

Department of Communications and Networking S-38  
AALTO UNIVERSITY OF SCIENCE AND TECHNOLOGY  
Faculty of Electronics, Communications and Automation

Miriam González

ANALYSIS OF WIRELESS INTERNET USAGE IN PUBLIC ACCESS  
POINTS

Otaniemi March 1, 2010

Thesis supervisor: Professor Jörg Ott  
Thesis instructor: Lic.Sc. Mikko Pitkänen

Author: Miriam González

Title: Analysis of Wireless Internet Usage in Public Access Points

Date: March 1, 2010

Language: English

Number of pages: 8+74

Faculty: Faculty of Electronics, Communications and Automation

Professorship: Networking Technology

Code: S-38

Supervisor: Professor Jörg Ott

Instructor: Lic.Sc. Mikko Pitkänen

Wireless local-area networks (WLANs) are increasingly common in public locations such as urban areas, universities, airports or shopping centers. A clear understanding of usage behavior in real hotspots is critical information for those who develop, deploy, and manage WLAN technology as well as those who develop systems and application software for wireless networks.

This thesis analyzes user behavior and network performance in public wireless networks in Helsinki, collecting traffic traces over several days at different access points. The analysis covers locations, which provide hotspots connectivity in different social environments including, libraries, museums or cafeterias. The capture process focuses on open access point infrastructures, where the access to the Internet is free and becoming ubiquitous. This enables the study of user behavior in urban environments.

The first part of the thesis introduces the research field on a general basis, a literature survey is given to present the current state-of-art. The survey covers technologies and issues in open WLAN networks, relevant terminology and privacy aspects concerning the traffic collection. The performed research and capture methodology is then presented. The focus is on the challenges in wireless monitoring and data collection.

The findings and part of the thesis discusses the results from the capture analysis to evaluate the Internet usage in a public wireless LAN environment. As an outcome, the thesis describes how free hotspots provide services to a diverse population that use the WLAN through open access points. In particular, web based applications are the most popular with the clients in this study. The traces show variable web response delay and different connectivity durations for browsing behavior and diversity of user attitudes depending on the location.

Keywords: WLAN user access, Analysis, Free WLAN Hotspot, User behavior

## Abbreviations and Acronyms

<b>BSSID</b>	Basic Service Set Identifier
<b>CDPD</b>	Cellular Digital Packet Data
<b>CSMA/CA</b>	Carrier Sense Multiple Access/Collision Avoidance
<b>DSSS</b>	Direct Sequence Spread
<b>ECC</b>	Electronic Communications Committee
<b>ESSID</b>	Extended Service Set
<b>ETSI</b>	European Telecommunications Standard
<b>FHSS</b>	Frequency Hopping Spread Spectrum
<b>FTP</b>	File Transfer Protocol
<b>IP</b>	Internet Protocol
<b>MAC</b>	Medium Access Control
<b>PDA</b>	Personal Digital Assistant
<b>PHY</b>	Physical Layer
<b>PCS</b>	Personal Communication Service
<b>RTT</b>	Round Trip Time
<b>SMTP</b>	Simple Mail Transfer Protocol
<b>SSH</b>	Secure SHell
<b>TCP</b>	Transfer Control Protocol
<b>VPN</b>	Virtual Private Network
<b>WPAN</b>	Wireless Personal Area Network

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Abbreviations and Acronyms</b>	<b>iii</b>
<b>Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Scenario . . . . .	2
1.2 Methodology-Wireless Packet Capture and Analysis . . . . .	3
1.3 Structure of the Thesis . . . . .	5
<b>2 Related Work</b>	<b>7</b>
2.1 Wireless Network Usage and User behavior . . . . .	7
2.2 Hotspot Environment . . . . .	9
<b>3 Wireless Access in Public WLAN Hotspots</b>	<b>11</b>
3.1 Introduction to IEEE 802.11 Technologies . . . . .	12
3.2 Limits of Wireless Networking . . . . .	14
3.2.1 Issues with TCP/IP over Wireless LANs . . . . .	16
3.3 Security Issues in WLAN Access . . . . .	17
3.4 Data Mining and Privacy Policies . . . . .	19
3.5 Wireless Network Access . . . . .	20
3.5.1 TCP Flow . . . . .	21
3.5.2 Round Trip Time . . . . .	22
3.5.3 Sessions . . . . .	23
<b>4 Capturing WLAN Traffic</b>	<b>27</b>
4.1 Challenges of Wireless Monitoring . . . . .	27
4.2 Data Collection . . . . .	28
4.3 Methodology Used for Analysis . . . . .	30

