networks (WLANs)

Similar to previously mentioned wireless technologies, WLANs are also prone to attacks on location privacy stemming mainly from the exposure of permanent physical unique identifiers to attackers monitoring communication radio channels.

As attacks on location privacy on more prominent in WLAN networks, particularly those based on WiFi/IEEE 802.11 due to their widespread and omnipresent usage caused by the democratization of smartphones, we dedicate our work in this

thesis on this area of research. In next chapters, we study WiFi networks and locations privacy problems and solutions in details.

1.5 Conclusion

In this chapter, we provided an overview on the main technologies of popular wireless networks in use, namely, wireless personal area networks (WPANs), and wireless wide area networks (WWANs). For each category we summarized the prevalent enabling technologies, then, we present their architectures. For these networking technologies, we presented the main weaknesses that led to attacks on location privacy and we showed that they are more prominent in networks linked with devices handheld and related to human users. We have shown that attacks on location privacy are less prominent in ZigBee networks and that also should apply to other device to device technologies where mobility is very low. In the next chapter, we will present WLAN technologies based on WiFi in details as the main work in this thesis has been carried out on location privacy in WiFi-based WLANs.

Chapter 2

IEEE 802.11 WLANs Fundamentals

2.1 Introduction

Wireless Local Area Network (WLAN) is a technology that allows one or many users to form a local network without relying on a physical wire. They use wireless technology, mainly radio, to connect users within an area of a few hundred meters of range. As wireless technologies evolve very rapidly, there has been a series of standardization efforts which led to the adoption by the Institute of Electrical and Electronics Engineers (IEEE) of the standard 802.11. The term IEEE 802.11, also commonly known as WiFi, is a generic term which is followed by one or two alphabetical letters indicating different variants of the 802.11 family. Examples of such variants include 802.11b, 802.11g, 802.11ac, etc. Each variant is distinguished from the others by many characteristics such as the frequency band used, the bit rates, and the range in an open environment, etc.

In this chapter we provide an overview of the main operation of the 802.11 protocol stack which defines the physical and link layer protocols. We particularly

focus on the MAC messages exchanged and information therein which have impact on location privacy in such technology.

2.2 Definitions and Architecture Overview

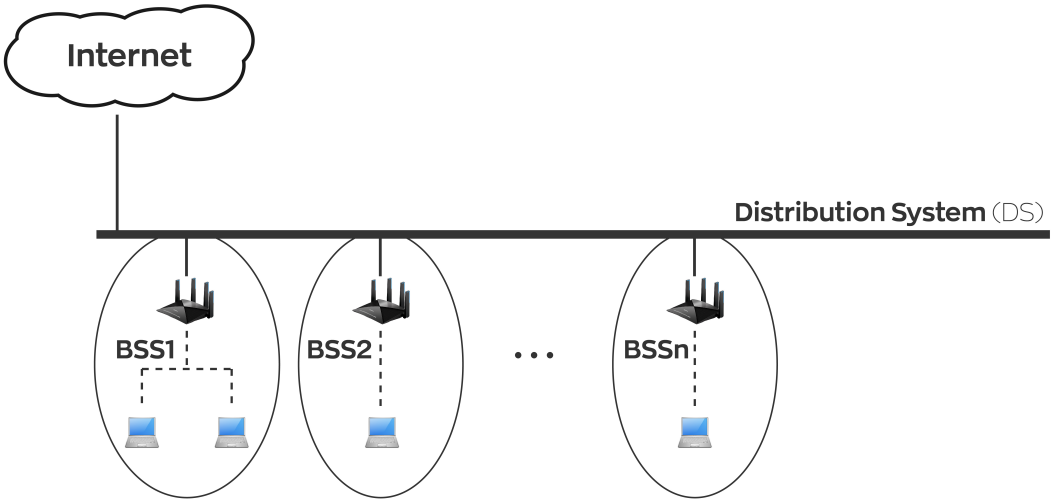

Figure 2.1: Topology of an IEEE802.11 WLAN.

The 802.11 standards have been approved for the first time in the year 1997 [79, 91, 133]. Commonly referred to as Wi-Fi, the 802.11 standards define an over-the-air interface between a client station and a base station access point (AP) or between two or more wireless client stations. The IEEE 802.11 standard is primarily designed for broadband connections, as extension of, or substitution for cable LANs.

Although the IEEE 802.11 standard defines two operation modes, the infrastructure mode which consists stations and access points connected to them, is more used than the ad-hoc mode which allows mobile stations to connect with each other without the need of an access point.

In infrastructure mode, the 802.11 WLAN consists of a set of APs and clients. An AP with clients associated to it form a Basic Service Set (BSS). Multiple BSS can be connected via a Distribution System (DS) and are usually connected to the Internet as shown in Figure 2.1.

The IEEE 802.11 standard focuses on the two lowest layers of the OSI reference model, namely, the Physical layer and the Data Link layers. This latter is itself subdivided into two sub-layers: the Logical Link Control (LLC) sub-layer, and the Medium Access Control (MAC) sub-layer. The physical layer is also subdivided into two sub-layers: the Physical Layer Convergence Procedure (PLCP), and the Physical Medium Dependent (PMD) sub-layers [63].

2.3 Addressing (MAC addresses)

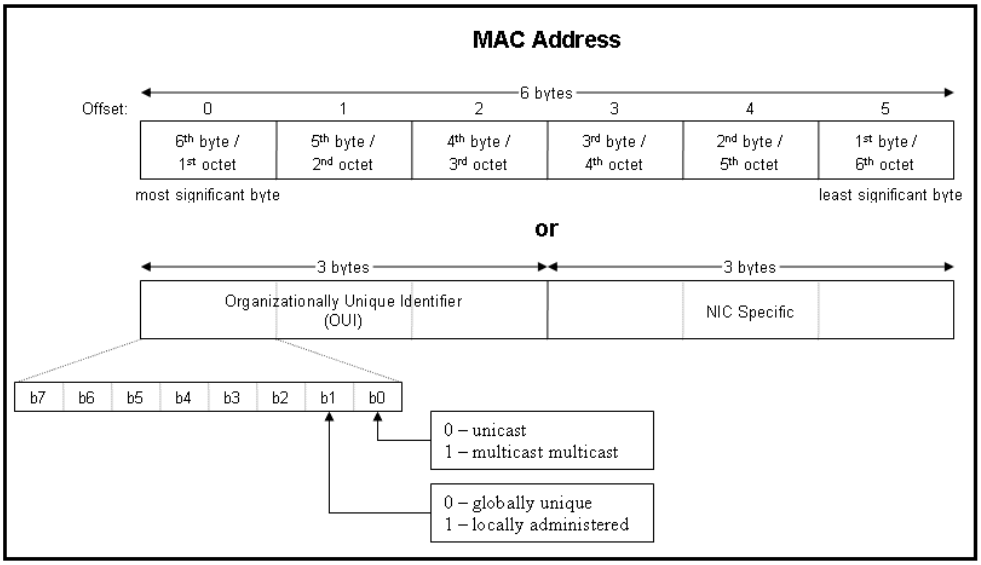

Figure 2.2: MAC address structure [92].

A MAC address, also known as the LAN/WLAN address or physical address is a link layer address, defined by the IEEE to uniquely identify devices. It is

composed of 48 bits (usually represented by hexadecimals so coded on 6 bytes), which yields a 2^{48} addressing space. As shown in Figure 2.2, the first three bytes (24 bits) of a MAC address are reserved, and only delivered by IEEE, this prefix is called Organization Unique Identifier (OUI), also known as Company Identifier (CID) [1]. The OUI can be used to identify manufacturers. The manufacturer is left responsible for assigning the remaining low order three bytes any value they wish when initializing devices to specify the model of Network Interface Card (NIC), provided they do not use the same MAC address twice to avoid addresses duplication.

Note that the MAC address is a permanent address, regardless of the position of the wireless device is, it always remains the same. Unlike the IP address which may change every time device moved from a subnet to another. As will be shown later, the MAC address is not encrypted and transmitted in plain text particularly with management frames, which opens doors for attackers to obtain it and use it to track users physical movements.

2.4 Frame Types and Structures

The 802.11 standard defines different frame types used by stations and access points for communicating, managing, and controlling the wireless medium. There are three main types of 802.11 frames: Data Frames, Management Frames, and Control Frames.

As shown in Figure 2.3, the general structure of a 802.11 frame contains the following fields.

- Frame Control: indicates the type of the frame (control, management, or data) and provides control information, which includes whether the frame is

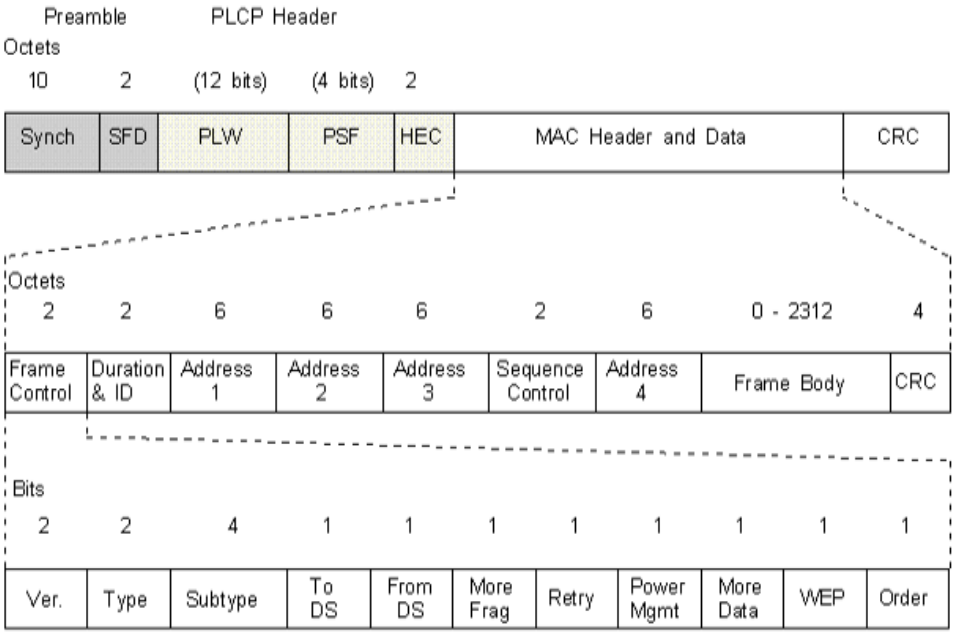

Figure 2.3: 802.11 Frame format [47].

to or from a DS (Destination), fragmentation information, etc. It contains the following fields: *Ver* which indicates the current protocol version number, *Type and Subtype* which indicate the frame type (Management, Control or Data frame) and the details of the frame type, *To DS and From DS* which indicates if the frame is to a destination or from destination, *More Fragments* which is set to 1 when more fragments are to follow, *Retry* which is set to 1 when the frame is a retransmission of an earlier frame, *Power Management* which indicates the power mode and set to 0 in the case of Power Saving Mode (PSM), *More Data* which is set by the AP to indicate that more frames are destined to a particular station that may be in power save mode, *WEP* which is set to 1 if the body of the frame is encrypted, *Order* which is set to 1 if the frame is being sent according to the Strictly Ordered Class.

- Duration/Connection ID: indicates the time (in microseconds) the channel

will be allocated for successful transmission of a MAC frame. In some control frames, this field contains an association, or connection identifier.

- Address 1: contains the MAC address of the receiver or Destination Address (DA) of the frame.
- Address 2: contains the MAC address of the initial sender or Source Address (SA) of the frame.
- Address 3: contains the MAC address of the intermediary recipient of the frame.
- Address 4: contains the MAC address of the intermediary sender of the frame.
- Sequence Control: is split in two parts: the first is a 4 bits fragment number subfield, used for fragmentation and reassembly (indicates the fragment number, if the frame is fragmented), and the second is a 12 bits sequence number, used to number frames sent between a given transmitter and receiver and is incremented with every transmitted frame.
- Frame Body: is variable in size, and contains the payload from higher-layer protocols.
- Frame Check Sequence: contains the result of applying CRC-32 polynomial on MAC header and frame body. It is used at the receiving station to check for transmission errors in the frame.

2.4.1 Management Frames

Management frames are used for establishing, maintaining, and terminating a connection between an AP and a wireless station. Some common 802.11 management frames subtypes include:

- Beacon: it includes capability information and parameters. The AP regularly sends a beacon frame to announce its presence and send information (SSID, timestamp, etc.)

- Probe Request: sent from a station when it requires information from another station (such as which APs are within range), or to join a wireless network. Figure 2.4 illustrates the structure of the frame.

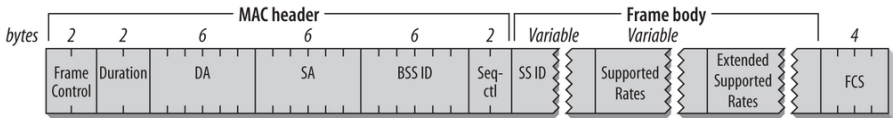

Figure 2.4: Probe request frame format [47].

- Probe Response: AP responds to a probe request with a probe response, detailing capability information such as supported data rates, etc.

- Authentication Request: in order to authenticate, a station sends authentication to an AP with its identity. The access point responds with an *Acknowledgement* frame.

- Deauthentication: sent from a station to an AP if it wishes to stop connection.

- Association Request: sent from a station to an AP to allocate resources and synchronize.

- Association Response: sent from an AP to a station with either an acceptance or rejection of an association request.

- Re-association Request: sent by station when it goes out of the range of the AP it is currently associated with when finding another AP with stronger signal.

- Re-association Response: sent from an AP containing the acceptance or rejection to a station reassociation request frame.
- Disassociation: sent from a station wishing to terminate the association.

2.4.2 Control frames

Control frames facilitate the exchange of data frames between stations and APs and are used to avoid/alleviate certain problems such as the hidden node problem.

- Request To Send (RTS): before sending a DATA packet, the station sends a RTS to the destination station.
- Clear To Send (CTS): upon reception of a RTS the destination replies with CTS.
- Acknowledgement (ACK): upon reception of a data frame, the receiving station sends an ACK to the source station if no errors are found.

2.4.3 Data frames

Data frames carry higher-level protocol data in the frame body. Depending on type of data frames, some fields may not be used.

2.4.4 Frame Classes

IEEE 802.11 frames are divided into three different classes, Class 1, Class 2, and Class 3. Table 2.1 shows a summary of frames of each class.

- Class 1: are used to provide the basic operations used by 802.11 stations, or to allow stations to find an infrastructure network and authenticate to it.

Class	Control	Management	Data
Class 1	RTS	Probe Request	Any frame with ToDS and FromDS set to 0
	CTS	Probe Response	
	Acknowledgment	Beacon	
	CF-End	Authentication	
	CF-End+CF-Ack	Deauthentication	
		ATIM (Announcement Traffic Indication Message)	
Class 2		Association Request/ Association Response	
		Reassociation Request Reassociation Response	
		Dissociation	
Class 3	PS-Poll	Deauthentication	Any frames including those with either the ToDS or FromDS bits set

Table 2.1: Class Frames [47].

- Class 2: are used to manage the association after a successful authentication with a network.
- Class 3: are used when a station has been successfully authenticated and associated with an access point.

2.5 Discovering and Joining a Network

A WLAN formed by an AP and mobile stations associated with it is called Basic Service Set (BSS). A BSS is identified by a 6 bytes identifier called BSSID. The BSSID is the MAC address of the access point (AP) running the BSS. Several BSS can be linked together using a Distributed System (DS) in order to form an Extended Service Set (ESS). Similarly an ESS is identified by an ESSID. An ESSID is commonly shortened to SSID which acts as the name of entire network. A DS

can be composed of two or more APs connected by LAN.

To associate with an access point, a mobile station passes by three phases: service discovery where devices search for and select an AP, authentication phase in which mobile stations identify themselves to the discovered AP, and the association phase where mobile stations associate with a successfully authenticated AP.

2.5.1 Service Discovery

Service discovery is the process of searching for the best access point. Mobile stations attempt to search for available wireless networks, and then attempt to associate with them. The IEEE 802.11 defines two modes of service discovery: passive, and active service discoveries.

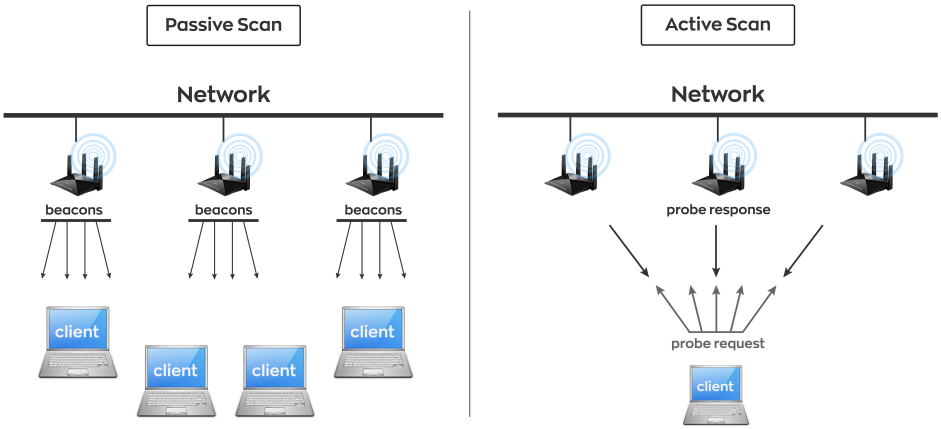

Figure 2.5: Passive and Active Service Discovery phase.

2.5.1.1 Passive Service Discovery

As shown in Figure 2.5, the passive service discovery, also commonly known as passive scan, does not require an action from the mobile station as it is the responsibility of APs announce their presence by periodically transmitting beacon

frames on a regular time interval (typically every 100 milliseconds). A mobile station searches for a network by just listening for beacons until it finds a suitable network to join. The mobile station listens to all the possible channels and waits to receive a beacon from an AP [100]. Beacon frames contain information about the AP, for instance, AP's MAC address, SSID, security policies and supported data rates.

2.5.1.2 Active Service Discovery

During active service discovery, also known as active scan, the mobile station is the one initializing the connection. The mobile station which seeks to join an AP broadcasts probe request frames periodically, for a number of times, on each channel. When an AP receives a probe request, it responds with a probe response back to the mobile station [31].

The active service discovery operates in two different modes according to weather or not an SSID is specified and carried in probe request frames. In the case, probe request does not carry an SSID, the mobile station broadcasts probe request frames on each channel. However, when and SSID is specified, the mobile stations unicast probe request frames to the requested AP.

Upon reception of a probe response, the mobile station makes a decision about joining the AP. If no transmissions are heard by probe delay period, the channel is declared empty and the mobile station restarts the process on another channel. Note that due timing delays, active scanning is faster than passive scanning and is frequently used by mobile stations particularly to speed hand-off between APs and keep the likes VoIP and other real time applications at acceptable user experience [100].

2.5.2 Mobile Station State Machine

The process of joining an AP is called *802.11 State Machine* (See Figure 2.6). A mobile station, STA, may be in one of the following states [148]:

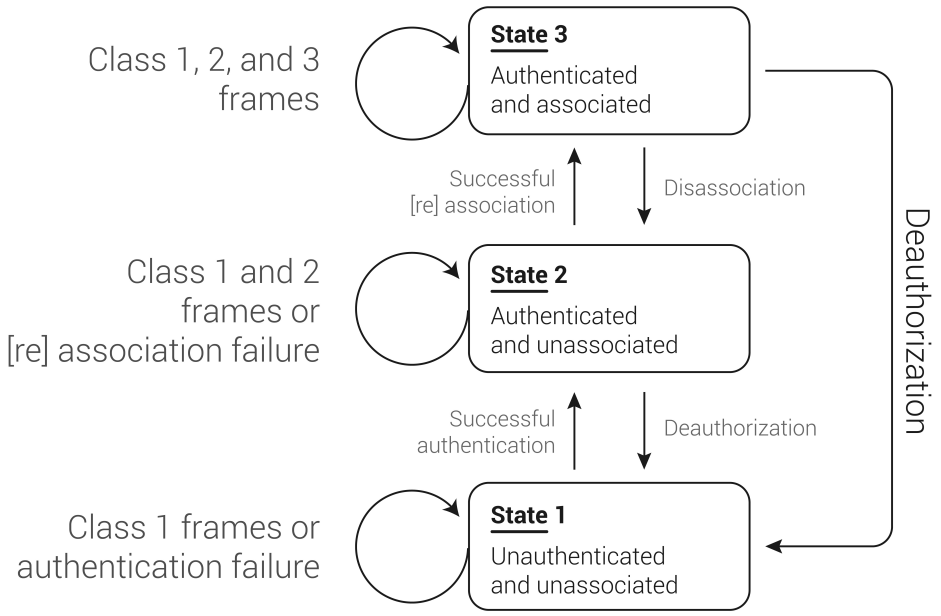

Figure 2.6: 802.11 State Machine automate [63, 92]

- Unauthenticated/Unassociated

  This state is also called State 1. In this state the mobile station is unauthenticated and unassociated, and it is looking for a network to join through active or passive scanning. After an authentication happens (either in open or encrypted systems), the mobile station moves to State 2. In State 1, only Class 1 frames are exchanged.

- Authenticated/Unassociated

The state is also called State 2. In this state the mobile station is authenticated to a network but not associated with the AP. Once in State 2, the mobile station can start an association procedure through the transmission of association request frames. In encrypted systems, association requests are encrypted. The mobile station can leave State 2 to either State 1 if the AP issues a deauthentication frame, or to State 3 if the AP issues and association response and accepts the association of the mobile station.

- Authenticated/Associated

  This state is also called State 3. In this state the mobile station is authenticated and associated. The mobile station reached this state if it receives a positive association response from AP and acknowledge it with an acknowledgment frame. In this state, the mobile station is able to exchange data frames with the AP. The station may leave State 3 to go either to State 2 if the AP sends a dissociation frame, or to State 1 if the AP sends a deauthentication frame.

Note that management frames are exchanged to transition from a state to another. These management frames are transmitted in clear, except for association/reassociation/dissociation in encrypted systems. Those management frames can be captured and exploited to track user's mobility and carry out attacks against location privacy.

2.6 Security Vulnerabilities

WiFi vulnerabilities may be divided into two categories. The first category can be seen as attacks on devices with or without intrusion, causing or not service disruption, and the second category as tracking devices passively to collect useful

information. Although vulnerabilities and attacks exploiting them in WiFi are numerous, we limit our study to those in relation with the work carried out in this thesis which is mostly focused on physical identification of devices.

2.6.1 Spoofing

Spoofing [10, 68] is a technique where an attacker masquerades the identity of a legitimate node and sends malicious packets with another source address to disturb normal operation of the network. Spoofing can be carried out to steal the identity of mobile stations or the AP, with different goals ranging from service disruption (Denial of Service (DoS)) to Man-In-the-Middle which targets getting access to confidential information exchanged between communication pairs. Spoofing has attracted a lot of research in the literature (e.g. see some surveys in [81, 121]) with solutions aiming at detecting spoofing (see [18, 136] as examples).

In MITM attacks the attacker is located in the network topology between two participants of the communication and acts as an intermediary routing traffic through itself. Typically, the attacker uses a spoofed MAC address of the AP and sends deauthentication frame to the victim station [27]. The attacker then forward back and forth frames from and to both communication pairs. Spoofing attacks has typically been dealt with and avoided by the use of strong mutual authentication protocols [27].

2.6.2 Tracking

As it will be detailed in the next chapter, passive monitor can be installed to capture frames transmitted by mobile stations. A part of these frames can be used to identify their holders and thus can be exploited to track their movements [130]. These captured frames can also be combined with other captured frames such as

from Bluetooth to further refine the tracking [4]. Major solutions to tracking will be presented in the next chapter.

2.7 Conclusion

In this chapter, we provided a detailed overview on the IEEE 802.11 standard focusing on operations generating message transmission by a mobile station. We showed that various types of messages are transmitted for reasons of management: discovery of service (i.e. APs), authentication, association, power saving, etc. We have shown that a number of these messages may carry sensitive information and are transmitted in clear. We have shown that one of the most exploited personally identifying information that could be used to carry out attacks on location privacy is the MAC address of the mobile station.

As it will be shown in the next chapter 3, there have been in the literature a lot of attacks that have been designed to track users movements by installing passive monitoring APs with the goal of capturing messages transmitted by mobile stations, storing them in databases and exploiting them to build a movement history of mobile users. In Chapter 3, we review main contributions, the key idea driving their design, and show when possible the counter measures have been proposed to alleviate their effect and preserve users mobility privacy.

Chapter 3

Location Privacy in 802.11 WLANs

3.1 Introduction

The number of people carrying mobile devices equipped with a WiFi has been constantly increasing, which has consequences on their privacy including their location privacy. For instance, an adversary can track the location of mobile device by installing passive WiFi monitors or controlling a number of APs to capture packets transmitted by mobile users.

Over the years, there have been numerous papers published about location privacy. These papers propose a set of solutions to cope with this problem and prevent attackers from tracking users. In this chapter we focus on location privacy problem.

We start by introducing the privacy concept in Section 3.2, then we present the major sources of the problem in WLANs in Section 3.3, and the main recently proposed solutions to assure location privacy in Section 3.4. In Section 3.5, we present the entropy metric that can be used to measure the degree of location

privacy, and we conclude in Section 3.6.

3.2 Concept of Privacy and Examples

3.2.1 Privacy as a General Concept

Privacy is recognized as a right by many countries in the world. Some have even codified it in their laws. For example, England's 1361 Justices of the Peace Act is the first known piece of privacy legislation, in which the arrest of eavesdroppers and stalkers has been legislated [11]. Also, the Fourth Amendment to the US Constitution proclaims citizens' right to privacy, and in 1890 US Supreme Court Justice Louis Brandeis stated that *the right to be left alone* is one of the fundamental rights of a democracy [23]. According to [131] the privacy is *the ability and/or right to protect our personal secrets, the ability and/or right to prevent invading our personal space*. A person has privacy when two factors are in place:

- must have ability to control information about themselves.
- must exercise that control consistent with their values.

3.2.2 Example of Privacy Issues in the Digital World

The use of Internet services leaves traces that making possible for a third party to collect, organize and analyze personal data. In fact, the collected information can be used for malicious purposes, such as the disclosure of location, political orientation, religious beliefs, lifestyle choice, and even identity.

One of the major sources of privacy breaching in the Internet is caused by the combination of the IP and HTTP protocols. While IP leaves traces about the communicating pair (source and destination IP addresses) available to attackers,

HTTP leaves much more additional information on the user's system and even the user's profile through the use of Cookies. Cookies are files that store information, unique identifiers, recent activities at a website, credit card details, or site password information. Cookies are sent by web servers to web browsers and which may then be sent back to the server each time the browser requests a page from the server and create the opportunity for more automated interaction between a web server and a client. Cookies may also allow a third party to disclose a user's activities if they have access to their computer and their cookie files. The use of Cookies has been increasingly legislated recently to limit the amount of information they collect on users and to inform them on what is being collected while browsing a particular website.

The use of encryption to hide sensitive and personally identifying information has helped a great deal to minimize privacy breaching in the Internet as the content of the exchanged messages can be encrypted end-to-end to make its content only available to communicating pairs. Although encryption helps to hide message contents, it is not sufficient to hide source and destination of the messages which can be extracted from the IP addresses of the source and the destination of the messages that are located at the IP header of the packets. There are techniques which users can use to make possible to know the source of the IP address by using anonymous proxy servers that change the IP address of the source with another so that the source of the packets can not be known easily (sometimes it is impossible). With encryption at the application layer such as SSL/TLS and IP address anonymity, privacy of users can be significantly enhanced. However, other threats are also possible at lower layers by exploiting a number of characteristics that make is possible to identify personally identifiable information about users.

3.2.3 Location Privacy

With mobile phones and devices being in use, another dimension of privacy is also considered, namely the location privacy. In fact, the location privacy means the ability to prevent other parties from learning one's current or past locations. As shown in Chapter 1, almost all wireless communication technologies have vulnerabilities which make them prone to attacks on location privacy. With the democratization of WiFi and the availability of open software and hardware [2], attacks on location privacy on WiFi have been much easier to perform [33, 94, 101, 108].

In fact, it is relatively inexpensive to deploy enough WiFi passive monitor to cover large areas for location tracking, compared with covering the same area with a cell-based tracking systems. The received signal strength of the mobile user at different APs can be used with triangulation to find the approximate location of the originating node. The knowledge of location, combined with other patterns, can be used to link a MAC address to a user (e.g. if a MAC address is identified to be in particular office during working hours, then it is probably the MAC address of one of the devices used by the person working at the office) [67]. When a MAC is linked to an identity, past and future movements will be easily mapped.

3.3 Sources of Location Privacy Problems in WLANs

WiFi technology has been mainly designed to offer wireless Internet access to mobile station. Privacy in general and location privacy in particular have not been included in the design of the standard. As a result many weaknesses have been exploited by attackers to carry out attacks on privacy. Regarding location privacy breaching, the major sources exploited by attackers are the following.

3.3.1 Wireless Channel on an ISM Band

One of the risks of WiFi is due to the use of the free Industrial, Scientific, and Medical (ISM) 2.4GHz channel combined with an omnidirectional Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) channel access protocol. As a result, it is easier for attackers to get hold of low cost on-the-shelf 2.4 WiFi compatible radio chips and be in the range of the victims to be able to receive and decode, when non encrypted such as management frames, transmitted by all mobile stations in its range.

3.3.2 Non Encryption of PII

As shown in Chapter 2, even when encryption is used to secure the content of WiFi transmission, it only applies to the data, i.e. to the payload of the packet, and leaves the header and the information therein in clear. Header information include sensitive information which can be considered a Personally Identifying Information (PII) such as the MAC address. As it is shown below, the headers also carry other information which could lead, with combination with other pieces of information, to implicitly identify mobile stations.

In addition, many management frames are transmitted in clear and they carry PII (such as MAC addresses, and other sensitive fields which can be exploited as side information).

While most of the attacks are being made by passive monitor capturing transmitted frames, particularly probe request frames, it is also possible to trigger those silent mobile stations which are relying on passive scanning in the discovery of APs. Typically, an attacker that wants to trigger transmission from passive mobile station replays beacons transmitted by legitimate APs [29]. The transmission of beacons triggers a response from the mobile stations such as association or authentication

requests depending on the security configuration in use. When the mobile stations react to beacon transmissions, attackers can make use of the PII included in the frames transmitted in response to the transmitted beacons.

3.3.3 Availability of Side Information

Wireless device can reveal a lot of information about their owner. Even when the MAC address is unknown, other features, referred to as "side information", could be used to identify a particular user with a certain level of certainty depending on which side information is used [92]. Side information includes: content of the rest of fields of the MAC header particularly in management frames, list of SSIDs sought by a mobile station when trying to actively associate with preferred APs, mobility patterns of the mobile station, Radio Frequency (RF) signature of the wireless card used by the mobile station, received signal strength, etc.

For instance, in [66] detect the user behavior by characterizing the changes in probe request frame transmission patterns (based on the frequency of sent prob requests), which occur as an effect of user's smartphone usage. In [122], the authors propose a method that uses the probe requests to determine the type of the devices handheld (smart phones) or non-handheld devices (laptops). In [16], the authors extract the vendor ID embedded in the MAC address to determine the socio-economical status of users. In [124], the authors collected a data set of probe requests to infer statistics about devices models in order to estimate the number of devices from a specific vendor or operating system that are affected by a security vulnerability.

In general, side information is also known in the literature under the name of *Fingerprinting* which consists in collecting enough information to identify or classify a target based on some observable features. As stated earlier, several sensitive information elements leak when a mobile station is trying to or is connected to an

AP, e.g. time of arrival, list of preferred SSIDs, etc. These pieces of information can be used to a (unique) digital fingerprint [110, 129] for users. Fingerprints can be created from information specific to devices (device fingerprinting) [151], or from information transmitted by users (behavioral fingerprinting) [53], and also can be used to infer relation between the devices [30], and hence the owners of these devices and their relationship [16, 31, 37, 84, 116].

- Device Fingerprints: device fingerprinting can be performed through *Radio Frequency Fingerprinting* [55] by identifying the signal characteristics of radio transmitting devices [49], where the signals are different for each device. Device fingerprinting may also be performed with another information such as the timing pattern of the probe messages to identify the wireless interface's driver [41].

- Behavioral Fingerprints: a behavioral fingerprint is made by combination of "implicit identifiers" that may be found in the frames transmitted by users such as MAC address, SSID, activation patterns in emails [78], request header field in HTTP sessions [45, 53], and create a unique fingerprint.

Following are some examples on how side information might be used to identify mobile stations [129].

3.3.3.1 List of the SSID History

As shown in Chapter 2, a mobile station frequently performs active service discovery through the transmission of probe request management frames seeking association with preferred APs. The transmission of probe request frames is done in clear and the content of the message contains the list of all SSIDs preferred by the mobile station. These SSIDs are typically those with which the station associated with in the past. The list of these SSIDs is likely to differ from a user to another and

might in some situations lead to uniquely identify users. Therefore, even if the MAC address is hidden through a certain mechanism, the list of SSIDs sought by a mobile station could be seen a vulnerability which could be used by attackers to launch location privacy attacks [29, 34, 83].

In addition to the possibility of identifying a mobile station, the list of SSIDs could lead to revealing more personal information on the users carrying them. The list of SSIDs may include the name of the workplace of the user, the list of public places they visited in the past, including hotels, train stations, airports, restaurants, etc. In extreme situations where the home gateway has a unique identifier, it may lead to identifying the home of the user [126]. The availability of databases containing geographic locations of APs makes the situation even worse. For example WiGLE [157] is a Wireless Geographic Logging Engine storing more than 551 millions records (locations of APs in hotspots) (May 2019). Google location services also uses this feature by allowing AP owners share their physical locations with Google to help make localization services, particularly for indoor environments more accurate, for mobile user applications [64].

3.3.3.2 Mobility Patterns

The link of changing MAC addresses together can be performed by knowing the speed and direction of a user. If a user changes its address MAC from MAC1 to MAC2, and that both MACs have the same mobility pattern such as the same speed and direction of movements, then it is probably the same user [67, 129].

3.3.3.3 Regularity of Patterns

As it will be shown in Section 3.4, the hiding of the real MAC addresses of mobile stations through the use of pseudonyms is not sufficient if done naively. For example, if pseudonyms are changed at regular time intervals it becomes easy to link all

pseudonyms used by a particular user [41].

3.3.3.4 Network Interface Card Signal Signature (NIC)

The signal generated by a WiFi card can be used to identify that card. In [21], the authors developed a technique to identify those cards. That identification method has proven resilient against ambient noise and fluctuations of the wireless channel. In [54], the authors also used radio frequency fingerprinting to uniquely identify card interfaces based on the transient portion if the signal it generates (improvement by correlating several observations in time and using of Bayesian filtering) [125].

3.4 Countermeasures for Location Privacy in Wireless LAN

Protecting location privacy has been attracting extensive research to allow users to continue enjoying Internet services without giving up private and sensitive information such as their physical locations to unauthorized entities and organizations. The location of users is typically inferred from the physical addresses of the personal devices (phones, tablets) they carry. As expressed in Section 3.3, most of privacy vulnerabilities in WLANs stem from three sources which are:

- the use of a common wireless channel that can be monitored with low cost on-the-shelf wireless devices,
- the transmission without encryption of explicit identifiers such as the MAC address by 802.11 compatible radios,
- the possibility of using side information such as other MAC header information [92], list of SSIDs [83], mobility patterns [129], wireless card RF

signature [125], etc. for implicitly identifying a mobile station.

In general, location privacy can be enhanced by tackling these vulnerabilities. Therefore, to provide more privacy to users, existing solutions tend to find a way to make the exploitation of these PIIs of little utility to the trackers. Most of the existing solutions are generally based on the following approaches: the use of pseudonyms, silent periods, link-layer header encryption, SSIDs list removal from probe requests, and traffic manipulation to reduce pattern-recognition-based attacks [29].

In the following we present the existing solutions to overcome privacy problem, and in table 3.1 we summarize the major advantages and disadvantages for each solution.

3.4.1 Random MAC Addresses/Pseudonyms

The use of pseudonyms or random MAC addresses from time to time aims to reduce identification risk with temporary interface identifiers, while making sure it is difficult to link new identifiers with previously used ones. Solutions based on this approach aim to solve location tracking attacks by using temporary addresses instead of the physical address (e.g. [38, 52, 98, 99, 117, 143, 149]).

This idea has also been used in cellular networks where the physical identifier, technically known as the International Mobile Subscriber Identifier (IMSI), is replaced with a temporary identifier called the Temporary Mobile Subscriber Identifier (TMSI) to avoid continuously identify a user (see Chapter 1), as well as in upper layers where pseudonyms have been used with IPv6 identifiers such as in [56] to provide more anonymity to users. In the case of WiFi, the physical identifiers (i.e. MAC addresses) are replaced with temporary addresses that are typically concocted using cryptographic functions (e.g. hash function). Solutions based on this approach are gaining success and have been adopted by the industry

(Apple's solution in iOS 8 although attacks have been already designed to defeat this solution [149] when traditional hash methods are being used). Other papers also showed it is possible to deanonymize devices, even when they use MAC address randomization [34, 42, 92, 94, 125].

Even though using temporary addresses seem to be an accepted solution against location attacks, recent research has shown that the change of MAC addresses should happen according to a well-designed way to avoid correlation attacks, aiming at linking new identifiers with old ones, leading to enabling tracking eventually even with when concocted temporary addresses are used. Usually the change of pseudonyms is combined with additional techniques such as the use of silent periods, traffic manipulation through the injection of fictitious packets, or according to particular locations known as Mix-Zones.

Limitations: although this is one of the best solutions adopted by the industry (Apple's iOS 8 [149]) and by recent research works [153], the temporal use of disposable MAC addresses creates some design problems such as address selection, uniqueness, and integration with port authentication, and may interfere with the operation of several protocols of the communication stack such as ARP and DHCP. Furthermore, the use of temporal MAC pseudonyms may also affect certain applications such as those with QoS constraints.

There are several challenges to randomly changing the MAC address:

- the selected MAC address should be valid under 802.11 standard (48 bits long, start with an OUI [62]), and unlikable to the old chosen address. As shown in [52], they use MD5 hash to generate the address by MD5 hash, and the remaining bits can then be filled in such that it is a valid IEEE 802.11 address [52].

- when several clients randomly choose their addresses, collisions are possible i.e. the chosen MAC address is already in use. Although the probabilities

for collisions in small networks are small, the problem of duplicate addresses becomes more apparent and the probabilities of address collision become high when the network extends over hundred of devices connected to several APs. Several duplicate address detection mechanisms are proposed to mitigate the collision problem, such as the work presented in weak Duplicate Address Detection (DAD) which is an important component of the address resolution protocol (ARP) that can detect duplicate addresses, even if the nodes that are assigned duplicate addresses initially belong to different partitions [140, 145].

- in some situations where the AP allows certain MAC addresses to connect to the network such in companies that have their WiFi set-up, a device that decides to randomly change its MAC address may lose its rights to connect. To mitigate this problem, an authentication protocol is proposed in [52] based on a symmetric key protocol. The user and home authenticator share a symmetric secret key k, which is exchanged at the time of subscription.