<b>5</b>	<b>Measurements</b>	<b>33</b>
5.1	Definitions . . . . .	33
5.2	User View . . . . .	34
5.2.1	User Duration . . . . .	34
5.2.2	User Devices . . . . .	34
5.2.3	Traffic Patterns per User . . . . .	35
5.3	Traffic Statistics of Public WLAN Traffic . . . . .	38
5.3.1	IP Traffic Load . . . . .	38
5.3.2	Analysis of User TCP traffic Load . . . . .	40
5.3.3	Round Trip Time Measurements . . . . .	43
5.4	Internet Usage in Wireless Hotspot . . . . .	51
5.4.1	Analysis of User Applications . . . . .	51
5.4.2	Analysis of HTTP traffic . . . . .	57
<b>6</b>	<b>Conclusions and Future Work</b>	<b>64</b>
	<b>References</b>	<b>67</b>
	<b>Acknowledgements</b>	<b>72</b>

## List of Figures

1.1	A map of the zone in the center with greater Helsinki area coverage. . . . .	2
1.2	A description of Warwalking process. . . . .	4
3.1	An example wireless access scenario. . . . .	11
3.2	IEEE 802.11 Protocol Layers. . . . .	13
3.3	TCP flow control. . . . .	22
3.4	An example web application session. . . . .	25
4.1	Monitoring wireless traffic, illustration based on [49]. . . . .	28
4.2	Flow chart used to obtain the results. . . . .	31
5.1	User traffic patterns in <i>Kamppi</i> location with four APs providing service. . . . .	35
5.2	User traffic patterns in <i>Library-Post</i> location with four APs providing service. . . . .	36
5.3	User traffic patterns in <i>Stockmann</i> location with four APs providing service. . . . .	37
5.4	Traffic IP pattern in <i>Kamppi</i> location with four APs providing service. . . . .	38
5.5	Traffic IP pattern in <i>Library-Post</i> location with six APs providing service. . . . .	39
5.6	Traffic IP pattern in <i>Stockmann</i> location with four APs providing service. . . . .	39
5.7	A TCP flow measurement from <i>Kamppi</i> . Distribution across 10 hours analysis with 51 connected clients (MAC) address in <i>Helsinki kaupungin WLAN</i> and 29 addresses in <i>Lasipalatsi WLAN</i> . . . . .	40
5.8	A TCP flow measurement from <i>Library-Post</i> . Distribution across 10 hours analysis with 25 MAC addresses in <i>Helsinki Kaupungin WLAN</i> and 55 MAC address in the <i>Stadinetti WLAN</i> . . . . .	41
5.9	A TCP flow measurement from <i>Stockmann</i> location. Distribution across 10 hours of capture with 48 MAC addresses in the <i>Kafka WLAN</i> and 50 addresses in the <i>Helsingin kaupungin WLAN</i> . . . . .	42
5.10	Analysis of the RTTs in the <i>Kamppi</i> location. . . . .	44
5.11	Analysis of the RTTs in <i>Library-Post</i> location. . . . .	45
5.12	Analysis of the RTTs in the <i>Stockmann</i> location. . . . .	46
5.13	Web domain connection time example. . . . .	48
5.14	Analysis of HTTP response delays in the <i>Kamppi(a)</i> and <i>Library-Post (b)</i> locations during a continuous trace of 3h. . . . .	49

5.15	Analysis of HTTP response delays in the <i>Stockmann</i> location during a continuous trace of 3h. . . . .	50
5.16	Analysis of web domain connection times in the <i>Kamppi</i> and <i>Library-Post</i> locations during a continuous trace of 3h. . . . .	50
5.17	Analysis of web domain connection delays in the <i>Stockmann</i> location during a continuous trace around 3h. . . . .	50
5.18	Traffic protocols distribution from 10 hours trace in the <i>Kamppi</i> , <i>Library-Post</i> and <i>Stockmann</i> locations. . . . .	51
5.19	Analysis of the use of applications by volume connected to the <i>Kamppi</i> AP location. . . . .	52
5.20	Analysis of the use of applications by volume connected to the <i>Library-Post</i> AP locations. . . . .	53
5.21	Analysis of the use of applications by volume connected to the <i>Stockmann</i> AP locations. . . . .	54
A.1	Traffic summary in <i>Kamppi</i> location . . . . .	74
A.2	Traffic summary in <i>Stockmann</i> location . . . . .	75
A.3	Traffic summary in <i>Library-Post</i> location . . . . .	76

## List of Tables

1.1	Overall statistics for the trace. . . . .	3
4.1	Main locations. . . . .	29
4.2	Summary of collected WLAN traffic. . . . .	30
5.1	The most common applications by number of users in the <i>Kamppi</i> location, incoming traffic (MB), outgoing traffic (MB). . . . .	55
5.2	The most common applications by user in the <i>Library-Post</i> location, incoming traffic (MB), outgoing traffic (MB). . . . .	56
5.3	The most common applications by user in the <i>Stockmann</i> location, incoming traffic (MB), outgoing traffic (MB). . . . .	56
5.4	The most popular HTTP hosts in the <i>Kamppi</i> location. . . . .	58
5.5	The most popular HTTP hosts in the <i>Library-Post</i> location. . . . .	58
5.6	The most popular HTTP hosts in the <i>Stockmann</i> location. . . . .	59
5.7	Statistics of HTTP response by type in the <i>Kamppi</i> location. . . . .	60
5.8	Statistics of Http response by type in the <i>Library-Post</i> location. . . . .	60
5.9	Statistics of Http response by type in the <i>Stockmann</i> locations. . . . .	61
5.10	Statistics of web resource sizes in <i>Kamppi</i> location. . . . .	62
5.11	Statistics of web resource sizes in the <i>Library-Post</i> location. . . . .	62
5.12	Statistics of web resource sizes in the <i>Stockmann</i> location. . . . .	62