- in many systems restricting access to service according to MAC addresses, the DHCP identifies the user by its MAC address, and provide an IP address to get access to the Internet accordingly. Consequently, if the MAC address is not recognized by the DHCP, then no IP address will be assigned to the mobile station.

3.4.2 Mix-Zones

The concept of Mix-Zones has been introduced in [19], and has been used in other works such as [109], and in [160] to enhance location privacy in Vehicular Social Network (VSNs), and in [9] for protecting privacy in Vehicular Ad hoc NETworks (VANETs). These Mix-Zones are used by users to change their addresses to reduce location attack risks. In these Mix-Zones a large number of users change

their addresses at the same time, which makes continuous tracking through the correlation between old and new addresses even more difficult. This is because the attackers cannot link users going into the Mix-Zone with users coming out of it, and thus, the mapping between old and new addresses cannot be easily made.

To enter the Mix-zones the device saves the location of the AP and its SSID. When the device is going to send a probe request frame, it first compares its current location to the registered location of the AP. Then, it decides whether it sends the SSID in the probe request or not.

The Mix-zones are characterized by having a high privacy entropy (see Section 3.5 for details on Entropy) to make it difficult for an attacker to link new and previous pseudonyms. For example, to preserve location privacy in the context of location based services, the authors of [12] propose a MAC swapping protocol to allow two mobiles users to swap their MAC addresses to mislead location tracking attacker. Note that the use of Mix-Zones has been increasing recently in particular in the case of vehicular communications [95].

Limitations: this method will not prevent an attacker from identifying user particularly in low density Mix-Zones or when other side information data can be exploited.

3.4.3 Silent Periods

an arbitrary change of pseudonyms might not be optimal for location privacy preservation as it is still possible for an attacker to link changing addresses together through correlations (can exploit mobility patterns of the users), which enables them to continuously track the user, and link new pseudonyms to old ones. It has been shown through experimentation that correlations are difficult to establish when the address change is not performed instantaneously, but after a certain period of silence, known as the *Silent Period* in the literature [67, 80], which is

introduced deliberately after a pseudonym ceases to be in use and before another pseudonym is used. With silent periods in use, each user that wants to change their address, do so after entering a random silent period during which they do not transmit any message. This method prevents the attacker from linking messages to the same device based on the time stamp.

Typically, a number of users are required to enter in silent periods mutually, so that when they resume message transmission with using their newer addresses they can be mixed together. For instance, if there are two or more devices enter and exit the silent period, and thus possess a new pseudonym, the eavesdropper or attacker will not be able to determine to which device the new pseudonym belongs. Therefore, using a silent period is considerably more effective than using a constant time between pseudonyms. It is thus an effective method against tracking.

Note that the concept of silent periods is similar to that of Mix-Zones since silent periods can be viewed as *temporal* Mix-Zones in contrast to other traditional spatial Mix-zones. To increase privacy, silent periods should have certain characteristics such as being random and longer than a certain threshold.

Limitations: The downside of silent periods is that they may affect applications with QoS requirements or those requiring long sessions in addition to requiring a relatively high density for nodes to be able to perform better.

3.4.4 Traffic Manipulation

In addition to the use of silent periods and mix-zones to change pseudonyms, some location attacks can still be carried out by observing traffic patterns, which can be used in some situation to uniquely identify a user, even with the use of sophisticated anonymisation techniques such as the ones described above. To tackle these attacks, traffic manipulation techniques such as those proposed in [113, 158, 161] aim at breaking the pattern of their traffic by injecting additional fictitious packets, with

the goal of making those patterns unrecoverable.

Limitations: although these solutions might operate in certain traffic conditions, it is not clear how the injection of new packets can be envisaged and be efficient in high traffic loads, and whether injected fictitious packets are effective in hiding the signature of the existing traffic pattern.

3.4.5 Hiding the List of Sought SSIDs

The authors in [86] also tried to hide the content of probe request by sending them without the list of the sought SSIDs. The suggested modification are in the pre-connection phase and protects it by using encryption. Although this solution seems to be appealing, it is not clear whether this way would continue to provide fast hand-off to mobile stations which is one of the main reasons why probe requests are used.

With the aim of continuing to offer the benefits of probe request transmissions, i.e. fast hand-off between APs, without revealing the entire list of SSIDs to which the mobile stations associated with in the past, the authors of [73] proposed LAPWiN. LAPWiN which stands for Location Aided Probing for Protecting User Privacy in WiFi Networks, uses the concept of mix-zone to change MAC pseudonyms and restricts the list of SSIDs transmitted to only those for which the APs are in the close vicinity of the mobile station. Thus only those APs are probed actively through the transmission of probe request frames.

Limitations: LAPWiN is proposed under assumption that APs are fixed and not mobile. Although this assumption is common for the most of APs, signal fluctuation might indicate that some APs are out of range while they are still in the vicinity of the mobile station. LAPWiN needs to take into account signal fluctuation to avoid excluding APs that are still in the nearby and detectable. In addition, it does not solve the entire problem which is basically the transmission in

clear of MAC addresses or non optimal way of using pseudonyms.

3.4.6 Encryption of Entire Frames

In [14, 51], the authors proposed to encrypt the entire packet transmitted at the link layer, including the header of the packet (contains the physical address), that it has always been transmitted without encryption in traditional solutions and in the IEEE 802.11 standard. This approach should be used in conjunction with a pseudonym mechanism to prevent tracking.

Limitations: Although this solution seems to protect location privacy, it is not clear how probe request packets that are necessary for the operation of 802.11 can still be broadcast to multiple access points with symmetric cryptography as it requires that mobile stations share cryptographic keys with APs.

Solution	Advantages	Disadvantages
Random MAC	- Protect real MAC Address.	- Design problem (selection, uniqueness, ...)
Mix Zones	-Protect real MAC address. -Prevent the link of previously used pseudonyms to newly used ones.	- Does not protect users in low density Zones.
Silent Period		- Affect applications with QoS requirement (real time applications).
Traffic Manipulation	- Make the pattern unrecoverable.	- Is not efficient in high traffic load.
Encrypt Entire Packet	- Protect MAC header.	- Encrypt Probe Request packets. - Necessity of modification in the standard 802.11.

Table 3.1: Overview of defence methods.

3.5 Quantifying Privacy: The Entropy Metric

In order to assess the effectiveness of the proposed solutions aiming at enhancing privacy preservation, there is a need for a metric that measures the level of privacy preservation of a given solution. The *Entropy* [135] is one of the metrics that is widely used for this purpose.

The concept of *Entropy* was first introduced in Information Theory [135] to quantify information, and measure the expected value of information contained in each message. The Entropy can be used in other contexts to measure other metrics such as disorder, lack of information or ignorance, as well as freedom. It has been used in many disciplines such as physics, biology, economics, etc. [22]. In our context, the Entropy is used to measure the uncertainty an adversary has in attempting to link previously used identifiers to newly used ones.

Hence, the uncertainty of an attacker to link a new pseudonym of an outgoing user to its old identifiers is measured by entropy which can also be seen as the amount of information required to identify the anonymity. Thus, the higher the privacy entropy value, the more attackers will be uncertain of the user location inference, and hence the better privacy protection the system will offer. The Entropy has been extensively used in the literature to measure the effectiveness of location privacy preserving solutions [19, 96, 112, 137, 164].

In order to measure the efficiency of a proposed solution, privacy entropy must be quantified as follows. Given an attacker and a set of all mobile stations U, let λ be the observation of the attacker about the user at some location L.

Given observation λ, the attacker computes a probability distribution P over users $u \in U$. The privacy entropy of this observation λ is [67]:

$$H_\lambda = -\sum P_{u,\lambda} \log_2(P_{u,\lambda}). \tag{3.1}$$

A detailed description on the use of the Entropy to assess the effectiveness of location privacy preservation solution will be presented in Chapter 4.

3.6 Conclusion

Although the location privacy of users in Wireless Networks is a very attractive domain, there are still considerable challenges to be overcome to ensure full privacy of users. In this chapter, we have presented the major problems in WiFi networks and the main countermeasures to cope with them. We have described major existing solutions followed by the strengths and shortcomings of each one of them.

Most of the solutions proposed in the literature and described in this chapter focus on solving the privacy problem without taking into account the effect of the proposed solutions on the quality of service perceived by the users, particularly those which are using real-time applications such as Voice-over-IP (VoIP) or high bandwidth demanding video streaming applications. In the next chapter, we present our first contribution in this thesis, which consists in accurately modeling and quantifying the effect of privacy on the quality of service, and in the next chapter, we present decentralized solution based on our proposed model is able to provide a trade-off between privacy preservation degree and the quality of service perceived by the users.

Chapter 4

On QoS-Aware Location Privacy in Mobile Networks

4.1 Introduction

In this thesis we deal with the threats to user privacy in the context of wireless local networks. We focus on location privacy where an adversary tries to learn a user's past and current locations. The current WiFi standard is vulnerable to location privacy and mobility profiling attacks due to the transmission of personally identifying information such as the MAC address in plain text. Many systems have been proposed to estimate the occupancy in indoor and outdoor spaces and count the number of users [25, 87, 88, 101, 114, 115]. This process is achieved through the capture of the transmitted packets of WiFi users, and locating their position after extracting their MAC addresses that are sent in clear. Such process also reveals the downside of WiFi, where transmitting users MAC address in clear can be considered as a problem and cause some sensitive information leak.

As it has been shown in the previous chapter (Chapter 3), many studies demonstrated that WiFi equipped smart phone or other devices endanger their

owners' privacy. Many studies have also proposed a set of solutions that aim to make it difficult for attackers to track the movements of users.

Such works use temporal addresses (pseudonym) instead of the real MAC address. Cryptographic hash functions are often used to hide the identifiers. However, these techniques are not efficient to protect WiFi users, because it still possible to correlate the used pseudonyms. In order to solve this problem, other research efforts introduced the concept of silent periods between the pseudonym change, during which the user device remains silent and do not transmit any packet. While extensive effort has been made to increase the difficulty of user device location tracking, little has been done on quantifying the effect of these solutions on the quality of service perceived by end users.

In this chapter, we present our first contribution consisting in developing a novel and comprehensive mathematical model and analysis that provide a clear understanding of the relation between privacy entropy, the use of pseudonyms, silent periods, and side information expressed as the mobility pattern, and the quality of service.

We evaluate our proposal with numerical simulation and mobility traces collected from WiFi users in an office environment. Our results show that the introduction of silent periods are very efficient in improving privacy but beyond a certain threshold the extent of improvement is less significant. Based on this result, we elaborated our solutions that seeks to find the best trade-off between privacy and throughput.

The remainder of this chapter is organized as follows. In Section 4.2, we define the entropy metric, and present our proposed mathematical model for WiFi users that allows them to quantify their privacy level using the Entropy metric for different cases. Afterwards, we introduce our proposed algorithm that can be used for each user independently from other users to achieve the privacy with less effect on the quality of service in Section 4.3, and present the results of the evaluation in

Section 4.4. Finally, we conclude in Section 4.5.

4.2 System Model and Privacy Measurement

4.2.1 Privacy Entropy

As shown in Chapter 3, the location privacy is very difficult to measure. However, to evaluate the performance of solutions, the entropy has been one of the most used metrics. In our context, we use the entropy to measure the uncertainty an adversary has in attempting to link previously used pseudonyms to newly used ones. The higher the entropy, the better the solution [43, 93, 137, 164].

In the following, we present our model to calculate the privacy entropy in the basic case where pseudonyms are used without silent periods, the second case where the attacker has mobility pattern information, and the case where silent periods are introduced.

4.2.2 Entropy Calculation without Silent Periods

Let n be the number of users, and let $\mathcal{U}^t = \{u_1^t, u_2^t, \cdots, u_n^t\}$ be the set of pseudonyms used by the n users at time t, and $\mathcal{U}^s$ the set of pseudonyms that were used at time s ($s < t$). We want to calculate the probability $p(u_i^s, u_j^t)$ that a user device has been using pseudonym u_i^s at time s and u_j^t at time t. This probability expresses the linkability of those pseudonyms and is expressed as follows:

$$p(u_i^s, u_j^t) = \begin{cases} 1 & \text{if } u_i^s = u_j^t \\ 0 & \text{if } u_i^s \neq u_j^t \text{ and } u_i^s \in \mathcal{U}^t \\ 1/|\mathcal{U}^s \setminus \mathcal{U}^t| & \text{elsewhere} \end{cases} \tag{4.1}$$

- The first line of Eq. (4.1) reflects the fact the same user device has not changed its pseudonym, therefore the probability is equal to 1.
- The second line reflects the linkability of the new pseudonym to another pseudonym that is already in use, in which case the linkability probability is equal to 0.
- In the third line, we establish the set of suspected pseudonyms which is the set $\mathcal{U}^s \setminus \mathcal{U}^t$ which reflects all pseudonyms that where present at time s and disappeared at time t. Anyone from those pseudonyms could be the one that we are considering at time t. When no side information is known, the probabilities of anyone being the considered pseudonyms are all equal and their values are equal to $1/|\mathcal{U}^s \setminus \mathcal{U}^t|$.

The entropy H related to a given user at a given time, say u_j^t, depends on the current time of observation t and the previous time of observation s and is given by the following expression:

$$H_s(u_j^t) = -\sum_{u_i^s \in \mathcal{U}^s \setminus \mathcal{U}^t} p(u_i^s, u_j^t) \log_2 \Big(p(u_i^s, u_j^t) \Big) \tag{4.2}$$

The entropy measures the weakness of a new pseudonym. If the entropy is low, the new pseudonym is likely to be linked easily to a previous pseudonym. However, when the entropy is high, the new pseudonym assures high privacy for user u_j^t.

4.2.3 Entropy Calculation with Mobility Patterns

In this case, we assume that the attacker has established a mobility pattern model characterizing user mobility behaviors. To establish such a model, we consider that the service area $\mathcal{L}$ is partitioned into m subareas l_i, $\mathcal{L} = \{l_1, l_2, \cdots, l_m\}$.

We assume a Markovian model that captures the probabilities $\pi_{i,j}$ where $1 \leq i, j \leq m$ of movements from l_i to l_j. To calculate those probabilities, the attackers needs to monitor the service area over a long period and logs all time-stamped location information about users.

We assume that users are not changing their pseudonyms (or consider only those users that did not change their pseudonyms). These probabilities depend on time Δt where a user would go after that time. Therefore, we have:

$$\pi_{i,j}^{\Delta t} = \frac{l_{ij}^{\Delta t}}{\sum_{k=1}^{m} l_{kj}^{\Delta t}} \tag{4.3}$$

where $l_{ij}^{\Delta t}$ is the number of users that moved from location l_i to location l_j within a time interval of Δt. For the sake of simplicity, we consider that the interval Δt has the same length as $[s, t)$. Thus we drop Δt from the value of $\pi_{i,j}$. Therefore, the linkability probability of pseudonyms $p(u_i^s, u_j^t)$, where the user $u_i^s \in [s, t)$, becomes as follows:

$$p(u_i^s, u_j^t) = \begin{cases} 1 & \text{if } u_i^s = u_j^t \\ 0 & \text{if } u_i^s \neq u_j^t \text{ and } u_i^s \in \mathcal{U}^t \\ \pi_{ij}/|\mathcal{U}^{\Delta t} \setminus \mathcal{U}^t| & \text{elsewhere} \end{cases} \tag{4.4}$$

The corresponding entropy also depends on Δt and is equal to:

$$H_{\Delta t}(u_j^t) = -\sum_{u_i^s \in \mathcal{U}^{\Delta t}} p(u_i^s, u_j^t) \log_2 \left(p(u_i^s, u_j^t) \right) \tag{4.5}$$

4.2.4 Entropy Calculation with Silent Periods

In this case, we consider that users wait for a silent period SP before changing pseudonyms. We assume that the silent period is chosen uniformly randomly in

the interval $[\mathrm{SP}_{\min}, \mathrm{SP}_{\max}]$. We define the following intervals (see Figure 4.1):

$$\mathcal{S} = (t - \mathrm{SP}_{\max}, t - \mathrm{SP}_{\min}] \tag{4.6}$$

$$\mathcal{T} = (t - \mathrm{SP}_{\min}, t] \tag{4.7}$$

We use the notation $\mathcal{U}^{\mathcal{S}}$ (resp. $\mathcal{U}^{\mathcal{T}}$) to refer to the set of pseudonyms in use during time interval $\mathcal{S}$ (resp. $\mathcal{T}$). Note that the number of pseudonyms k in use during period Δt might be greater than n, $(k \geq n)$ if $\Delta t \geq \mathrm{SP}_{\min}$.

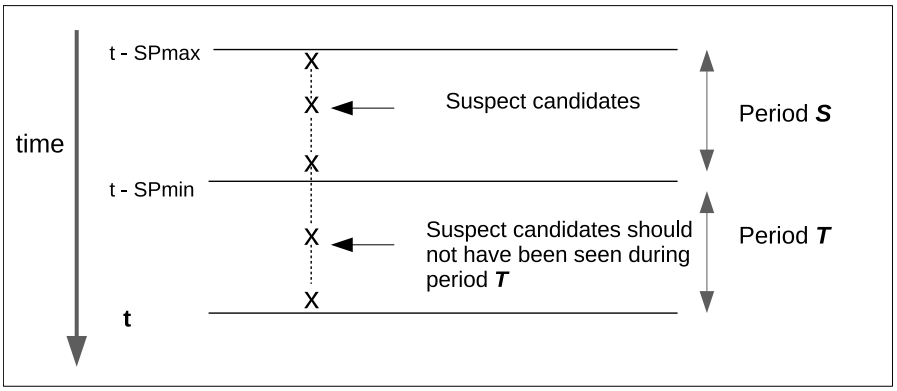

Figure 4.1: Notations used in our mathematical model.

If we consider a user u_j at time t, the list of suspected candidates u_i is equal to all candidates with pseudonyms appearing only in the time interval $\mathcal{S}$ but not in $\mathcal{T}$

The probability $p(u_i^s, u_j^t)$ where user $u_i^s \in \mathcal{S}$ and user $u_j^t \in \mathcal{T}$. This probability can be easily deduced from Eq. (4.1) and Eq. (4.4):

$$p(u_i^s, u_j^t) = \begin{cases} 1 & \text{if } u_i^s = u_j^t \\ 0 & \text{if } u_i^s \neq u_j^t \text{ and } u_i^s \in \mathcal{U}^{\mathcal{T}} \\ \pi_{i,j}/|\mathcal{U}^{\mathcal{S}} \setminus \mathcal{U}^{\mathcal{T}}| & \text{elsewhere} \end{cases} \tag{4.8}$$

- The set $\mathcal{U}^{\mathcal{S}}$ represents all potential candidates that might be suspected to have changed their pseudonyms.
- The $\mathcal{U}^{\mathcal{T}}$ represents the impossible candidates.
- The set $\mathcal{U}^{\mathcal{S}} \setminus \mathcal{U}^{\mathcal{T}}$ represents the set of suspect candidate.

Thus, the corresponding entropy becomes:

$$H_{\mathcal{S}\setminus\mathcal{T}}(u_j^t) = -\sum_{u_i^s \in \mathcal{U}^{\mathcal{S}}\setminus\mathcal{U}^{\mathcal{T}}} p(u_i^s, u_j^t) \log_2 \left(p(u_i^s, u_j^t)\right) \tag{4.9}$$

4.3 QoS-Aware Privacy Preservation

4.3.1 Calculation of Throughput reduction

The introduction of silent periods to preserve a certain level of privacy has an impact on the throughput that will be available to the user. By assuming that the average communication session duration is CS and the average silent period is SP and by using Θ_{priv} (resp. Θ_{nopriv}) to refer to the throughput available to the system when silent periods are (resp. are not) used, the ratio $\Theta_{\text{priv}}/\Theta_{\text{nopriv}}$ is equal to $\text{CS}/(\text{CS} + \text{SP})$. Therefore, the throughput reduction ratio is given by:

$$\frac{\Theta_{\text{nopriv}} - \Theta_{\text{priv}}}{\Theta_{\text{nopriv}}} = 1 - \frac{\text{CS}}{\text{CS} + \text{SP}} \tag{4.10}$$

As shown in Eq. (4.10), silent periods (i.e. the value of SP) need to be minimized to reduce the effect on the throughput available to users. However, reducing the silent periods will reduce the privacy entropy. The effect of silent period length on the privacy entropy is different for each individual user.