# 1 Introduction

Wireless local-area networks (WLANs) are increasingly common, particularly in university campus areas, urban places and shopping centers. Standardized technology such as IEEE 802.11 [1] is now broadly deployed and its usage is increasing constantly. Wireless Internet service providers (WISPs) have established Wi-Fi hotspots in increasing numbers at public venues, providing local coverage to traveling users with the ability to access email, web, and other Internet applications on the move.

A clear understanding of usage behavior in real WLANs is interesting knowledge for those who develop, deploy, and manage WLAN technology. This hotspot networking equipment provide wireless Internet access in popular public places such as airports, shops and cafés, where individuals spend a considerable amount of their time outside of home and work. An understanding of how these hotspot networks are used can provide input to design, hotspot deployments and the development of technologies to be used in WLANs. In addition to the convenience of untethered networking, today's wireless LANs provide relatively high data connectivity and are easy to deploy in public settings.

Internet data brings benefits to learning, teaching and research, but a number of problems still plague Internet connectivity and usage. The objective of this thesis is to evaluate the usage patterns of Wireless Internet users in public WLANs. These usage patterns can feed input for academic research at Aalto University School of Science and Technology. It is clear that using the web is an important activity for many people so there is consensus regarding the apparent representative pattern in a large amount of data. We consider it important to collect traces of user behavior and search for characteristics of typical mobile usage to evaluate statistics in order to better understand user sessions in public WLAN traffic as well as client diversity in WLAN hotspots. This research continues earlier work on wireless networks, that has explored user behavior and traffic characteristics inside a university department network [2] and on a college campus [3].

This thesis presents and analyzes Internet traffic in a public-area Wireless network in Helsinki using traces recorded over several days at different access points. These traces consist on monitoring packet headers of traffic from several WLANs serving the urban area. The goal of the study is to analyze statistics based on these traces. We examine user behavior in order to understand the diversity of remote hosts which were contacted, traffic volumes, traffic flows size per applications, diversity of user

population and application Round Trip Times (RTTs).

## 1.1 Scenario

The City of Helsinki has set up a free wireless Internet network in a number of public spaces in the city center since summer 2008 [4]. Wireless Internet access at no cost is available to anyone with a device with a WLAN interface in downtown Helsinki both outdoors and in several public spaces including city libraries and cultural centers. Currently, around 50 base stations and 30 libraries maintained by the city already exist in downtown Helsinki.

The primary challenges for WLAN measurement include geographic diversity of WLAN deployment, and physical proximity required for WLAN packet capture. Coupled with the heterogeneity of user equipment (e.g, different devices, operating systems, and protocol stacks) and recent trends in Internet usage. Geographically Public WLANs present a challenging environment for network measurement. We collected traffic from several APs in the Helsinki center. This preliminary analysis will help us identify areas best suited to focus our attention in the analysis of information from these access points.

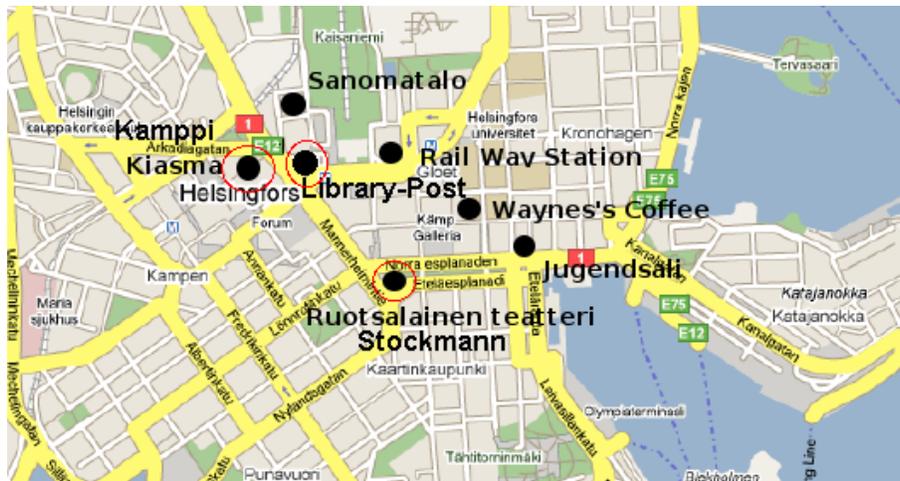


Figure 1.1: A map of the zone in the center with greater Helsinki area coverage.

Figure 1.1 shows a map plotting a zone in the center of Helsinki that is covered by the preliminary state of investigation with free Internet service provided by the city of Helsinki. The black points indicate some of the places offering open networks, which were selected in order to represent networks for customers, free library networks,

and free APs. However, some of these locations had insufficient traffic volumes to generate meaningful data for analysis.

The APs provided overlapping coverage during the capture process and share different Service Set Identifier (SSIDs). They were operating at a data rate of 11Mbps, using different channels and power of 100mW. When moving from one AP to another with different SSID within the network, a user must re-authenticate to obtain the Internet access at the new AP.

The wireless user population was different between the APs, but consisted of about several distinct users (MAC address). User wireless hardware was heterogeneous, such as PDAs, mobile-phones, mobile VoIP and laptop clients. The packet trace, a timestamped sequence of packets captured with a sniffer, includes traffic from 14 different APs and they have been analyzed taking into account to which of them the captured traffic was related.

Table 1.1: Overall statistics for the trace.

<b>Attribute</b>	<b>Values</b>
Number of wireless users	651
Total duration of traces	50 hours
Total bites transmitted	1.27 GB

The final traces were collected from three different locations (see Figure 1.1, red circles) in Helsinki center called *Kamppi* (4 APs), *Library-Post* (6 APs) and *Stockmann* (4 APs). These locations provide hotspots related to different social environments like libraries, museums and cafeteria. Table 1.1 summarizes the collected traffic and characteristics. The trace consists of a continuous data around 3 hours collected from each of the APs selected over a total period of 10 hours in each location. Moreover, the period was chosen for data collection based on the most active hours per day. We have collected the captures during different times in different days to obtain a large quantity of traffic to evaluate the user behavior in these APs.

## 1.2 Methodology-Wireless Packet Capture and Analysis

The following summarizes the methodology used to capture the user traffic during the trace analysis. The process basically consists of searching for free Wi-Fi networks by a person walking without any vehicle, using a portable computer. This act is

called *warwalking* and it has the inconveniency of lower speed of travel, but models well pedestrians using mobile Internet services. The wireless networking equipment used for Warwalking includes a laptop computer, a wireless access adapter that optionally supports an external antenna and a packet sniffing software on the laptop to control the wireless access adapter. The methodology is motivated by the idea that with wireless networking devices and open AP infrastructures, Internet access would become ubiquitous and free. The observed wireless access points were free public access hotspots, set up by individuals or organizations in the city of Helsinki explicitly for the purpose for providing free Internet access to anyone with the proper wireless networking equipment. In this set up the users do not need to first establish a payment relationship before being allowed to access the Internet through a wireless access point, i.e. hotspot.

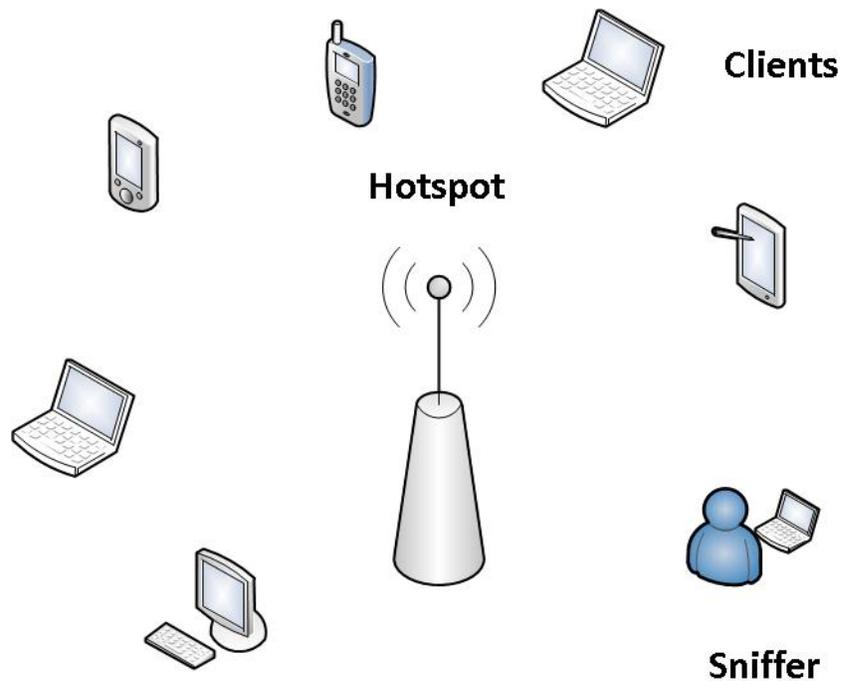


Figure 1.2: A description of Warwalking process.