Assume that a user device wants to maintain a certain entropy value. The user device needs to find the lowest silent period that achieves the target entropy.

As the entropy depends on global information such as the previous and current location of the other users, the number of users in the area, etc., the user device that wants to find the optimal silent period value needs to rely on a central entity that has this global information. Such a solution has weaknesses and vulnerabilities due to its centralized nature.

4.3.2 Proposed Privacy Preservation Algorithm

In this section, we present our decentralized solution (see Algorithm 1) that aims at estimating global information locally, in a way that allows a user device to estimate the entropy it will get, depending on the silent period used before using another pseudonym.

- Let u_i^s be the user that finished its CS at time s and wants to go into SP for a duration of sp, and reappear at time t ($t = s + sp$).
- Let h be the privacy level desired to be achieved at t.

As shown in Section 4.2.4, the entropy depends on the number N of suspect candidates, which are neighbors to the user that may be linked to u_j^t. As shown in Section 4.2.4, the suspect candidates belong to the set $\mathcal{U}^s \setminus \mathcal{U}^t$. We use u_i, $i \in [1, N]$ to refer to those candidates.

We have $\mathcal{U}^s \setminus \mathcal{U}^t = \{u_1, u_2, \ldots, u_N\}$ and :

$$N = |\mathcal{U}^s \setminus \mathcal{U}^t| \tag{4.11}$$

We use $p_i = \Pr(u_j^t = u_i)$. Therefore, we have:

$$h = H_s(u_j^t) = -\sum_{i=1}^{N} p_i \log_2 p_i \tag{4.12}$$

In the case of no side information, all the probabilities p_i are equal. Thus,

$$p_i = 1/N \tag{4.13}$$

By replacing p_i with its value from Eq. (4.13), we obtain:

$$h = \log_2 N \tag{4.14}$$

Thus, we get N as:

$$N = 2^h \tag{4.15}$$

To maintain a privacy entropy value of h, the user device has to run Algorithm 1. It starts by choosing the desired entropy (h) (see Line 1). After finishing its communication session, the user device enters into a silent period, and captures all appearing and disappearing users in its region, until it gets N as a number of users disappearing (see Lines 1 to 4). Once that number is obtained, the user device can reappear with a new pseudonym.

We defined $sp_{\max}$ as the maximum duration that a user device can spend in the silent period. So, during the silent period, the user device determines at each unit of time the set of pseudonyms ($\mathcal{U}^t$) that appear at the current time (t), and the set ($\mathcal{U}^s$) that appeared at the previous time (s), then, it determines the set $\mathcal{U}^s \setminus \mathcal{U}^t$, which represents the pseudonyms that disappeared from time s to time t to be the suspect candidates (see Lines 11 to 18). After that, it compares the number of suspect candidates to N (see the condition in Line 11). If they are not equal, and the spent time in silent period is less than $sp_{\max}$, the user device has to stay in the silent period (Lines 11 to 18). If is not the case (the duration exceed $sp_{\max}$, or the user device has got N), it cuts the silent period, changes its pseudonym, and starts

Algorithm 1 Find the optimal sp value for a target entropy h

1: ▷ Let h be the desired entropy.
2: ▷ Let s be the time at which the user device finishes its CS and enters into SP.
3: ▷ Let N be the number of users that are needed to enter into SP with the current user device since time s, so the current user device gets the desired privacy level.
4: $N \leftarrow 2^h$
5: ▷ Let $sp_{\max}$ be the maximum time (number of seconds) a user device can spend in SP in case the entropy could not be satisfied within this time interval.
6: ▷ Let *count* be the number of users that entered into SP since the current users entered into SP.
7: $count \leftarrow 0$
8: ▷ Let sp be the number of seconds the current user device spends in SP
9: $sp \leftarrow 0$
10: ▷ Wait until *count* reaches N to obtain an entropy h or sp reaches $sp_{\max}$.
11: **while** $(count < N)$ and $(sp < sp_{\max})$ **do**
12: $t \leftarrow current\_time$
13: $sp \leftarrow sp + (t - s)$
14: Determine the set $\mathcal{U}^t$ as defined in Section 4.2.4.
15: Determine the set $\mathcal{U}^s$ as defined in Section 4.2.4.
16: $count \leftarrow count + |\mathcal{U}^s \setminus \mathcal{U}^t|$
17: $s \leftarrow t$
18: **end while**
19: Generate a new pseudonym.
20: Reappear at current time.

a new communication session (Lines 19 and 20).

4.4 Validation and Performance Evaluation

4.4.1 Methodology

In order to validate the model and evaluate the performance of our method proposed in [104], we run a series of numerical simulations. We considered a fixed number of users moving within a simulation area. For the mobility model, we used the Bonn-Motion tool [15] with a Manhattan Grid model to simulate the movement of

users in the built environment, which includes the movement of users in shopping malls, office environments, etc.

We considered both cases:

1. Without mobility model: an attacker does not have any idea about the mobility pattern of users,

2. With mobility model: the mobility model represented by an offline mobility database, is known to the attackers.

The mobility data generated by the Boon-Motion tool has the following format: `<time, id, loc>`, where:

- `id`: represents a user device unique pseudonym.

- `loc`: represents the user device location with the pseudonym (`id`) at time "`time`".

- `time`: the time.

We generated multiple scenarios and varied the number of users (from 50 to 150 users) to reflect diverse user density configurations. We ran simulations for a duration of 5 hours in an area of 300m×300m where the average speed of users is 6mph (approximately 10m/h). And, we divided the simulation area into 9 sub-areas arranged in a 3x3 grid. Each sub-area represents a geographic location. Simulation parameters are summarized in Table 4.1.

In our simulation we considered studying the effect of the SP and the CS lengths on the privacy entropy of the system and the throughput that can be achieved according to the desired entropy.

Parameter	Value
Number of Users in the Area	50-150 nodes
Duration of the simulation	5 hours
Area	300m x 300m
Number of sub-area	3X3 (9 sub-areas)
Node average speed	6mph

Table 4.1: The parameters of our simulation.

4.4.1.1 Effect of the CS Length on the Privacy Entropy

In the first set of simulations, we fix the SP length to 120 seconds and vary the CS from 10 to 40 minutes to reflect different applications. Each user device keeps using the same pseudonym during the same session.

- When the communication session ends, the user device waits for a silent period of 120 seconds before deciding to start another communication session.
- When the user device starts another session, it does so with another pseudonym picked randomly.

Figure 4.2 and Figure 4.3 show the variation of the entropy of the system, taken as the average on all users' individual entropies, in function of the CS length in the cases when prior mobility information is known or not, as shown in figures Figure 4.2 and Figure 4.3, respectively.

As it is expected, the entropy of the system decreases when the communication session increases, when the user density decreases, or when prior mobility information is known. Figure 4.2 and Figure 4.3 show that the effect of the user density or the duration of the CS is less strong on the entropy than that of the prior mobility information knowledge.

We conclude that obfuscating prior mobility information is more important than reducing the CS length or waiting until the user density becomes higher. However,

obfuscating prior mobility information can be challenging in some configurations when global mobility patterns can be clearly distinguished.

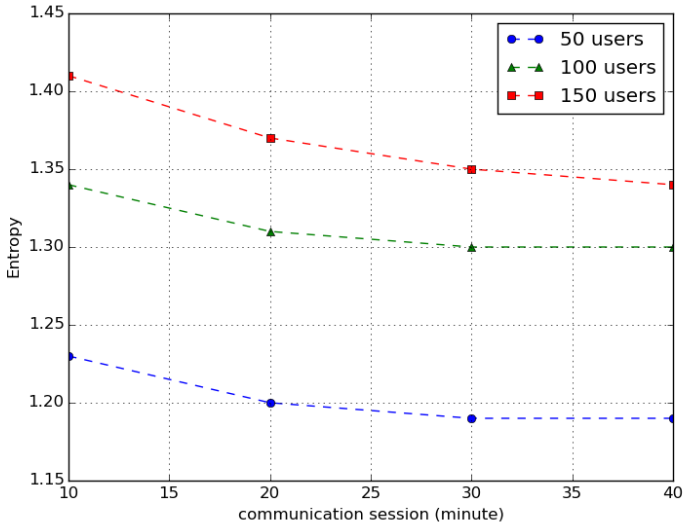

Figure 4.2: Impact of CS duration on the privacy entropy With side information (SP = 120 seconds).

4.4.1.2 Effect of the SP Length on the Privacy Entropy

In another set of simulations, we set the CS to 10 minutes and vary the SP from 0 to 120 seconds. To simulate a variable length SP, we make every user device uniformly pick a random number in the interval $[\mathrm{SP}_{\min}, \mathrm{SP}_{\max}]$. To efficiently calculate the entropy associated with every new pseudonym, we continuously observe users. Each time we detect a change in the list of pseudonyms, say at time t, we start by finding the list of *suspects* which are the possible previous pseudonyms that can be linked to the new pseudonym.

The list of suspects contains all the users that were present in the interval $[t - \mathrm{SP}_{\max}, t - \mathrm{SP}_{\min}]$ and disappeared after that.

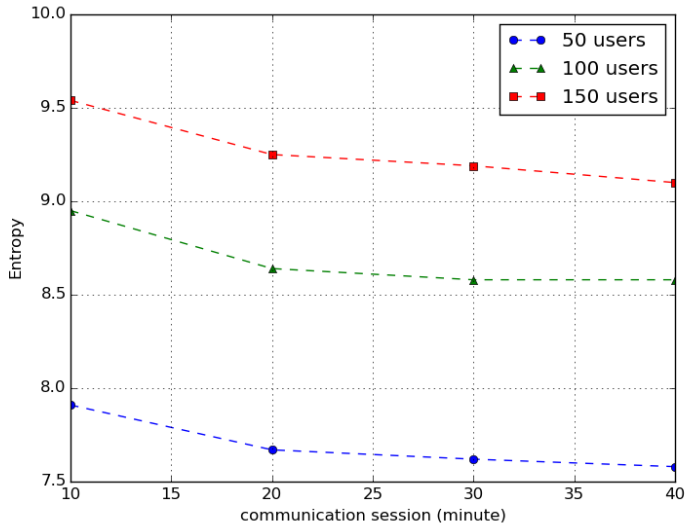

Figure 4.3: Impact of CS duration on the privacy entropy Without side information (SP = 120 seconds).

1. **No Prior Mobility Information Available**

   In the case where no side information is known (e.g. no mobility model is available), all the suspects are equivalent and have the same probability to be linked to the new pseudonym.

   The entropy we calculate is equal to the one expressed in equation (4.2). By putting $\Delta t = [t - \mathrm{SP}_{\max}, t - \mathrm{SP}_{\min}]$, the entropy we calculate is equivalent to $\mathcal{H}_{\Delta t}(\mathcal{U}^t)$.

   In the first experiment we fixed the CS to 10 minutes and change the SP from 0 to 120 seconds with the changing of the number of users. Then we calculate the entropy for each occurrence of a new pseudonym.

   The results are illustrated in Figure 4.4 which shows the effect of SPs on the entropy value when no side information is available. When the SP is equal to zero it means the nodes did not use silent periods which is the case described in Section 4.2.2.

We show that the increasing of the SP length increases the entropy which reaches a maximum value of 10.2 (resp. 9.8, 8.8) with 150 (resp. 100, 50) users when SP is 120 seconds. This is because the increase of SP gives more opportunity to other users to enter silent periods during the same period where suspects are collected, which increases the number of suspect candidates thereby making it difficult to link the new appearing pseudonyms to the correct user. Figure 4.4 also shows the impact of user density on the privacy preservation level measured according to the entropy metric. We show that the entropy increases with the increase of the number of users. This is because when the user density is higher, an attacker cannot easily identify and locate a user device when they are frequently changing their pseudonyms.

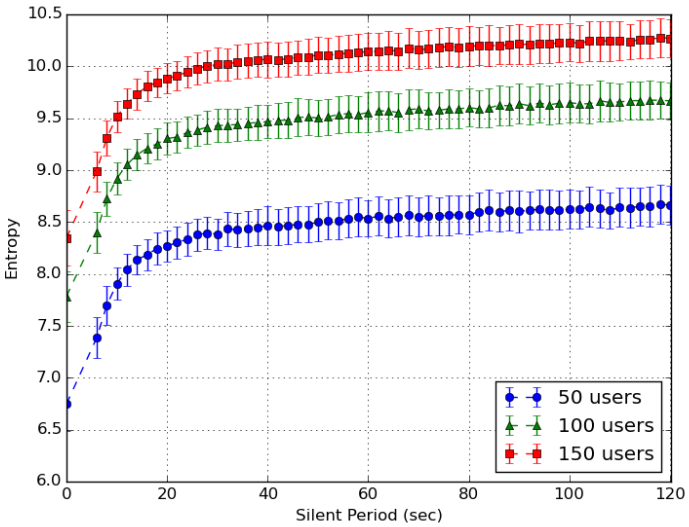

Figure 4.4: Impact of SP duration on the privacy entropy, Without side information (CS = 10 min).

2. **Prior Mobility Information Available**

To assess the effect of user density and SP variations on privacy entropy in the context where mobility data has been collected in advance and mobility profiles have been established for users, we generated multiple scenarios and used part of the data we generated as the offline data to simulate the prior mobility knowledge and the other part as the test data. The test data will be used during the online stage based on the models established during the offline stage. We repeated the same experiment as described above. We set the CS length to 10 minutes and we varied the SP length and the number of users.

The obtained results are illustrated in Figure 4.5 which equally shows the increase of the entropy values with the increase of the SPs. We also show the effect of the number of users on the entropy compared to that of the SP length, when the SP is 120 seconds the entropy is 1.3, (resp. 1.4, 1.47) for a number of users of 50 (resp. 100, 150) which means that the entropy is also influenced by the number of users.

In Figure 4.4 and Figure 4.5, we show that moving from a situation where no SP is used (SP = 0) to another situation where SPs are used increases the entropy significantly. However, beyond a certain *threshold* (20 seconds in our simulations) for the SP length, the amount of the gain obtained for the entropy does not increase with the same intensity and it is not as significant as for the first SP lengths.

4.4.1.3 Effect of the SP length on the Quality of Service (QoS)

The introduction of SP is efficient in increasing the privacy of users; however, it affects the QoS perceived by applications particularly those with strong delay requirements and those with long communication sessions. For applications with

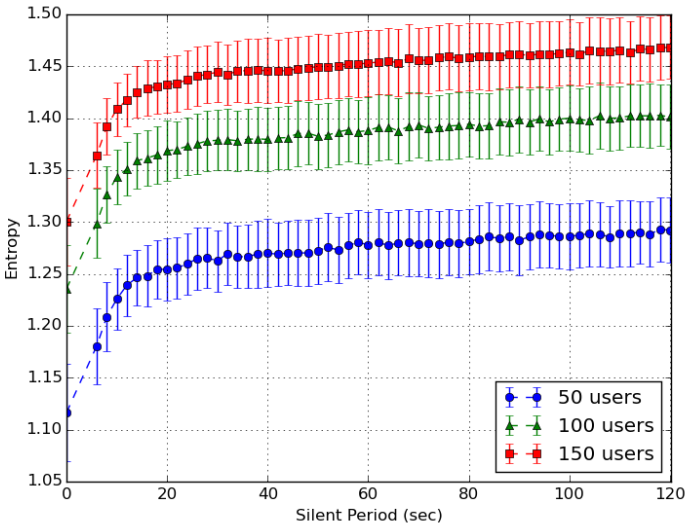

Figure 4.5: Impact of SP duration on the privacy entropy, With side information (CS = 10 min)

long CSs, the effect introduced by the use of SPs decreases with the increase of the length of CSs as shown in Section 4.4.1.1. Therefore, it is required for applications to choose the appropriate SP length that provides a good trade-off between privacy and QoS. The effect of the SP length is particularly severe for applications with short CSs as the throughput reduction will be higher.

To show the relation between the privacy entropy and the available throughput, we plot Figure 4.6 and Figure 4.7. The figures show the ratio of the throughput that becomes available for a targeted privacy entropy value. Both Figure 4.6 and Figure 4.7 show that after a certain *threshold* value, it costs a significant amount of throughput reduction to reach higher privacy entropy values. Therefore, a good trade-off between location privacy and the quality of service should be made.

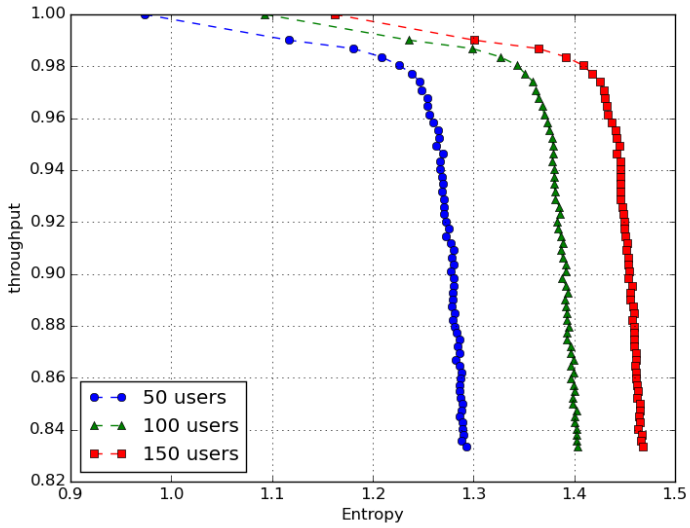

Figure 4.6: Available throughput percentage in function of required entropy, With side information.

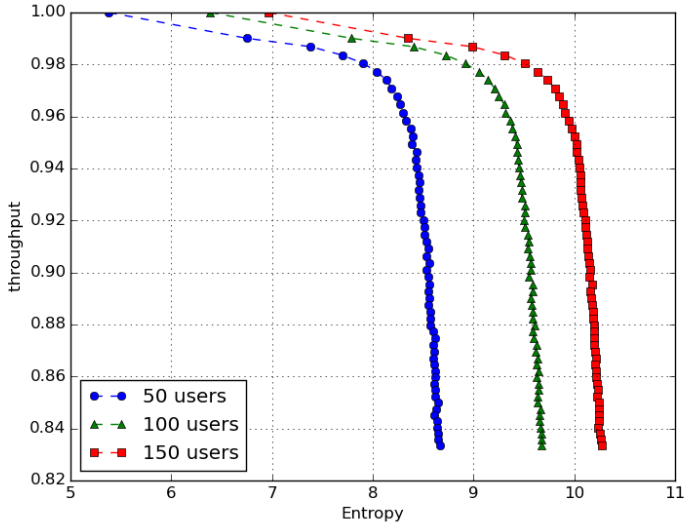

Figure 4.7: Available throughput percentage in function of required entropy, Without side information

4.4.2 Validation

4.4.2.1 QoS-aware Privacy Entropy Improvement

Our proposed solution makes it possible for each user device to achieve the privacy entropy they want with a minimum effect on the QoS they perceive characterized

by the available throughput to them. The solution is decentralized and thus relies on local information available to the user only, i.e. without the need of a central entity that tells the user which parameters to use. The aim of reaching a certain entropy is to prevent an attacker to link previous pseudonyms to the new one used by the user.

We showed that we mainly achieve higher entropy values by using a silent period which we need to minimize so that we do not affect the throughput available to the user.

To evaluate our solution that is based on adaptive changing of the SP length for each user device to achieve a target entropy with minimum throughput reduction, we compare its performance against traditional solutions that rely on a central entity that is based on calculating the global entropy of the system and providing each user device with the SP value it should use to maximize the entropy of the system globally. The main weakness of using the same interval for SP values for the entire system, i.e. the same interval $[\mathrm{SP}_{\min}, \mathrm{SP}_{\max}]$ for all users, is that an attacker can guess the values of both $\mathrm{SP}_{\min}$ and $\mathrm{SP}_{\max}$ and thus can construct the set of suspect candidates in an easier way than the case where all users have different SP lengths.

Indeed, when each user device chooses their silent period independently of the others, an adversary can not guess the silent period that the user device chose because it is related to the number of its neighbors and the desired privacy level. Therefore, the interval of suspect candidates will increase and include more suspect users which increases the entropy value. To assess the difference between our method and the one that uses a constant SP for all users, we run simulations based on the same mobility database generated in this section (Section 4.4.1) and set the CS length to 10 minutes and the number of users in the simulations to 150 users.

The obtained results show that the entropy measured at an arbitrary user is

significantly higher with our method due to the use of independent and adaptive SP value for each user.

4.4.2.2 Experimentation with Real WiFi Traces

In this section, we rely on real WiFi traces to evaluate the effect of using our algorithm on the privacy level that could be achieved and the associated costs in terms of the mean power consumed and the throughput available. In this set of experiments we developed a data collection algorithm that captures WiFi packets transmitted in a Lab Office environment and keeps those packets whose MAC addresses correspond to hand-held devices (smart phones, tablets) based on the manufacturer of their WiFi chips obtained from their MAC OUI list [62]. In our evaluations, we considered two scenarios.

- **First scenario:** In this scenario the user device does not use any elaborate technique to calculate the optimal SP value it should use. In this case, the user device detects the end of CSs based on the inter-packet transmission time, a longer time means the user device finished a CS and entered another one after that. Most of the current hand-held devices use an aggressive power-save mode where WiFi chips are switched to sleep mode automatically when the device is not in active use or there are no pending transmissions. During this mode, we consider the device as in a natural SP. Once the device is activated again or there is data to be transmitted, the device terminates the power-save mode and starts a new CS. We emulate pseudonym change and we consider that the user device appeared with a different MAC address.

- **Second scenario:** In this scenario the device does not automatically start a new CS once activated again or when new packets are to be transmitted. Instead, we emulate adaptive SPs according to Algorithm 1 where the current

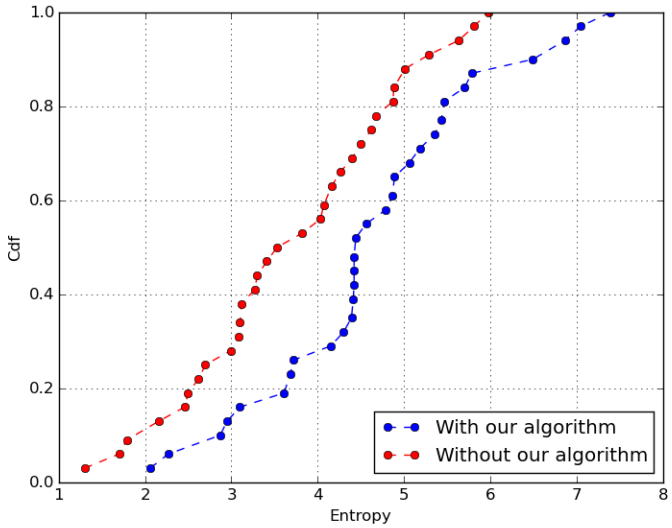

Figure 4.8: The CDF of the entropy with and without decentralized SP length computation and pseudonym change.

user relies on keeping track of the other users that are within the same area and the times when they entered into their corresponding SPs. Based on this information, the user device can decide if the natural SP needs to be extended so that privacy entropy reaches the threshold expected to offer a better privacy, or the device can restart a new CS without extending the natural SP in the case the current measured entropy satisfies the required privacy threshold.

In our experiments, we consider that the data is collected from WiFi devices whose chips can deliver a bandwidth of 54Mbps and whose active (resp. sleep) mode mean power consumption is 434mW (resp. 7mW) [24, 141]. We consider the maximum value for SPs to be 60 seconds as the maximum time a user with pending packets to be transmitted can wait with the hope to reach the desired privacy level. Even if the privacy threshold can not be attained, the user device restarts a new CS because it cannot keep waiting indefinitely. In the case there

are no new packets pending for transmission, the SP can naturally be longer. We chose a value of 60 seconds because the simulations we run in the previous section showed that beyond that value the entropy does not improve significantly.

The obtained results are plotted in Figure 4.8, Figure 4.9, and Figure 4.10.

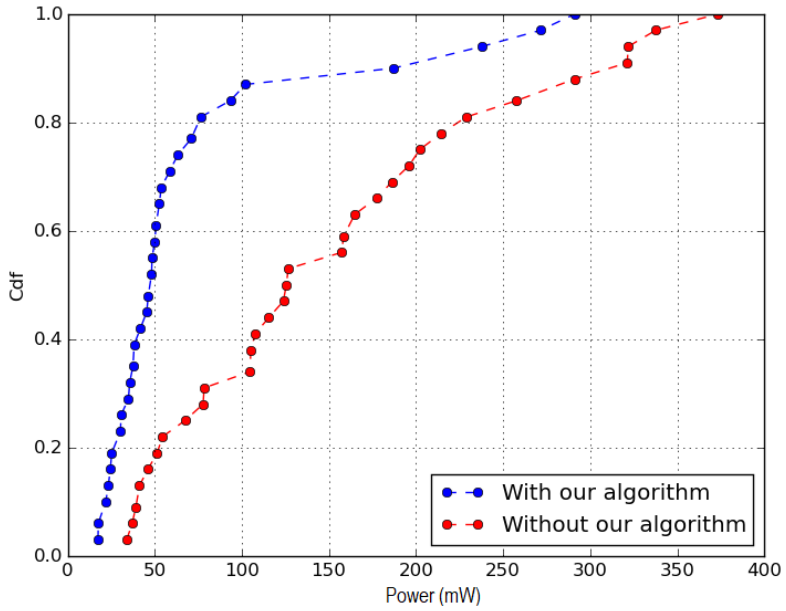

Figure 4.9: The CDF of the mean power consumption with and without decentralized SP length computation and pseudonym change.

In Figure 4.8, we plot CDF of the entropy observed over all users achieved in both scenarios when:

1. The SP length calculation and pseudonym change are done in a decentralized way as described in our Algorithm.

2. The pseudonym change is performed after each natural SP.

As we can see in this figure, over 60% of the users have an entropy value of around 5 or higher with our algorithm compared with the traditional cases where silent periods are only natural, in which case the entropy values are smaller than 4.

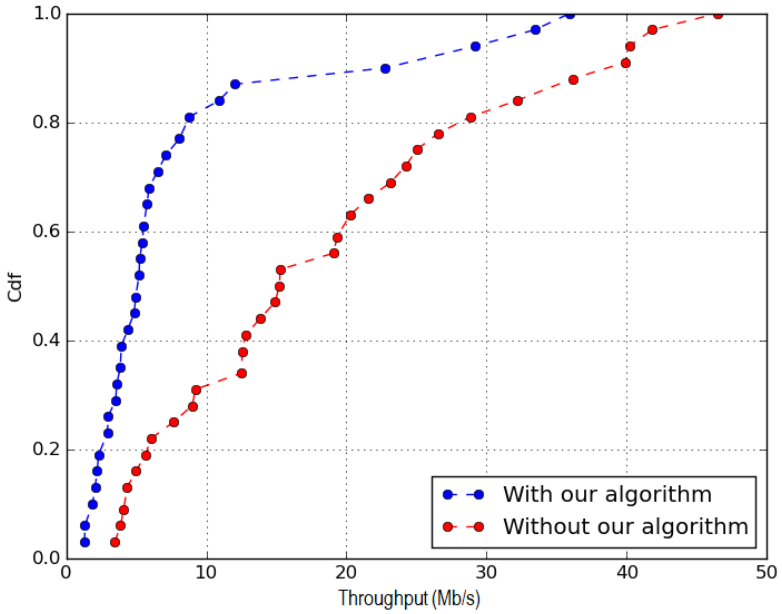

Figure 4.10: The CDF of the throughput with and without decentralized SP length computation and pseudonym change.

This confirms our findings that using dynamic and user-specific SPs to change the pseudonym result in a higher entropy compared to changing pseudonym after each natural SP.

In Figure 4.9 and Figure 4.10, we plot the CDF of the mean power consumption and the throughput available to users to show the effect of our decentralized SP calculation on those parameters. We show in Figure 4.9 that our method consumes less power compared to when natural SPs are used. This is because in some cases, the natural SP length is not sufficient to reach the target privacy entropy in which case the extra time added to the natural SP allows the devices to stay longer in sleep time and thus save more power. For example, about 60% of the users consume 50mW on average whereas when no dynamic SP is used about the same percentage of nodes consume around 150mW.

Figure 4.10, show that the throughput achieved by the use of longer SPs is smaller than the one obtained with natural SPs. This is the cost to be paid to

achieve higher user privacy while preserving a certain quality of service. It was possible to achieve higher privacy by increasing the SP length but that would not meet QoS constraints of applications.

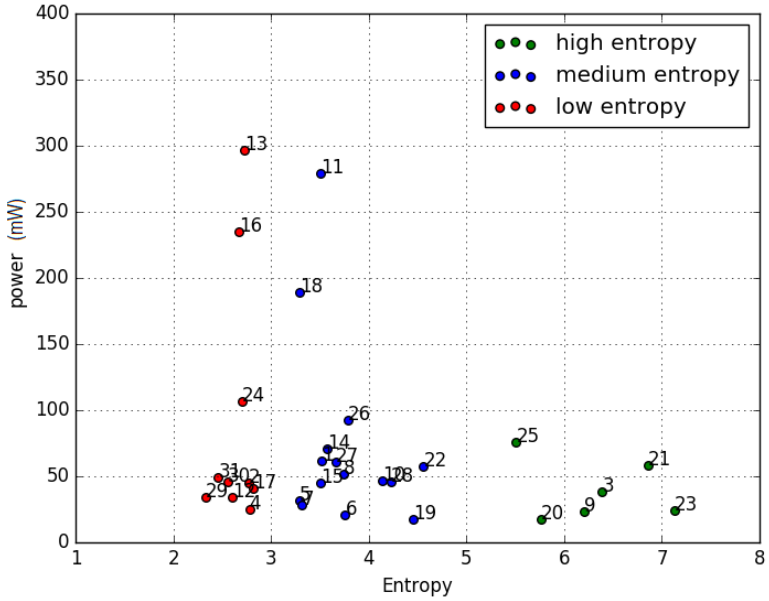

Figure 4.11: Classification of users according to their privacy entropy circumstances (power).

Note that the specific relation between entropy, mean power consumption, and throughput is different for each user, because there are other parameters that determine how these affect each other, such as the location and the communication pattern of the neighboring users. Some users can be in configurations where the location they are in and their neighbors communication pattern allow them to reach very high levels on privacy even with small SP lengths whilst other users in other configurations may need to use very long SPs to only slightly increase their privacy levels a bit. This is shown in Figure 4.11 and Figure 4.12 where we classified users into three classes: high, medium, and low entropy classes which we obtained by running a $k - means$ algorithm to cluster users according to their

entropy value.

Therefore, users in a low entropy class will need higher SP lengths which would degrade their performance in term of throughput. However, users in a high entropy class can achieve high levels of privacy entropy with much smaller effects on the throughput available to them.

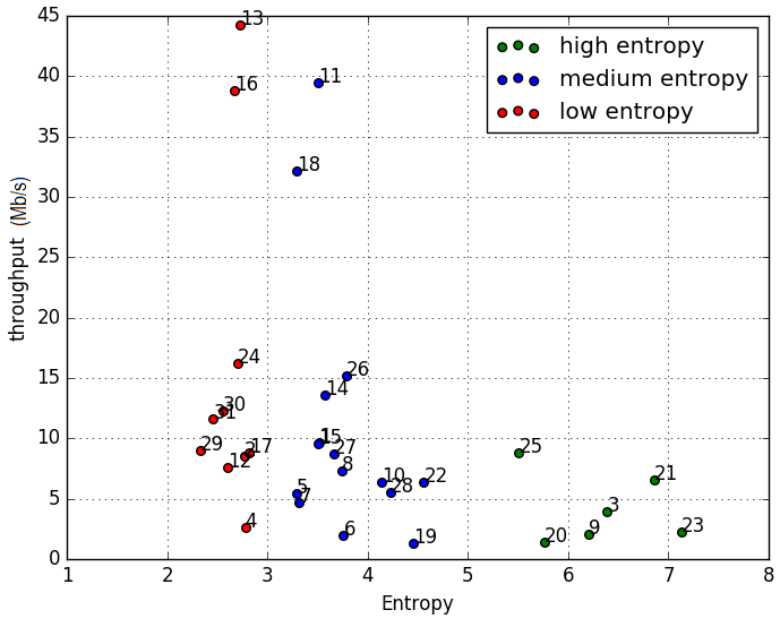

Figure 4.12: Classification of users according to their privacy entropy circumstances (throughput).

4.5 Conclusion

In this chapter we presented the privacy metric which is commonly used to measure the privacy preservation of a given solution. In addition, we proposed a decentralized solution for location privacy in WiFi networks, in which each user device makes a decision on which strategy to use to achieve their objective in terms of privacy and quality of service. We have provided a generic mathematical model that provides a

clear understanding of the relation between privacy entropy, the use of pseudonyms, silent periods, and side information expressed as the mobility pattern. However, it allows users to attain their desired levels of privacy while lowering its effect on the QoS perceived by them.

We evaluated our proposal with numerical simulations and mobility traces collected from WiFi users in an office environment, and the results showed that the introduction of silent periods are very efficient in improving the location privacy but beyond a certain threshold the extent of improvement is less significant.

In the next chapter (Chapter 5), we will present our second contribution, which is an enhanced silent period-based solution that aims at maximizing the privacy of users while ensuring the quality of service required by the applications they use. It allows each user device to estimate the most suitable silent periods and communication sessions for a best trade-off between privacy and QoS for a typical video streaming application (e.g. YouTube).

Chapter 5

Enhancing Location Privacy and QoS for Video Streaming Over WLANs

5.1 Introduction

With the democratization of smart phones and mobile broadband services, the usage of applications requiring a guaranteed QoS such as VoIP, offline streaming, or IPTV and live streaming have seen a big surge. In multimedia services it is very important to offer better quality of service to users compared to other best effort services. However, as shown in the previous chapter, ensuring a good location privacy in WiFi networks may require the introduction of silent periods which would affect the perceived QoS by end users.

It has been particularly shown that a strong privacy depends on the user density during the silent period. In the case of low user density, users may need to use longer silent periods to make sure the density of users increases during the silent period. Solutions based on this approach are gaining success and have been adopted by

the industry (Apple's solution [149]). The authors in [67] show that silent periods should be chosen randomly, because the attacker can correlates two pseudonyms that used exactly the same silent period.

While these solutions have proven efficient in privacy preservation for many applications such as e-mailing, and web surfing [67], their effect on real-time and audio/video streaming applications has not been assessed thoroughly in the literature. In fact, the introduction of intentional silent periods will increase the privacy of users but will also affect both delay and bandwidth, which would hamper the well functioning of these applications as show in the previous chapter (Chapter 4).

In this chapter, we study the possibility of finding a trade-off between QoS and privacy, which can be achieved by choosing appropriate values for silent periods and communication sessions. Therefore, to find the best values for silent periods, we propose a decentralized solution that allows each user device to estimate the most suitable silent periods and communication sessions for a best trade-off between privacy and QoS for a typical video streaming application (e.g. YouTube).

Our solution is based on the exploitation of our elaborate mathematical model quantifying the privacy and QoS, and designing a decentralized algorithm which allows users to select adequate silent and communication periods, depending on the network parameters. We evaluated our solution on a set of different configurations based on numerical simulations. The obtained results demonstrated the efficiency of our approach.

The remainder of this chapter is organized as follows. In Section 5.2, we present our proposed mathematical model, and algorithm for privacy and QoS preservation, and in Section 5.3, we present the validation and performance evaluation of our solution. Finally, in Section 5.4, we conclude and highlight the benefits obtained from our proposed solution.

5.2 Proposed Model for Privacy and QoS

Our approach is based on the frequent change of MAC addresses (pseudonyms), and the use of silent periods after each pseudonym change. However, as aforementioned in the introduction, the use of silent periods usually decreases the QoS perceived particularly audio/video streaming ones. In this section, we present our enhanced silent-period-based solution that aims at maximizing the privacy of users while ensuring the quality of service required by the applications in use. This solution is based on our mathematical model, presented in the previous chapter, which allows to quantify the period that the user device has to stay in the silence period and in the active session while ensuring both privacy and the QoS.

5.2.1 System Model

Figure 5.1 depicts the notation and the different symbols used in our mathematical model. We consider T_{cs} as the duration of a communication session and T_{sp} as the duration of a silent period. The T_{spBuffer} represents the time duration where the user can consume data during a silent period without any degradation on the QoS. In our model, for a given user and depending on various parameters (e.g. bandwidth, user density, privacy), we can have different periods for the communication sessions (T_{cs1}, T_{cs2}, ...), and different silent periods (T_{sp1}, T_{sp2}, ...), as well as different periods for the T_{spBuffer} ($T_{\text{spBuffer1}}$, $T_{\text{spBuffer2}}$, ...).

The goal of this work is achieving the desired privacy and make the duration of T_{spBuffer} larger than T_{sp} to guarantee that users continue to play video content without interruption while still preserving their privacy.

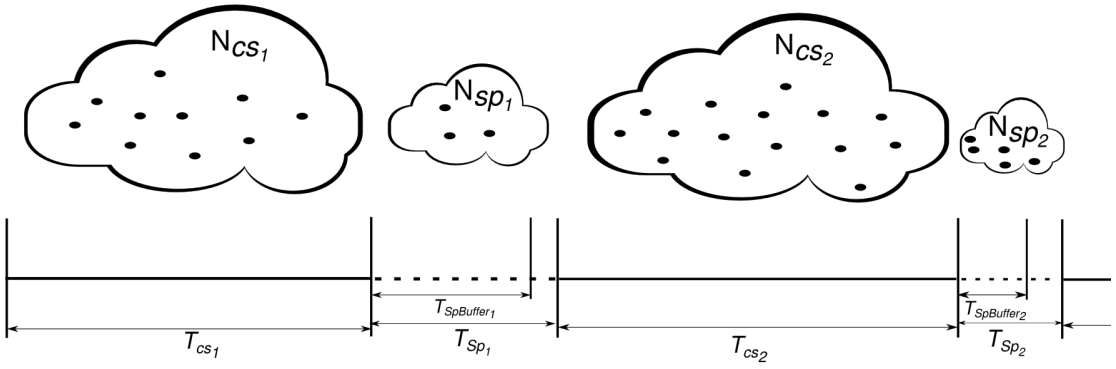

Figure 5.1: Notation and symbols used in the mathematical modelling.

5.2.2 Privacy and QoS Model : Case of Video Streaming

As we have explained in the previous chapter (Chapter 3), the entropy is one of the most used metrics to measure the location privacy. It measures the uncertainty an adversary has in attempting to link previously used pseudonyms to newly used ones. It is related to the number of users that can not be distinguished. Therefore, the higher the entropy, the better the privacy protection the system offers [137].