Figure 1.2 shows the warwalking process that is used to capture the traffic from the free hotspots. The type of users connected to the APs can differ over the course of the traffic capture process, including devices such as mobile phones with WLAN access, PDAs, laptops and iPods. For the study, *tcpdump* [5] and *airodump* packet sniffers are installed to the laptop to capture any wireless traffic that passes through the APs. During the traffic collection we followed steps, which are applied to collect the traffic captures. The following lists the task during the collection process:

- Configuring network card in monitor mode for allowing the monitoring of all traffic received from the wireless network.
- Detecting and identifying the wireless network.
- Testing for active communication channels and SSID.
- Capturing network traffic by overhearing communication. For this task we use a network sniffer tool, which can be configured to collect packets that are interesting to our study.
- MAC address collection of access points and clients.

The public WLAN infrastructure is structured so that there is no convenient central point for capturing all traffic. However, there exist approximations about where the APs are situated and what radio range they can provide. This helps to evaluate different points in order to capture the traffic in the best possible success. The sniffer device that is used to collect the data for the analysis is a VGN Sony Vaio laptop with Linux OS and an Intel Wireless WiFi 4965AGN card that operates in both the 2.4GHz and 5.0GHz spectrum.

Ensuring the anonymity of users was a priority for us. The MAC address enables us to distinguish between users, but not identify a particular client. The IP addresses in the trace are temporary and cannot reveal the identity of individual users. Finally, we do not search for any private information contained in the payload. Moreover, we anonymize sensitive information to protect the privacy of the users.

We analyze header packets in the trace using Wireshark [6], freely available software suite for analyzing monitored network traffic. For post-processing, we implement python scripts that parse the Wireshark output for relevant analysis.

### 1.3 Structure of the Thesis

This thesis is structured as follows. Section 2 gives an overview on related work in the field. The section describes the most relevant studies made in a campus, public, and metropolitan areas.

Section 3 explains important aspects of wireless access in public hotspots and presents an introduction of different wireless technologies. In addition, privacy policies related to the traffic capture process are introduced. A description mechanism to keep

the confidentiality of user information safe is provided. Finally, we define important terminology related to the study.

A procedure is used to collect and analyze the traffic of the proposed WLANs structures and the description is presented in Section 4. Moreover, the challenge of wireless monitoring, global summary on collected data and the used tools in the study are briefly introduced.

Section 5 discusses the measurements conducted in this thesis. The result analysis evaluates the behavior of the wireless Internet usage in our scenarios.

Finally, most important findings are summarized and conclusions are drawn and explained in section 6 to provide an overview of the results. In addition, suggestions for future work and possible improvements to the presented study are summarized in this Section.

## 2 Related Work

Researchers and associations have performed a number of useful studies about wireless network usage. This section presents an overview on the work done in the past in this research area. In particular, user behavior and network performance in public wireless access points has been analyzed. Moreover, studies about different scenarios and research on trace collections are summarized. Several studies have been performed on wireless university campus networks and public networks. In this thesis, we complement the previous research by presenting results from our trace collected in the city environment in public places.

In many types of local networks, the amount of data transferred often reveals some information about network usage. Several studies have characterized wireless network usage in a variety of environments. Moreover, conferences such as the Passive and Active Measurement [7], provide current work from researchers and operations communities on topics including, active network measurements, traffic statistics, performance metrics, etc. An understanding about user behavior can help to guide the design of applications geared toward mobile environments, helps improve simulation tools by providing a better user mobility models, and helps to model user Internet behavior.

### 2.1 Wireless Network Usage and User behavior

Tang and Baker [2] analyzed the network for overall user behavior, overall network traffic, load characteristics and traffic characteristics from a user point of view during a twelve-week trace of a building-wide local-area wireless network in Stanford University. Their study provides a good qualitative description on how mobile users take advantage of a wireless network, explore information about network data traffic and what application mix is used. They found that the community they analyzed can be broken down into sub-communities, each with its own unique behavior when users are active and how much traffic the users generate. The study also finds that web-surfing and session application such as *ssh* and *telnet* are the most popular applications. Moreover, Tang and Baker [8] also characterized user behavior in a metropolitan-area network in order to find how users take advantage of a mobile environment. The network showed different performance characteristics valid for this particular network, but they can be usable for gaining an understanding of mobile network behavior in future analysis. The traces were obtained from a packet radio

infrastructure in the areas of San Francisco, Washington DC, and Seattle. They consisted of a seven week period during which no registrations were logged. Different patterns of mobility constituted an important goal analyzed in this study, for example, users who just use the radio at home and at the office. The result showed that not all the users execute their particular mobility pattern at the same time, but they can be divided into time patterns. The study is a previous demonstration about how people can take advantage of a mobile environment and they found that the movement of the users is a Gaussian distribution around the radius of the network.

Kotz and Essien [9] expanded upon their Wave LAN study by Tang and Baker. The study presented results from the largest and most comprehensive trace of network activity in a large, production wireless LAN, including all administrative, academic, and residential buildings. Their trace included detailed information about the amount and nature of the network traffic and the study has characteristics common to both residential and enterprise deployments. However, their residential university campus population may not reflect activity in a public space. The traces were obtained from Dartmouth College during 12 weeks and the study is based on understanding patterns of activity in the network. They found that many wireless cards are extremely aggressive when associating with access points, leading to a large number of short sessions. Their work concludes that there exists a need for solutions to support multi-subnet roaming.