In the rest of the chapter, we make the following assumptions :

- assume that a user device wants to get a privacy level of an entropy H. To reach this level of privacy, the user device needs to make sure that it has at least N_{sp} neighbors with which they can mix. As shown in [104], the value of N_{sp} should be equal to:

$$N_{\text{sp}} = 2^H \tag{5.1}$$

  To show the effect of the privacy needed by a user on the QoS, we take as example video streaming applications such as YouTube. This kind of applications enable the playback of the video before the content downloaded is completely finished. Depending on the content and quality of the video that

is chosen by the user, different data bitrates can be used by these applications.

Location privacy depends on the number of users in the monitored area. The more users there are, the better the privacy will be, but the worse the QoS they will experience, as these users will have to compete for the same medium and the available bandwidth needs to be divided between all network users. This of course can introduce interruptions in the playback (or in the downloading) of the video especially if the user requires a high-quality video.

In order to provide the highest possible quality (limited interruptions, and high video quality) of playback, we propose with our model a way to quantify the bandwidth, the bitrate, and period in which the user can enjoy playing the video without interruption.

- assume that there are N active users in a given region and that the overall available bandwidth to all users within that region is W.

- assume that the available bandwidth to each user W_u decreases when the number of neighboring users increases and that decrease in bandwidth is more pronounced when there are more users in the same area due to the effect of higher contention for the channel. During the communication session, the user downloads data with the available bandwidth W_u. If the available bandwidth W_u is higher than what is needed to play the video content without interruption, then the extra available bandwidth will be used to accumulate data in the buffer so it can be played during the silent periods.

We use D (resp. C) to refer to the amount of downloaded (resp. consumed) data during time interval T:

$$D = W_u \times T \tag{5.2}$$

$$C = \rho \times T \tag{5.3}$$

where ρ is the bit-rate used to play the video content. Using Eq. (5.2) and Eq. (5.3), we can calculate A which is the data amount of video content that can be accumulated in the buffer in the case the available bandwidth is higher than the bit-rate (i.e. when $W_u \geq \rho$). We have:

$$A = D - C \tag{5.4}$$

The bitrate ρ can be provided for some technologies such as YouTube[1] and could be calculated using *Kush Gauge's* formula[2]:

$$\rho = F \times P \times M \times K \tag{5.5}$$

where: F is the number of displayed frames per second, P is the frame size measured by the number of pixels, reflecting video quality. M is the motion rank.K is the codec constant, When a user device enters a silent period with data already accumulated in the buffer (i.e. $A > 0$), it can continue playing video content without interruption during the silent period for a time duration T_{spBuffer}. We have:

$$T_{\text{spBuffer}} = \frac{A}{\rho} \tag{5.6}$$

By combining Eq. (5.2), Eq. (5.3) and Eq. (5.4) with Eq. (5.6), we get:

$$T_{\text{spBuffer}} = \left(\frac{W_u}{\rho} - 1\right) \times T \tag{5.7}$$

As we can see in Eq. (5.7), reducing the value of ρ, i.e. decreasing the video quality

[1]see https://support.google.com/youtube/answer/2853702?hl=en
[2]https://quadrophone.github.io/kush-gauge/

consumed by the user, allows the value of T_{spBuffer} to grow, which allows the user to continue playing video content for longer periods while waiting for more users to enter into silent periods. When the number of users in silent periods reaches the required value, users can leave the silent period and resume communication with new pseudonyms thereby they continue playing videos without affecting their privacy.

5.2.3 QoS and Privacy Preservation Algorithm

In this section, we present our decentralized algorithm (Algorithm 2) which allows each user device to maintain a given privacy while reducing the effect on the perceived QoS of the video content played. We characterize the QoS by the number and duration of interruptions occurring, as well as the degradation in video quality (e.g. from 720p to 360p).

5.2.3.1 Communication Session (CS)

During the communication session, the user device calculates at each unit of time the following parameters:

- The number of users N_{cs} required to reach the entropy h.
- The available bandwidth W_u.
- The downloaded D, consumed C and accumulated A data volumes (see lines 6 to 26).

In this period the user can simultaneously download data (with rate W_u) and consume data (with rate ρ). When the W_u is greater than ρ, data downloading is faster than its consumption. Thus, the downloaded data will be accumulated in

the buffer A (see line 16), so that it can be consumed later in next sessions (silent periods and communication sessions) (see lines 21 and 35).

The user device interrupts a communication session (this represents T_{cs}) and enter into a silent period if one of the following conditions occurs (in Line 6):

- The user device exceeds the maximum duration of a communication session (T_{csMax}).
- The number of connected users (N_{cs}) is greater than or equal to the needed number to get the privacy (reaches N_{sp}) and the $T_{\text{spBuffer}} \geq T_{\text{spBufferMin}}$, where, $T_{\text{spBufferMin}}$ represents the minimum possible duration that allows the user device to change its pseudonym in the silent period without any interruption (in our experimentation, this corresponds to 5 seconds). It is considered as a transition period between using new and old pseudonyms. We argue that during this short period, the probability that N_{sp} decreases is minimal.
- The calculated T_{spBuffer} is greater than or equal to maximum duration that the user device can stay in the silent period (T_{spMax}).

In all the cases, the user device has to stay a minimum duration in the communication session corresponding to T_{csMin} (defined in the experimentation).

Algorithm 2 Maximize perceived QoS for a given privacy entropy H and a required video quality q.

1: ▷ Algo is run when the user device switches from SP to CS. D, C, A and T_{spBuffer} are set to 0 when the video starts.
2: Generate a new pseudonym.
3: $\rho \leftarrow$ getBitRate(q)
4: videoSize $\leftarrow$ getVideoSize(q)
5: $N_{\text{sp}} = 2^H$;$N_{\text{cs}} \leftarrow 0$; $T_{\text{cs}} \leftarrow 0$;$T_{\text{sp}} \leftarrow 0$;

6: **while** (($T_{cs} < T_{csMax}$) and (($N_{cs} < N_{sp}$) or ($T_{spBuffer} < T_{spBufferMin}$) or ($T_{cs} < T_{csMin}$)) and (($T_{spBuffer} < T_{spMax}$) or ($T_{cs} < T_{csMin}$))) **do**
7: $N_{cs} \leftarrow$ getNbrOfConnectedUsers()
8: $W_u \leftarrow$ calculateWu(N_{cs})
9: $T_{cs} \leftarrow$ getSpentTimeInCs()
10: **if** ($W_u > \rho$) **then**
11: **if** ($D <$ videoSize) **then**
12: $D \leftarrow D + (W_u \times T)$
13: **end if**
14: **if** ($D > C$) **then**
15: $C \leftarrow C + (\rho \times T)$
16: $A \leftarrow D - C$
17: **end if**
18: **else**
19: **if** ($A \geqslant (\rho \times T)$) **then**
20: $C \leftarrow C + (\rho \times T)$
21: $A \leftarrow A - (\rho \times T)$
22: **end if**
23: **end if**
24: $T_{spBuffer} \leftarrow \frac{A}{\rho}$
25: wait(T) ▷ Wait for some time T before looping.
26: **end while**
27: ▷ Here the user device is switching from CS to SP
28: $T_{spBuffer} \leftarrow \frac{A}{\rho}$
29: count $\leftarrow 0$ ▷ counting the number of users during SP
30: **while** ((count $< N_{sp}$) and (($T_{spBuffer} \leqslant T_{spMax}$ and $T_{sp} < T_{spMax}$) or ($T_{spBuffer} > T_{spMax}$ and $T_{sp} < T_{spBuffer}$))) **do**
31: count $\leftarrow$ getNbrOfConnectedUsers()
32: $T_{sp} \leftarrow$ getSpentTimeInSp()
33: **if** ($A \geqslant (\rho \times T)$) **then**
34: $C \leftarrow C + (\rho \times T)$
35: $A \leftarrow A - (\rho \times T)$
36: **end if**
37: **if** ($T_{spBuffer} < T_{spMax}$) **then**
38: $q \leftarrow$ degradingVideoQuality(A)
39: $\rho \leftarrow$ getBitRate(q)
40: **end if**
41: wait(T) ▷ Wait for some time T before looping.
42: **end while**

5.2.3.2 Silent Period (SP)

During the silent period (SP), the user device calculates T_{spBuffer} (see Algorithm 2 Line 28). During this period (see Algorithm 2 Lines 30 to 42), at each interval of time, the user consumes data from buffer A and updates the number of connected users (count). The user device interrupts its silent period SP and starts a new CS with new pseudonym when the number of connected users count is greater than N_{sp} (count $\geq N_{\text{sp}}$) (reaching the privacy) or this user device has exceeded the maximum duration of silent period (T_{spMax}=60s) [104], where the entropy will not be much improved after this duration.

In fact, we distinguish the following different cases:

- $T_{\text{spBuffer}} \geq T_{\text{spMax}}$: the user device can stay in SP (see Algorithm 2 Line 30) without any interruption in playing the video and hoping to reach the privacy (count=N_{sp}).

- $T_{\text{spBuffer}} < T_{\text{spMax}}$: this means we have an interruption in the playback which degrades the QoS.

- $T_{\text{sp}} \leq T_{\text{spBuffer}}$: this means that the user device has reached the desired entropy (reach N_{sp}) so it can interrupt its SP before completely consuming T_{spBuffer}, and enters in a new communication session (cs_{new}) without any interruption in consuming data from buffer. This is considered as the ideal scenario for a user (getting the desired entropy and the QoS). As consequence to this scenario, the accumulated data (which is not yet consumed during T_{sp}) should be consumed during this new *cs* (or the next *sp*). This scenario allows also to increase the value of T_{spBuffer} during the next silent period.

- In the case of $T_{\text{sp}} > T_{\text{spBuffer}}$: the user did not reached the desired entropy during T_{spBuffer}, so, it needs to stay in the silent period after the T_{spBuffer} until

getting the desired entropy or reaching T_{spMax}. But, during this period (in the interval $[T_{\text{spBuffer}}, T_{\text{sp}}]$), the user can not consume the data in the buffer (it is empty), which causes an `interruption` in the playback and so degrading the QoS.

The idea behind this algorithm is:

- The use of our mathematical model to quantify the T_{spBuffer} during and after each communication session.
- Have a dynamic T_{cs} and dynamic T_{sp} that depends on the desired privacy and the calculated T_{spBuffer}.
- Avoid any interruption by optimizing the current buffer size and accepting the video quality degradation during the silent period. Even if we have interruptions, their length must be greatly reduced.

  Indeed, equations Eq. (5.5) and Eq. (5.6) provide the video quality that the user can play from the buffer and satisfy the condition $T_{\text{spBuffer}} \geqslant T_{\text{spMax}}$ (see Lines 37 to 40).

5.3 Validation and Performance Evaluation

5.3.1 Methodology

We conducted several experiments to evaluate our solution proposed in [132]. In our numerical experiments, we quantify the user device privacy satisfaction rate (P_{satRate}) by measuring the QoS in terms of interruption length (I_{length}) during a video streaming session. Next, we ran the same experimentation, but with the use of static CSs and random SPs as proposed in [67]. Then, we compared the two

results. We define P_{satRate} as follows:

$$P_{\text{satRate}} = \frac{P_{\text{satisfaction}}}{A_{\text{changes}}} \times 100 \tag{5.8}$$

where, $P_{\text{satisfaction}}$ is the number of times the user device reaches the desired privacy after each pseudonym change. A_{changes} is the total number of pseudonym changes.

We selected two configurations: $[1, 45]$ and $[45, 80]$ in which we vary the number of users randomly. We use the first interval to simulate a region that contains an average density where the maximum number of users is 45, and the second interval to simulate a situation with a higher density where the maximum number of users is 80. In our simulations, we avoided extreme values such as situations where the number of users is in the extremes of the interval. Also to obtain stable values we repeated each simulation 10 times and we calculated the average values presented in tables and plots, particularly for P_{satRate} and I_{length}.

The chosen video length we used is 90min. The selected entropy values are $H = 4$ (medium privacy) and $H = 6$ (high privacy). Thus, 16 neighbors are needed to reach $H = 4$ and 64 neighbors to reach $H = 6$ (see Eq. (5.1) above). We used random SPs similar to those used in [67] varying in the interval $[T_{\text{spMin}}, T_{\text{spMax}}]$ where T_{spMin} is deterministic and T_{spMax} is random. In these circumstances, [67] has shown that the best silent period to achieve the maximum privacy is 24min. To show that this period has a negative impact on QoS for streaming applications, we have used the following intervals to vary the silent periods: [10sec., 30min.], [10sec., 5min.], and [10sec., 60sec.].

We ran the simulations with two network technologies: IEEE802.11g and IEEE802.11ac. The theoretical and practical bandwidth for IEEE802.11g are 54Mb/s and 25 Mb/s [74]. For IEEE802.11ac they are 1300Mb/s and 433Mb/s [74]. We used the practical bandwidths.

5.3.2 Privacy Measurement Results

We present the quantified P_{satRate} for the selected entropy values. Table 5.1 shows the results for W=25Mb/s, and Table 5.2 shows the results of W=433Mb/s.

As we can see, all the obtained P_{satRate} values using our approach are much better than those obtained with the random silent-period-based approach. In fact, using our approach, we can even stream a video in HD (720p) and reach 93% in P_{satRate} for $H = 4$ (Table 5.1).

Config.	**H**	**QoV**	P_{satRate} **Without our approach**			P_{satRate} **With our approach**
			[10s,60s]	**[10s,5m]**	**[10s,30m]**	**Dynamic SP**
[1, 45]	4	144p	50%	57%	75%	92%
		240p	50%	57%	75%	92%
		360p	50%	50%	75%	92%
		480p	53%	50%	62%	93%
		720p	57%	64%	62%	93%
		1080p	57%	64%	62%	90%
[45, 80]	6	144p	37%	50%	50%	83%
		240p	31%	33%	37%	60%
		360p	31%	33%	33%	55%
		480p	31%	33%	33%	55%
		720p	31%	33%	33%	55%
		1080p	31%	33%	33%	34%

Table 5.1: Obtained P_{satRate} (W = 25Mb/s).

As expected, without our approach the best P_{satRate} are obtained when we use the interval [10sec, 30min] to randomly generate SPs. The obtained P_{satRate} for this interval and $H = 6$ (in Table 5.1) are close to our approach's results. However, for W=433Mb/s, the P_{satRate} is much improved using our approach (in Table 5.2). This is so because in our approach the high bandwidth allows the user to accumulate more data during CSs, which increases T_{spBuffer}, then, it can stay for a long duration

in a SP without any degradation in QoS and waiting for more neighboring users to come in and reach the privacy. Once we have the needed number of users, the user device can interrupt its SP. All the obtained results confirm our intuition that privacy can be improved with our method while maximizing the perceived QoS.

ConFigure.	**H**	**QoV**	$P_{\mathbf{satRate}}$ **Without our approach**			$P_{\mathbf{satRate}}$ **With our approach**
			[10s,60s]	**[10s,5m]**	**[10s,30m]**	**Dynamic SP**
[1, 45]	4	144p	61%	64%	69%	98%
		240p	61%	64%	69%	98%
		360p	61%	64%	69%	98%
		480p	61%	64%	69%	98%
		720p	61%	64%	69%	98%
		1080p	61%	65%	69%	98%
[45, 80]	6	144p	35%	38%	41%	93%
		240p	35%	38%	41%	93%
		360p	35%	38%	41%	93%
		480p	35%	38%	41%	92%
		720p	35%	38%	39%	79%
		1080p	34%	38%	40%	62%

Table 5.2: Obtained P_{satRate} ($W = 433$Mb/s).

5.3.3 QoS Measurement Results

We present here the impact of a high privacy ($H = 6$) on the QoS. The obtained results are plotted in Figure 5.2 for W=25Mb/s and in Figure 5.3 for W=433Mb/s. We can see in these figures that the use of the interval [10sec, 30min] to randomly

generate SPs gives a low QoS, even if we use a high bandwidth (W=433Mb/s).

As indicated above, when $N \in [1, 45]$, we can not reach $H = 6$. So, with our approach the user device has to stay the maximum duration in SP. Despite that, our obtained QoS values are much better than the other approaches. In addition, when $N \in [45, 80]$, thus, it is possible that the user reaches $H = 6$, but this huge number of users will increase the interruption length and degrades the QoS (as depicted in figures 5.2 and 5.3). Despite that, with our approach we obtained the highest QoS.

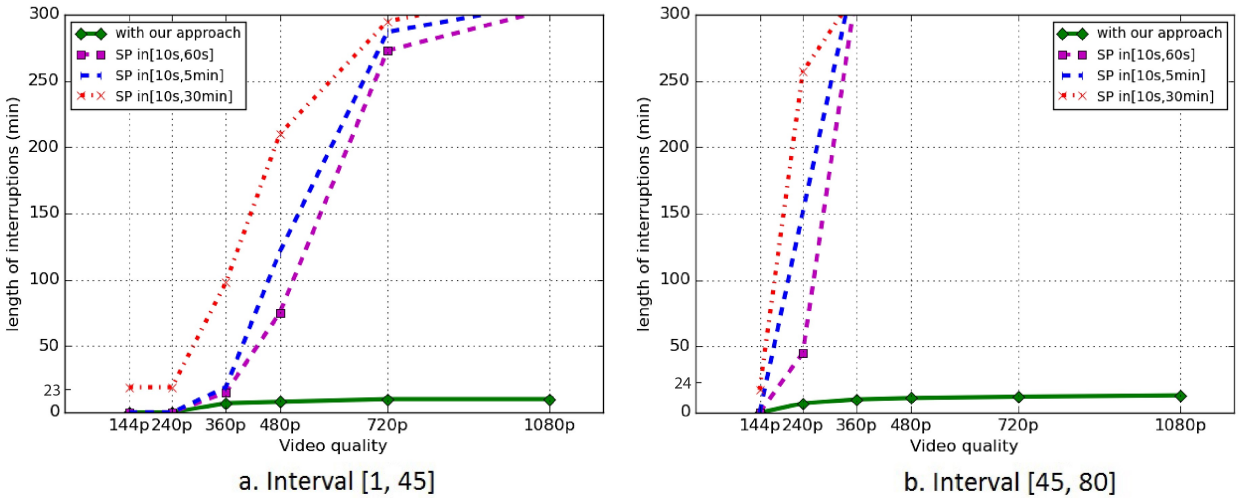

Figure 5.2: Obtained QoS measurements ($H = 6$ and $W = 25$Mb/s)

In Figure 5.3, the use of a high bandwidth W=433Mb/s (IEEE802.11ac), with our approach, we can stream a video without any degradation in QoS at 1080p and 720p (see Figure 5.3.a and Figure 5.3.b). We can also observe in Figure 5.2, that there is a steep increase in the interruptions duration with traditional methods. With our proposal, we can stream a video with a better QoS, but with a lower quality of video. This due to our technique of preserving QoS and degrading video quality during the silent periods.

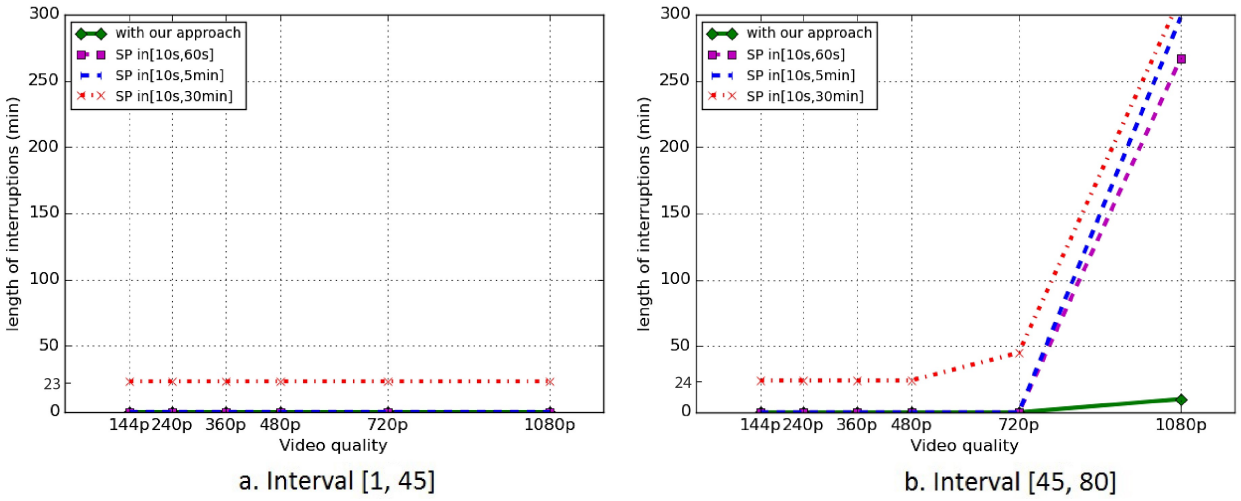

Figure 5.3: Obtained QoS measurements ($H = 6$ and $W = 433$Mb/s)

5.3.4 Discussion

We have experimentally demonstrated that our approach provides the best trade-off between privacy and QoS for video streaming applications compared to the existing solutions. With our proposal, we use make a trade-off between QoS and privacy. In addition, to further reduce interruptions cumulative length, our approach performs video quality downgrading when it is expected that the targeted privacy cannot be reached with the initial video quality.

5.4 Conclusion

Earlier solutions to location privacy in wireless local networks have mainly aimed at increasing location entropy by using pseudonyms which change from time to time, typically according to temporal and spatial conditions. These solutions did not necessarily take into consideration the effect of maximizing location entropy on other communication parameters such as the quality of service perceived by

applications.

In this chapter, we investigated the effect of using random silent periods as a temporal condition for pseudonym change to increase location privacy according to the entropy metric. We presented a mathematical model that quantifies both location entropy and QoS for video streaming and provides the best strategy to use to maximize the perceived QoS of service to reach a target level of privacy.

General Conclusion

Summary

Nowadays, wireless networking technologies have become increasingly popular in the computing era and users experience. One of the very successful and widespread technologies is WiFi, which gives its users the possibility to connect to the Internet and other WiFi local networks from anywhere, anytime. Due to its design, WiFi networks suffer from a certain vulnerabilities making them prone to threats and attacks, including attacks on location privacy.

One of the most common issues is that the signal transmitted by a mobile device identified by its MAC address can be captured by an attacker (or a set of attackers) which will use it to infer their location thereby making it possible to track the user's movement. In addition to the MAC addresses, there are other side information leaking from devices which can it identifiable even in the absence of MAC addresses (example when pseudonyms are used).

In this thesis we focused on how to quantify and protect the location privacy of mobile users in WiFi where an adversary tries to learn a user's past and current locations. We have presented a literature review in which we list major state-of-the-art contributions in this area and discussed their merits and limitations. In particular, we have presented two contributions for location privacy protection area of research, which are summarized as follows.

In first contribution [104], we have provided a mathematical model that allows

WiFi users to quantify the level of their location privacy using one of the most used metrics, namely the Entropy. We have provided a clear representation of the relation between privacy entropy, the use of pseudonyms, silent periods, and side information expressed as the mobility pattern. We also provided a decentralized algorithm to achieve the desired privacy level, in which each user device makes a decision on which strategy to use to achieve their objective in terms of privacy and quality of service. Privacy is ensured through the frequent change of pseudonyms (MAC addresses) using silent period at each change, in order to make it even harder for attackers to distinguish users based on their MAC addresses. We ran several experiments to assess the effect of various design parameters on the privacy preservation, and we evaluated our proposal with numerical simulations and mobility traces collected from WiFi users in an office environment. Our results proved the efficiency of our algorithm in providing the desired privacy level by the user.

In the second contribution [132], we have also looked into enhancing the algorithm in order to keep a better QoS for real time applications. Based on our first contribution [104], we have proposed another algorithm for QoS demanding applications, where, we have taken as an example the video streaming such as YouTube video sharing platform. In order to allow users to preserve their privacy without losing the quality of service (measured by the length of interruptions that may caused during the privacy protection process), we have proposed an enhanced *silent period*-based solution that allows to maximize the perceived QoS for a preset privacy. Our solution allows to: (i) quantify the desired user device location privacy and QoS, (ii) compute the silent and active periods that cope with these two objectives depending on the network parameters, and (iii) offer users the best trade-off between privacy and QoS. We experimented our proposal with a set of numerical simulations, with different configurations, and measured the impact of the introduction of privacy requirements on the perceived QoS with a privacy

preserving solution based on the use of random silent periods. The obtained results demonstrated the efficiency of the proposed solution compared to existing solutions.

Future Work

In our algorithm, each user chooses its silent period independently of the others which makes our solution decentralized and robust to networks dynamics. However, at the same time, it does not take into account the behavior of the neighbors which could affect the privacy of the user. As a future work, our algorithm could be significantly improved by taking into consideration the behavior of the neighbors of the user. The user can make use of game theory techniques, such as reinforcement learning to measure the effect of its action on the obtained privacy level and learn from experience the actions, i.e. *the silent periods* that need to be used to optimize its entropy in the future. In addition, although the silent period is an effective method against tracking, it does not provide protection against fingerprinting, because, the list of preferred networks of a mobile station, which is a list of all the networks the device previously has been associated with, is often still transmitted and the combination of silent periods with pseudonyms does not efficiently solve affect that. In fact, in the future we would like to solve this problem by combining our proposed solution with another solution that aims to protect or hide the implicit identifiers such as the history of visited networks.

In addition, the DHCP identifies the user device by its unique MAC address, and provide an IP address to get access to the Internet. Consequently, if the MAC address is not recognized by the DHCP server, then no IP address will be given to the user. Hence, as future solution, we aim to take into consideration the situation mentioned above, because, after the change of the MAC address the DHCP should be able to recognize the new MAC address and offer a corresponding IP address.

Bibliography

[1] *Guidelines for Use of Extended Unique Identifier (EUI), Organizationally Unique Identifier (OUI), and Company ID (CID)*. IEEE, 2017.

[2] *Aircrack-ng*, `https://aircrack-ng.org/` (Accessed: February 2020).

[3] *ZigBee Alliance*, `https://zigbeealliance.org/` (Accessed: February 2020).

[4] Naeim Abedi, Ashish Bhaskar, and Edward Chung. Bluetooth and wi-fi mac address based crowd data collection and monitoring: benefits, challenges and enhancement. In *Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia*, 2013.

[5] Arun Agarwal, Kabita Agarwal, Sumanshu Agarwal, and Gourav Misra. Evolution of mobile communication technology towards 5g networks and challenges. *American Journal of Electrical and Electronic Engineering*, 7(2):34–37, 2019.

[6] Maurizio Aiello, Enrico Cambiaso, Silvia Scaglione, and Gianluca Papaleo. A similarity based approach for application dos attacks detection. In *2013 IEEE Symposium on Computers and Communications (ISCC)*, pages 000430–000435. IEEE, 2013.

[7] Wahhab Albazrqaoe, Jun Huang, and Guoliang Xing. Practical Bluetooth Traffic Sniffing. In *Proceedings of the 14th Annual International Conference*

on Mobile Systems, Applications, and Services - MobiSys '16, pages 333–345, New York, New York, USA, 2016. ACM Press.

[8] Wahhab Albazrqaoe, Jun Huang, and Guoliang Xing. A practical bluetooth traffic sniffing system: Design, implementation, and countermeasure. *IEEE/ACM TRANSACTIONS ON NETWORKING*, 27(1):71–84, 02 2019.

[9] Belal Amro. Protecting privacy in vanets using mix zones with virtual pseudonym change. *arXiv preprint arXiv:1801.10294*, 2018.

[10] Gaeil An and Shinhyo Kim. Mac spoofing attack detection based on evm in 802.11 wlan. In *The Seventh International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, pages 163–167, 2013.

[11] a.R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[12] O Arana, F Garcia, J Gomez, and V Rangel. Msp: Providing location privacy in wlan networks with a mac swapping protocol. *Computer Networks*, 2018.

[13] Henrick Arfwedson and Rob Sneddon. Ericsson's bluetooth modules. *Ericsson Review*, 76(4):198–205, 1999.

[14] Frederik Armknecht, Joao Girao, Alfredo Matos, and Rui L Aguiar. Who said that? privacy at link layer. In *Proc. of INFOCOM*. IEEE, 2007.

[15] Nils Aschenbruck, Raphael Ernst, Elmar Gerhards-Padilla, and Matthias Schwamborn. Bonnmotion: a mobility scenario generation and analysis tool. In *Proceedings of the 3rd international ICST conference on simulation tools and techniques*, page 51. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2010.

[16] Marco V. Barbera, Alessandro Epasto, Alessandro Mei, Vasile C. Perta, and Julinda Stefa. Signals from the crowd: Uncovering social relationships

through smartphone probes. *Proceedings of the 2013 Conference on Internet Measurement Conference*, pages 265–276, 2013.

[17] Paolo Baronti, Prashant Pillai, Vince WC Chook, Stefano Chessa, Alberto Gotta, and Y Fun Hu. Wireless sensor networks: A survey on the state of the art and the 802.15. 4 and zigbee standards. *Computer communications*, 30(7):1655–1695, 2007.

[18] Chafika Benzaïd, Abderrahman Boulgheraif, Fatma Zohra Dahmane, Ameer Al-Nemrat, and Khaled Zeraoulia. Intelligent detection of mac spoofing attack in 802.11 network. In *Proceedings of the 17th International Conference on Distributed Computing and Networking*, pages 1–5, 2016.

[19] Alastair R Beresford and Frank Stajano. Mix zones: User privacy in location-aware services. In *Proceeding of PerCom.* IEEE, 2004.

[20] Bram Bonné, Arno Barzan, Peter Quax, and Wim Lamotte. Wifipi: Involuntary tracking of visitors at mass events. In *2013 IEEE 14th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pages 1–6. IEEE, 2013.

[21] Vladimir Brik, Suman Banerjee, Marco Gruteser, and Sangho Oh. Wireless device identification with radiometric signatures. In *proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, MobiCom '08, pages 116–127, New York, NY, USA, 2008. ACM.

[22] Jean-Bernard Brissaud. The meanings of entropy. *Entropy*, 7(1):68–96, 2005.

[23] Louise Cadoux and Pierre Tabatoni. Internet et protection de la vie privée. *Commentaire*, 23(89):57–66, 2000.

[24] Aaron Carroll, Gernot Heiser, et al. An analysis of power consumption in a smartphone. In *USENIX annual technical conference*, volume 14, pages 21–21. Boston, MA, 2010.

[25] Hao Chen, Yifan Zhang, Wei Li, and Ping Zhang. Non-cooperative wi-fi localization via monitoring probe request frames. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–5. IEEE, 2016.

[26] Hiten Choudhury, Basav Roychoudhury, and Dilip Kr Saikia. Enhancing user identity privacy in lte. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 949–957. IEEE, 2012.

[27] Mauro Conti, Nicola Dragoni, and Viktor Lesyk. A survey of man in the middle attacks. *IEEE Communications Surveys & Tutorials*, 18(3):2027–2051, 2016.

[28] Peter Cope, Joseph Campbell, and Thaier Hayajneh. An investigation of bluetooth security vulnerabilities. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 1–7, 01 2017.

[29] Mathieu Cunche. I know your mac address: Targeted tracking of individual using wi-fi. *Journal of Computer Virology and Hacking Techniques*, 10(4):219–227, 2014.

[30] Mathieu Cunche, Mohamed-Ali Kaafar, and Roksana Boreli. Linking wireless devices using information contained in wi-fi probe requests. *Pervasive and Mobile Computing*, 11:56–69, 04 2014.

[31] Mathieu Cunche, Mohamed Ali Kaafar, and Roksana Boreli. I know who you will meet this evening! Linking wireless devices using Wi-Fi probe requests. *2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2012 - Digital Proceedings*, 2012.

[32] Adrian Dabrowski, Nicola Pianta, Thomas Klepp, Martin Mulazzani, and Edgar Weippl. Imsi-catch me if you can: Imsi-catcher-catchers. In *Proceedings of the 30th annual computer security applications Conference*, pages 246–255, 2014.

[33] Cuthbert Daniel and Wilkinson Glenn. Snoopy: Distributed tracking and profiling framework. *44Con 2012*, 2012.

[34] Adriano Di Luzio, Alessandro Mei, and Julinda Stefa. Mind your probes: De-anonymization of large crowds through smartphone WiFi probe requests. *Proceedings - IEEE INFOCOM*, 2016-July(Section IV), 2016.

[35] Tim Dittler, Florian Tschorsch, Stefan Dietzel, and Björn Scheuermann. Anotel: Cellular networks with location privacy. In *2016 IEEE 41st Conference on Local Computer Networks (LCN)*, pages 635–638. IEEE, 11 2016.

[36] Tomislav Dragičević, Pierluigi Siano, SR Prabaharan, et al. Future generation 5g wireless networks for smart grid: a comprehensive review. *Energies*, 12(11):2140, 2019.

[37] Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278, 2009.

[38] Yanfei Fan, Bin Lin, Yixin Jiang, and Xuemin Shen. An Efficient Privacy-Preserving Scheme for Wireless Link Layer Security. In *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, pages 1–5. IEEE, 2008.

[39] Syeda Gauhar Fatima, Syeda Kausar Fatima, Syed Abdul Sattar, Syed Adil, and Khaja Fouzan Ahmed. Home automation using zigbee technology and iot. *Technology*, 10(2):92–96, 2019.

[40] Pierre-Alain Fouque, Cristina Onete, and Benjamin Richard. Achieving better privacy for the 3gpp aka protocol. *Proceedings on Privacy Enhancing Technologies*, 2016(4):255–275, 2016.

[41] Jason Franklin, Damon McCoy, Parisa Tabriz, Vicentiu Neagoe, J Van Randwyk, and Douglas Sicker. Passive data link layer 802.11 wireless device driver fingerprinting. In *USENIX Security Symposium*, volume 3, pages 16–89, 2006.

[42] Julien Freudiger. Short: How Talkative is your Mobile Device? An Experimental Study of Wi-Fi Probe Requests. *WiSec '15 Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, pages 1–6, 2015.

[43] Julien Freudiger, Maxim Raya, Márk Félegyházi, Panos Papadimitratos, et al. Mix-zones for location privacy in vehicular networks. *ACM Workshop on WiN-ITS*, 2007.

[44] Launay Frédéric and Perez André. *LTE Advanced Pro*. Wiley, 2019.

[45] Norihiro Fukumoto, Shigehiro Ano, and Shigeki Goto. Passive smart phone indentification and tracking with application set fingerprints. *Proceedings of the Asia-Pacific Advanced Network*, 36:41–48, 2013.

[46] Pimmy Gandotra and Rakesh Kumar Jha. Device-to-device communication in cellular networks: A survey. *Journal of Network and Computer Applications*, 71:99–117, 2016.

[47] Matthew Gast. *802.11 wireless networks: the definitive guide*. "O'Reilly Media, Inc.", 2005.

[48] MK Geir et al. Privacy enhanced mutual authentication in lte. In *2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 614–621. IEEE, 2013.

[49] Ryan M. Gerdes, Thomas E. Daniels, Mani Mina, and Steve F. Russell. Device identification via analog signal fingerprinting: A matched filter approach. In *proceedings of NDSS*, page 78, 2006.

[50] Taher Ahmed Ghaleb. Techniques and countermeasures of website/wireless traffic analysis and fingerprinting. *Cluster Computing*, 19(1):427–438, 2016.

[51] Ben Greenstein, Damon McCoy, Jeffrey Pang, Tadayoshi Kohno, Srinivasan Seshan, and David Wetherall. Improving wireless privacy with an identifier-free link layer protocol. In *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, MobiSys '08, pages 40–53, New York, NY, USA, 2008. ACM.

[52] Marco Gruteser and Dirk Grunwald. Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers: A Quantitative Analysis. *Mobile Networks and Applications*, 10(3):315–325, 06 2005.

[53] Xiaodan Gu, Ming Yang, Congcong Shi, Zhen Ling, and Junzhou Luo. A novel attack to track users based on the behavior patterns. *Concurrency and Computation: Practice and Experience*, 29(6):e3891, 2017.

[54] Jeyanthi Hall. Enhancing intrusion detection in wireless networks using radio frequency fingerprinting. *Communications, internet, and information technology*, pages 201–206, 2004.

[55] Jeyanthi Hall, Michel Barbeau, and Evangelos Kranakis. Detection of transient in radio frequency fingerprinting using signal phase. *Wireless and Optical Communications*, pages 13–18, 2003.

[56] Seungyeop Han, Vincent Liu, Qifan Pu, Simon Peter, Thomas Anderson, Arvind Krishnamurthy, and David Wetherall. Expressive privacy control with pseudonyms. *ACM SIGCOMM Computer Communication Review*, 43(4):291–302, 2013.

[57] Shaikh Shahriar Hassan, Soumik Das Bibon, Md Shohrab Hossain, and Mohammed Atiquzzaman. Security threats in bluetooth technology. *Computers & Security*, 74:308–322, 2018.

[58] Martin Herfurt and Collin Mulliner. Remote device identification based on bluetooth fingerprinting techniques. *Trifinite Group, White Paper*, 2004.

[59] Byeongdo Hong, Sangwook Bae, and Yongdae Kim. Guti reallocation demystified: Cellular location tracking with changing temporary identifier. In *25th Annual Network and Distributed System Security Symposium, NDSS*, 2018.

[60] Syed Hussain, Omar Chowdhury, Shagufta Mehnaz, and Elisa Bertino. Lteinspector: A systematic approach for adversarial testing of 4g lte. In *25th Annual Network and Distributed System Security Symposium, NDSS*, 2018.

[61] Syed Rafiul Hussain, Mitziu Echeverria, Omar Chowdhury, Ninghui Li, and Elisa Bertino. Privacy attacks to the 4g and 5g cellular paging protocols using side channel information. In *Annual Network and Distributed System Security Symposium, NDSS*, 2019.

[62] IEEE. IEEE Public OUI list. `http://standards-oui.ieee.org/oui.txt`. Accessed: 2017-04-01.

[63] IEEE. 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)*, pages 1–3534, 12 2016.

[64] Google Inc. Geolocation api. `https://developers.google.com/maps/documentation/geolocation/intro`.

[65] Taher Issoufaly. *Physical Tracking: menaces, performances et applications.* PhD thesis, La Réunion, 2019.

[66] Shuja Jamil, Sohaib Khan, Anas Basalamah, and Ahmed Lbath. Classifying smartphone screen on/off state based on wifi probe patterns. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 301–304. ACM, 2016.

[67] Tao Jiang, Helen J. Wang, and Yih-Chun Hu. Preserving location privacy in wireless lans. In *Proc. of MobiSys*. ACM, 2007.

[68] Paria Jokar, Nasim Arianpoo, and Victor CM Leung. Spoofing detection in ieee 802.15.4 networks based on received signal strength. *Ad hoc networks*, 11(8):2648–2660, 2013.

[69] Paria Jokar, Nasim Arianpoo, and Victor CM Leung. Spoofing prevention using received signal strength for zigbee-based home area networks. In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 438–443. IEEE, 10 2013.

[70] Mohammed Shafiul Alam Khan and Chris J Mitchell. Improving air interface user privacy in mobile telephony. In *International Conference on Research in Security Standardisation*, pages 165–184. Springer, 2015.

[71] Mohammed Shafiul Alam Khan and Chris J Mitchell. Trashing imsi catchers in mobile networks. In *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 207–218. ACM, 2017.