Furthermore, Henderson and Kotz returned to the same network after it had matured in 2003/2004 [10]. The study presents the results of the largest WLAN trace to date, and the first analysis of a large, mature WLAN to measure geographic mobility as well as network mobility. They found dynamic increases in usage, and changes in the applications and devices used on the network. They found that the applications used on the WLAN increased in peer-to-peer technologies, streaming multimedia, and voice over IP (VoIP) traffic. The study shows that VoIP has been used little on the wireless network, most VoIP calls are made on the wired network and calls last less than a minute. On the other hand, they found great heterogeneity in the types of clients, such as PDAs and mobile VoIP clients, which are wireless devices and have different mobility characteristics than laptops.

The previous studies outscored above have presented patterns of user mobility and network usage characteristics for one particular domain. However, Balazinska and Castro [11] complement these studies by presenting results from a four week trace

gathered on a corporate WLAN. They focus on load distribution across access points, population characteristics, user level of activity and user mobility. One important result from this study specifically refers to user access, that is clients use wireless services in a fraction of days and a fraction of time. They propose two user models to classify their mobility, which are persistence and prevalence. The first one means how long is a user AP association and the second one reflects how frequently a user visits various locations. The resulting users spend most of the time at a single location, but periodically visit other locations.

## 2.2 Hotspot Environment

Because the wireless hotspots are widely used in business and research, they are an attractive scenario to analyze. However, there is not a large number of studies concerning this field. Blinn and Henderson [12] have studied a wireless hotspot network examining five weeks of SNMP traces from the Verizon Wi-Fi HotSpot network in Manhattan. They analyzed the network in terms of users, access points (APs) and traffic. This analysis seeks to expand the previous studies, finding that most users access the network infrequently and spending their time at a single AP. Although half of the traffic was caused by 5% of valid users, it varied across days and exhibited unusual hourly characteristics. As in previous analysis, the authors found that users access the network infrequently and they spend most of the time in a single AP. Their research has tried to find similarities between this study and campus datasets, but operator hotspot data is somewhat harder to obtain than campus WLAN data, so there were limitations concerning what users were doing.

Balachandran, Voelker and Bahl [13] have observed how the mobile computing landscape has changed both in terms of the number and type of hotspot venues, presenting experiences of commercial hotspots in the areas of security, coverage, management, location services, and interoperability. These constitute deployment-related problems to provide Wi-Fi Hotspot connectivity. The study considers the performance benefits of WLANs as a platform for networking in public places and the authors discuss the experiences of commercial hotspot providers in the context of the needs of a typical business traveler. The conclusions of the authors suggest improving end-user facilities in order to offer end-users an easy and fast mechanism to access the network in a transparent way and independent manner. That means a hotspot network must provide a reliable and robust third party authentication to serve the performance demands of the users.

Afanasyev, Chen and Voelker [14] have presented a study about the usage of the Google WiFi network deployed in California using different client device types such as laptops and WiFi-capable smartphones like the Apple iPhone. They use a trace of 28 days to analyze the temporal activity of clients, accounting information collected by the central Google WiFi RADIUS server. The study shows traffic demands on the network, the mobility, the diversity and coverage of users. The authors analyze an urban environment with several access technologies, including wired broadband networks, 3G cellular, and WiFi Hotspots. The users are characterized into three different populations, residential, commercial and transportation. They limit their traffic analysis to high-level application classification based on protocols and port numbers. The study found that static modem users place the highest demand on the network, Hotspots were concentrated in public areas with moderate mobility and smartphone users were numerous and concentrated in travel corridors.

McNett and Voelker [15] analyzed the mobility patterns in a campus wireless network during 11 weeks of network activity. An interesting observation about this study is the analysis of traces of user-oriented PDAs. The authors characterize the high-level mobility and access patterns of PDA users, evaluating the implications of these mobility patterns in new wireless communication architectures like ad-hoc networks. The authors found a variety of wireless applications, managing PDAs and the number of PDA users was greater than laptops and mobile phones in terms of access points monitored during the same time period.

Karvonen and Lindqvist [16] have published a paper on a study and usability analysis of WLAN access to perform the user's needs and the problems using the WLAN network offered by the hotspots. The authors present a study to answer to how usable and secure is the network offered by hotspots and they focus on whether the services offered to users are simple, usable and have good quality. Furthermore, they present and evaluate the methodologies used to study the usability of the hotspots and other types of WLANs and discuss usability issues in the controllability and visualisation embedded in current approaches. The study is presented with relevant work done in several locations in two cities in Finland, Helsinki and Espoo. The authors found several usability issues in the UIs handling of WLAN access management and they intend to do testing with real users on the new design methodology.