[72] Salam Khanji, Farkhund Iqbal, and Patrick Hung. Zigbee security vulnerabilities: Exploration and evaluating. In *2019 10th International Conference on Information and Communication Systems (ICICS)*, pages 52–57. IEEE, 2019.

[73] Yu Seung Kim, Yuan Tian, Le T. Nguyen, and Patrick Tague. LAPWiN: Location-aided probing for protecting user privacy in Wi-Fi networks. *2014 IEEE Conference on Communications and Network Security, CNS 2014*, pages 427–435, 2014.

[74] Samad S Kolahi and AA Almatrook. Impact of security on bandwidth and latency in ieee 802.11 ac client-to-server wlan. In *proceeding of ICUFN*. IEEE, 2017.

[75] Abhishek Kumar and Eep Gupta. Study on zigbee technology. *International Journal of Engineering & Research Technology*, 10 2013.

[76] T. Kumar and P. B. Mane. Zigbee topology: A survey. In *2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pages 164–166, 12 2016.

[77] Denis Foo Kune, John Koelndorfer, Nicholas Hopper, and Yongdae Kim. Location leaks on the gsm air interface. *ISOC NDSS (Feb 2012)*, 2012.

[78] Günther Lackner, Peter Teufl, and Roman Weinberger. User tracking based on behavioral fingerprints. In *International Conference on Cryptology and Network Security*, pages 76–95. Springer, 2010.

[79] Dennis Lee and Matthew Fischer. System and method for a flexible mac layer interface in a wireless local area network, 1997. US Patent 5,636,140.

[80] Leping Huang, Kanta Matsuura, Hiroshi Yamane, and Kaoru Sezaki. Enhancing wireless location privacy using silent period. In *IEEE Wireless*

Communications and Networking Conference, 2005, volume 2, pages 1187–1192. IEEE, 2005.

[81] Enos Letsoalo and Sunday Ojo. Survey of media access control address spoofing attacks detection and prevention techniques in wireless networks. In *2016 IST-Africa Week Conference*, pages 1–10. IEEE, 2016.

[82] C. Li, K. Dong, F. Jin, J. Song, and W. Mo. Design of smart home monitoring and control system based on zigbee and wifi. In *2019 Chinese Control Conference (CCC)*, pages 6345–6348. IEEE, 07 2019.

[83] Fenghua Li, Xinyu Wang, Ben Niu, Hui Li, Chao Li, and Lihua Chen. Tracku: Exploiting user's mobility behavior via wifi list. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.

[84] Jie Li, Fanzi Zeng, Zhu Xiao, Hongbo Jiang, Zhirun Zheng, Wenping Liu, and Ju Ren. Drive2friends: Inferring social relationships from individual vehicle mobility data. *IEEE Internet of Things Journal*, 2020.

[85] Zhenhua Li, Weiwei Wang, Christo Wilson, Jian Chen, Chen Qian, Taeho Jung, Lan Zhang, Kebin Liu, Xiangyang Li, and Yunhao Liu. Fbs-radar: Uncovering fake base stations at scale in the wild. In *NDSS*, 2017.

[86] Janne Lindqvist, Tuomas Aura, George Danezis, Teemu Koponen, Annu Myllyniemi, Jussi Mäki, and Michael Roe. Privacy-preserving 802.11 access-point discovery. In *Proceedings of the second ACM conference on Wireless network security - WiSec '09*, page 123, New York, New York, USA, 01 2009. ACM Press.

[87] Edoardo Longo, Alessandro EC Redondi, and Matteo Cesana. Pairing wi-fi and bluetooth mac addresses through passive packet capture. In *2018 17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, pages 1–4. IEEE, 2018.

[88] Edoardo Longo, Alessandro EC Redondi, and Matteo Cesana. Accurate occupancy estimation with wifi and bluetooth/ble packet capture. *Computer Networks*, 163:106876, 2019.

[89] Angela M Lonzetta, Peter Cope, Joseph Campbell, Bassam J Mohd, and Thaier Hayajneh. Security vulnerabilities in bluetooth technology as used in iot. *Journal of Sensor and Actuator Networks*, 7(3):28, 2018.

[90] Sejal S Lunawat, Pranjali S Dethe, and Paridhi M Jain. 5g wireless technology. *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)*, pages 86–89, 2018.

[91] Roger B Marks, Ian C Gifford, and Bob O'Hara. Standards in ieee 802 unleash the wireless internet. *IEEE microwave Magazine*, 2(2):46–56, 2001.

[92] Jeremy Martin, Travis Mayberry, Collin Donahue, Lucas Foppe, Lamont Brown, Chadwick Riggins, Erik C Rye, and Dane Brown. A study of mac address randomization in mobile devices and when it fails. *Proceedings on Privacy Enhancing Technologies*, 2017(4):365–383, 2017.

[93] Y. Matsuno, M. Ito, and K. Sezaki. Impact of time-varying population density on location privacy preservation level. In *Proc. of IEEE PerCom Workshops*, pages 1–6, 03 2016.

[94] Célestin Matte and Mathieu Cunche. Panoptiphone: How unique is your wi-fi device? In *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, pages 209–211, 2016.

[95] Imran Memon, Ling Chen, Qasim Ali Arain, Hina Memon, and Gencai Chen. Pseudonym changing strategy with multiple mix zones for trajectory privacy protection in road networks. *International Journal of Communication Systems*, 31(1), 2018.

[96] Joseph Meyerowitz and Romit Roy Choudhury. Hiding stars with fireworks: location privacy through camouflage. In *Proceeding of MobiCom*, pages 345–356. ACM, 2009.

[97] Nateq Be-Nazir Ibn Minar and Mohammed Tarique. Bluetooth security threats and solutions: a survey. *International Journal of Distributed and Parallel Systems*, 3(1):127, 2012.

[98] Zarrin Montazeri, Amir Houmansadr, and Hossein Pishro-Nik. Achieving perfect location privacy in markov models using anonymization. In *2016 International Symposium on Information Theory and Its Applications (ISITA)*, pages 355–359. IEEE, 2016.

[99] Zarrin Montazeri, Amir Houmansadr, and Hossein Pishro-Nik. Achieving perfect location privacy in wireless devices using anonymization. *IEEE Transactions on Information Forensics and Security*, 12(11):2683–2698, 2017.

[100] David Murray, Michael Dixon, and Terry Koziniec. Scanning delays in 802.11 networks. In *The 2007 International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 2007)*, pages 255–260. IEEE, 2007.

[101] ABM Musa and Jakob Eriksson. Tracking unmodified smartphones using wi-fi monitors. In *Proceedings of the 10th ACM conference on embedded network sensor systems*, pages 281–294. ACM, 2012.

[102] T. A. Myrvoll, J. E. Håkegård, T. Matsui, and F. Septier. Counting public transport passenger using wifi signatures of mobile devices. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6, 10 2017.

[103] M Tech Scholar Ashish Nagar and Shiva Bhatnagar. Zigbee on wireless sensor network for patient monitoring and home automation application. *International Journal of Scientific Research & Engineering Trends*, 5, 2019.

[104] Senoussi Nour-ElHouda, Bachir Abdelmalik, and Bouabdallah Abdelmadjid. On qos-aware location privacy in mobile networks. *Proceeding of Journal of Information and Computer Security (IJICS) 2018*, 2018.

[105] Olayemi Olawumi, Keijo Haataja, Mikko Asikainen, Niko Vidgren, and Pekka Toivanen. Three practical attacks against zigbee security: Attack scenario definitions, practical experiments, countermeasures, and lessons learned. In *2014 14th International Conference on Hybrid Intelligent Systems*, pages 199–206. IEEE, 2014.

[106] Ubertooth One. *Ubertooth*.

[107] Joseph Ooi. Imsi catchers and mobile security. *School of Engineering and Applied Science University of Pennsylvania*, 2015.

[108] Brendan O'Connor. Creepydol: Cheap, distributed stalking. *BlackHat US 13*, 2013.

[109] Balaji Palanisamy, Ling Liu, Kisung Lee, Aameek Singh, and Yuzhe Tang. Location privacy with road network mix-zones. In *Proceeding of MSN*. IEEE, 2012.

[110] Jeffrey Pang, Ben Greenstein, Srinivasan Seshan, and David Wetherall. Tryst: The case for confidential service discovery. *In HotNets VI: The Sixth Workshop on Hot Topics in Networks*, 2(3.2):1, 2007.

[111] Sagarkumar Patel, Vatsal Shah, and Maharshi Kansara. Comparative study of 2g, 3g and 4g. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, page 1962, 2018.

[112] Jonathan Petit, Florian Schaub, Michael Feiri, and Frank Kargl. Pseudonym schemes in vehicular networks: A survey. *IEEE communications surveys & tutorials*, 17(1):228–255, 2015.

[113] Anh Pham, Italo Dacosta, Bastien Jacot-Guillarmod, Kévin Huguenin, Taha Hajar, Florian Tramèr, Virgil Gligor, and Jean-Pierre Hubaux. Privateride: A privacy-enhanced ride-hailing service. *Proceedings on PET*, 2017(2):38–56, 2017.

[114] Francesco Potortì, Antonino Crivello, Michele Girolami, Emilia Traficante, and Paolo Barsocchi. Wi-fi probes as digital crumbs for crowd localisation. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8. IEEE, 2016.

[115] Pichaya Prasertsung and Teerayut Horanont. How does coffee shop get crowded? using wifi footprints to deliver insights into the success of promotion. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, page 421–426, New York, NY, USA, 2017. Association for Computing Machinery.

[116] Daniele Quercia and Licia Capra. Friendsensing: recommending friends using mobile phones. In *Proceedings of the third ACM conference on Recommender systems*, pages 273–276, 2009.

[117] Jeremy Quirke. Security in the gsm system. *Journal of AusMobile, May*, pages 1–26, 2004.

[118] N. Radio, Y. Zhang, M. Tatipamula, and V. K. Madisetti. Next-generation applications on cellular networks: Trends, challenges, and solutions. *Proceedings of the IEEE*, 100(4):841–854, 4 2012.

[119] Taibur Rahman and Swarnendu Kumar Chakraborty. Provisioning technical interoperability within zigbee and ble in iot environment. In *2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, pages 1–4. IEEE, 2018.

[120] SM Sohel Rana, M Robiul Hoque, and M Humayun Kabir. Evaluation of security threat of zigbee protocol to enhance the security of zigbee based iot platform. *Journal of Applied Science and Technology*, 11(01):37–45, 2019.

[121] Ahmad RazaCheema, Malek Alsmadi, and Salama Ikki. Survey of identity-based attacks detection techniques in wireless networks using received signal strength. In *2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)*, pages 1–6. IEEE, 2018.

[122] Alessandro Enrico Cesare Redondi, Davide Sanvito, and Matteo Cesana. Passive classification of wi-fi enabled devices. In *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 51–58. ACM, 2016.

[123] Paul Reichl, Beng Oh, Ravi Ravitharan, and Mark Stafford. Using wifi technologies to count passengers in real-time around rail infrastructure. In *2018 International Conference on Intelligent Rail Transportation (ICIRT)*, pages 1–5. IEEE, 12 2018.

[124] Pieter Robyns, Bram Bonné, Peter Quax, and Wim Lamotte. Poster: assessing the impact of 802.11 vulnerabilities using wicability. In *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, pages 217–218. ACM, 2016.

[125] Pieter Robyns, Bram Bonné, Peter Quax, and Wim Lamotte. Noncooperative 802.11 mac layer fingerprinting and tracking of mobile devices. *Security and Communication Networks*, 2017, 2017.

[126] Ian Rose and Matt Welsh. Mapping the urban wireless landscape with argos. In *proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pages 323–336. ACM, 2010.

[127] KVSSSS Sairam, N Gunasekaran, and S Rama Redd. Bluetooth in wireless communication. *IEEE Communications Magazine*, 40(6):90–96, 2002.

[128] T Scott Saponas, Jonathan Lester, Carl Hartung, Sameer Agarwal, Tadayoshi Kohno, et al. Devices that tell on you: Privacy trends in consumer ubiquitous computing. In *USENIX Security Symposium*, pages 55–70, 2007.

[129] Lorenz Schauer and Claudia Linnhoff-Popien. Extracting context information from wi-fi captures. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '17, page 123–130, New York, NY, USA, 2017. Association for Computing Machinery.

[130] Lorenz Schauer, Martin Werner, and Philipp Marcus. Estimating crowd densities and pedestrian flows using wi-fi and bluetooth. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 171–177. ACM, 2014.

[131] Vanja Seničar, Borka Jerman-Blažič, and Tomaž Klobučar. Privacy-enhancing technologies—approaches and development. *Computer Standards & Interfaces*, 25(2):147–158, 2003.

[132] Nour El Houda Senoussi, Mohamed Lamine Kerdoudi, Abdelmalik Bachir, and Abdelmadjid Bouabdallah. On enhancing location privacy and qos for video streaming over wireless networks. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2018.

[133] Mansoor Shafi, Akira Hashimoto, Masahiro Umehira, Shigeaki Ogose, and Takehiro Murase. Wireless communications in the twenty-first century: A perspective. *Proceedings of the IEEE*, 85(10):1622–1638, 1997.

[134] Altaf Shaik, Ravishankar Borgaonkar, N. Asokan, Valtteri Niemi, and Jean-Pierre Seifert. Practical Attacks Against Privacy and Availability in 4G/LTE Mobile Communication Systems. In *Proceedings 2016 Network and Distributed System Security Symposium*, number February, pages 21–24, Reston, VA, 2016. Internet Society.

[135] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[136] Yong Sheng, Keren Tan, Guanling Chen, David Kotz, and Andrew Campbell. Detecting 802.11 mac layer spoofing using received signal strength. In *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*, pages 1768–1776. IEEE, 2008.

[137] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. Quantifying location privacy. In *proceedings of IEEE Symposium on Security and Privacy (sp)*. IEEE, 2011.

[138] G. Shyam Kishore and Hemalatha Rallapalli. Towards 5g: A survey on waveform contenders. In *Advances in Decision Sciences, Image Processing, Security and Computer Vision*, pages 243–250. Springer, 2020.

[139] Ankush Singla, Syed Rafiul Hussain, Omar Chowdhury, Elisa Bertino, and Ninghui Li. Protecting the 4g and 5g cellular paging protocols against security and privacy attacks. *Proceedings on Privacy Enhancing Technologies*, 1:126–142, 2020.

[140] GuangJia Song and ZhenZhou Ji. Novel duplicate address detection with hash function. *PloS one*, 11(3), 2016.

[141] Li Sun, Ramanujan K Sheshadri, Wei Zheng, and Dimitrios Koutsonikolas. Modeling wifi active power/energy consumption in smartphones. In

Distributed Computing Systems (ICDCS), 2014 IEEE 34th International Conference on, pages 41–51. IEEE, 2014.

[142] Keen Sung, Brian Neil Levine, and Marc Liberatore. Location privacy without carrier cooperation. In *IEEE Workshop on Mobile Security Technologies, MOST*, page 148, 2014.

[143] Mohsen Toorani and A. Beheshti. Solutions to the GSM Security Weaknesses. In *2008 The Second International Conference on Next Generation Mobile Applications, Services, and Technologies*, pages 576–581. IEEE, 2008.

[144] Ivan Vaccari, Enrico Cambiaso, and Maurizio Aiello. Remotely exploiting at command attacks on zigbee networks. *Security and Communication Networks*, 2017, 2017.

[145] Nitin H Vaidya. Weak duplicate address detection in mobile ad hoc networks. In *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*, pages 206–216. ACM, 2002.

[146] Fabian Van Den Broek, Roel Verdult, and Joeri de Ruiter. Defeating imsi catchers. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, pages 340–351, 10 2015.

[147] Thanh Van Do, Hai Thanh Nguyen, Nikolov Momchil, et al. Detecting imsi-catcher using soft computing. In *International Conference on Soft Computing in Data Science*, pages 129–140. Springer, 2015.

[148] Mathy Vanhoef, Domien Schepers, and Frank Piessens. Discovering logical vulnerabilities in the wi-fi handshake using model-based testing. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 360–371. ACM, 2017.

[149] Edwin Vattapparamban, Bekir Sait Çiftler, İsmail Güvenç, Kemal Akkaya, and Abdullah Kadri. Indoor occupancy tracking in smart buildings using passive sniffing of probe requests. In *Proceeding of ICC*. IEEE, 2016.

[150] Niko Vidgren, Keijo Haataja, Jose Luis Patino-Andres, Juan Jose Ramirez-Sanchis, and Pekka Toivanen. Security threats in zigbee-enabled systems: vulnerability evaluation, practical experiments, countermeasures, and lessons learned. In *2013 46th Hawaii International Conference on System Sciences*, pages 5132–5138. IEEE, 2013.

[151] Otto Waltari and Jussi Kangasharju. The wireless shark: Identifying wifi devices based on probe fingerprints. In *Proceedings of the First Workshop on Mobile Data*, MobiData '16, page 1–6, New York, NY, USA, 2016. Association for Computing Machinery.

[152] Qijin Wang, Dandan Wang, and Xiaoxia Qi. An energy-efficient routing protocol for zigbee networks. In *IOP Conference Series: Earth and Environmental Science*, volume 295, page 052040. IOP Publishing, 2019.

[153] Shibin Wang, Nianmin Yao, Ning Gong, and Zhenguo Gao. A trigger-based pseudonym exchange scheme for location privacy preserving in vanets. *Peer-to-Peer Networking and Applications*, 11(3):548–560, 2018.

[154] W. Wang, G. He, and J. Wan. Research on zigbee wireless communication technology. In *2011 International Conference on Electrical and Control Engineering*, pages 1245–1249, 2011.

[155] Yan Wang, Jie Yang, Hongbo Liu, Yingying Chen, Marco Gruteser, and Richard P Martin. Measuring human queues using wifi signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 235–238. ACM, 2013.

[156] Jens Weppner, Benjamin Bischke, and Paul Lukowicz. Monitoring crowd condition in public spaces by tracking mobile consumer devices with wifi interface. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*, pages 1363–1371, New York, New York, USA, 2016. ACM Press.

[157] WiGLE. *Wireless Geographic Logging Engine*, (accessed September 5, 2016). `http://wigle.net`.

[158] Charles V Wright, Scott E Coull, and Fabian Monrose. Traffic morphing: An efficient defense against statistical traffic analysis. In *Proceeding of NDSS*, 2009.

[159] Shoubai Xiao. The research on data transmission application based on zigbee wireless network. In *2019 9th International Conference on Information and Social Science (ICISS 2019)*, 2019.

[160] Rong Yu, Jiawen Kang, Xumin Huang, Shengli Xie, Yan Zhang, and Stein Gjessing. Mixgroup: Accumulative pseudonym exchanging for location privacy enhancement in vehicular social networks. *IEEE Transactions on Dependable and Secure Computing*, 2016.

[161] Fan Zhang, Wenbo He, and Xue Liu. Defending Against Traffic Analysis in Wireless Networks through Traffic Reshaping. In *2011 31st International Conference on Distributed Computing Systems*, pages 593–602. IEEE, 06 2011.

[162] T Zillner. Zigbee exploited: The good, the bad and the ugly. In *IN DEPTH SECURITY VOL. II, Proceedings of the DeepSec Conferences*, pages 251–259, 2017.

[163] Mohammed Zubair, Devrim Unal, Abdulla Al-Ali, and Abdullatif Shikfa. Exploiting bluetooth vulnerabilities in e-health iot devices. In *Proceedings of*

the 3rd International Conference on Future Networks and Distributed Systems, pages 1–7. ACM, 2019.

[164] Rubina S. Zuberi and Syed N. Ahmad. Secure Mix-Zones for Privacy Protection of Road Network Location Based Services Users. *Journal of Computer Networks and Communications*, 2016(c):1–8, 2016.