### 3 Wireless Access in Public WLAN Hotspots

This chapter presents an overall description of wireless local area networks (WLANs). An introduction is given on relevant terminology, which is important to understand the trace data analysis. Furthermore, some limitations and privacy aspects concerning the wireless traffic collection are described to inform the reader about the drawbacks and relevant issues in the process.

Wireless LANs are usually designed to operate in license-free radio frequencies making their operation and maintenance cost less than cellular and Pc networks. The key advantages of wireless networks [17] as opposed to wired networks [18] are mobility, flexibility, ease of installation and maintenance, and reduced cost. The Helsinki establishment of the public WLAN network cost from around EUR 60,000 to 70,000 in extra equipment in 2009, the annual maintenance costs are estimated to be some EUR 20,000 to 25,000.

Figure 3.1 shows the WLAN access from the point of view of a user when he wants to retrieve content from the Internet. The client has to initiate a wireless access process in order to establish a WLAN session between the user device and the AP selected to offer the service. The public access points use session initiation where there is no need to perform any kind of user authentication to gain access to the service. The hotspot maintains an Internet domain connection with the wired network.

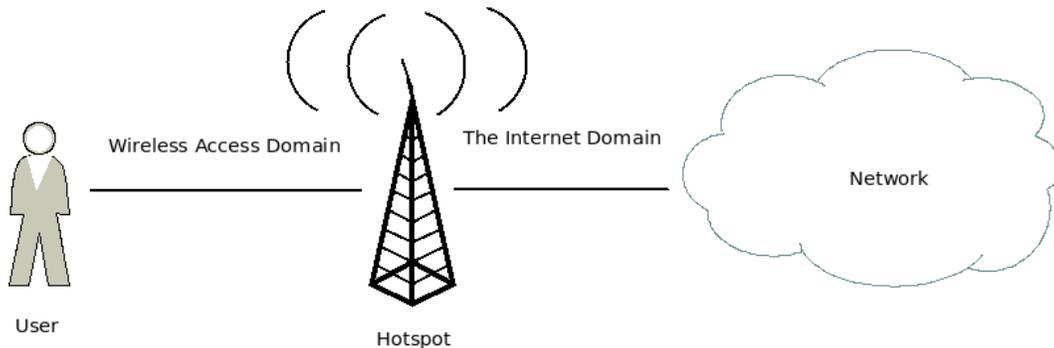


Figure 3.1: An example wireless access scenario.

All wireless devices are constrained to operate in certain frequency bands. The use of radio spectrum is rigorously controlled by regulatory authorities through licensing procedures. In Europe, the newly established Electronic Communications Committee (ECC) (CEPT [19]) is now responsible for spectrum allocation.

To achieve interoperability between WLAN devices supplied by different vendors,

the Institute of Electrical and Electronic Engineer (IEEE) designed the 802.11 standard, which had significant enhancements over WaveLAN, which was its predecessor. There include support for acknowledgments and retransmissions, contention free transmission using reservations, and an operating mode where a master host provides WLAN co-ordination.

### 3.1 Introduction to IEEE 802.11 Technologies

The IEEE 802.11 standard [1] specifies a 2.4 GHz operating frequency with data rates of 1Mbps and 2Mbps. In late 1999, the IEEE released two supplements to the IEEE 802.11 (1997) standard: IEEE 802.11a [20] and IEEE 802.11b [21]. The initial 802.11 standard mainly uses two types of physical layers: IEEE 802.11 FHSS (frequency hopping-spread spectrum) and IEEE 802.11 DSSS (direct sequence-spread spectrum). IEEE 802.11b is a data rate extension of the initial IEEE 802.11 DSSS providing data rates up to 11 Mbps in the 2.4 GHz ISM band using Orthogonal Frequency Division Multiplexing (OFDM) [22].

The WLAN technology discussed in this thesis is defined by the IEEE 802.11 family of specifications. There are currently four specifications in the family: 802.11, 802.11a, 802.11b, 802.11g. All four use the Ethernet protocol and carrier sense multiple access with collision avoidance (CSMA/CA) instead of Carrier Sense Multiple Acces/Collision Detection (CSMA/CD) for sharing access to a wireless medium.

- 802.11 – applies to wireless LANs and provides 1 or 2 Mbps transmission in the 2.4 GHz band using either the frequency hopping spread spectrum (FHSS) [23] or the direct sequence spread spectrum (DSSS).
- 802.11a – an extension to 802.11 that applies to WLAN provides up to 54 Mbps in the 5GHz band. 802.11a uses an orthogonal frequency division multiplexing (OFDM) encoding scheme rather than FHSS or DSSS. The 802.11a specification applies to wireless ATM systems and is used often in access hubs.
- 802.11b (also referred to as 802.11 High Rate or Wi-Fi) – an extension to 802.11 that applies to wireless LANS and provides 11 Mbps transmission (with a fallback to 5.5, 2 and 1 Mbps) in the 2.4 GHz band. 802.11b uses only DSSS. 802.11b was an extension to the original 802.11 standard, allowing wireless functionality comparable to the Ethernets standard.

- 802.11g – offers wireless transmission over relatively short distances at 54 Mbps in the 2.4 GHz band. The 802.11g also uses the OFDM encoding scheme.

For short range (less than 10 meters) and low power wireless communications among personal devices such as PDA and Bluetooth and subsequent IEEE standards (802.15) are taking effect. For long range wireless communications in the metropolitan areas, WiMax and IEEE 802.16 are the emerging standards.

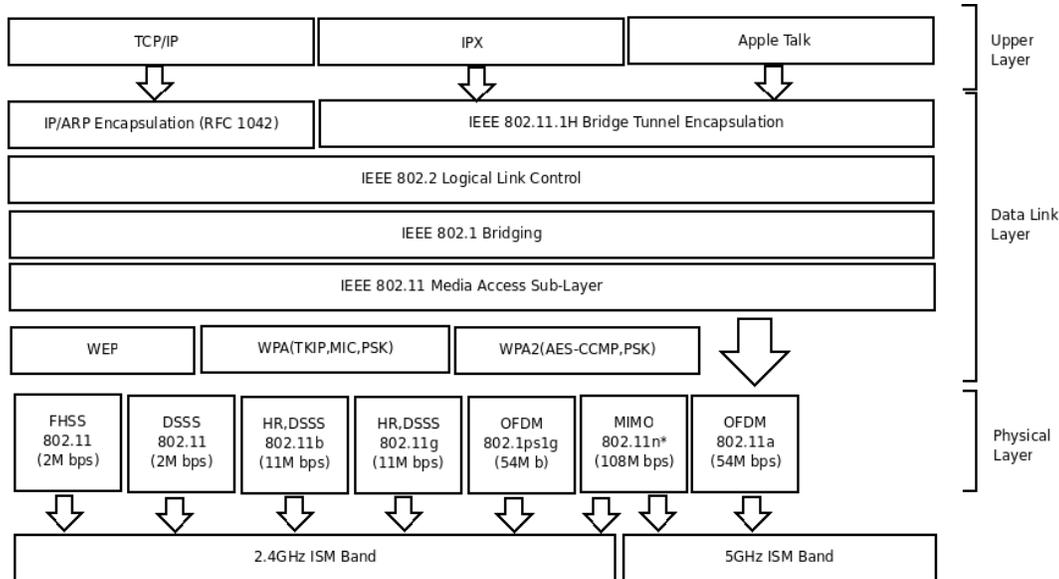


Figure 3.2: IEEE 802.11 Protocol Layers.

Figure 3.2 shows the layers in the IEEE 802.11 specification: layer one is called the Physical layer (PHY) and layer two is called the Media Access Control (MAC) layer. Layer one specifies the modulation scheme used and signaling characteristics for the transmission through the radio frequencies, whereas layer two defines a way of accessing the physical layer [24] and defines also services related to the radio resource and the mobility management. The standard defines three different physical layer specifications for WLAN: one infrared and two RF transmission methods, which are a direct sequence spectrum (DSSS) and a frequency hopping spread spectrum (FHSS).

The DSSS uses Different Bi and Quadrature Phase Shift Keying (DBPSK and DQPSK) modulation techniques. However, the FHSS uses 2-4 level Gaussian Frequency Shift Keying (GFSK) as the modulation schemes. Depending on the modulation scheme used, both the direct sequence and frequency hopping spread spectrum support data rates of 1 Mbit/s and 2 Mbit/s.

In order for wireless LAN devices to be interoperable, they have to conform to the same physical layer standard. Therefore a DSSS device is not capable of communicating with a FHSS based device.

## **3.2 Limits of Wireless Networking**

The main purpose of using wireless LANs is to take advantage of flexibility for serving mobile users. They do not replace fixed networks, and it is clear that devices such as servers and other data center equipment remain to be fixed. A brief description of the possible drawbacks of the wireless technology is provided below. These drawbacks can be potential barriers to making a good capture of information for later in our analysis.

### **Multipath propagation or attenuation**

Multipath propagation is a phenomenon where the receiver may receive multiple copies of the same transmitted signal due to reflection from physical objects such as office furniture or hills, etc. The phenomenon can cause signal corruption or interference. The delay spread is a term which describes the amount of delay related to the primary signal. The delay spread value indicates how distorted and possibly undetectable the signal is. This problem, however, is more pronounced in indoor usage scenarios despite the fact that the delay spread is smaller in indoor systems. Communications engineers often use signal-processing techniques such as rake receivers (radio receiver to counter the effects of multipath propagation), equalization, or antenna diversity in order to mitigate this problem [25].

### **Path Loss**

Path loss is the loss of power of the radio signal traveling (propagating) through space. It is expressed in dB. Path loss depends on the distance between transmitting and receiving antennas, the line of sight clearance between the receiving and transmitting antennas as well as the antenna height.

## **Radio Signal Interference**

Most of the wireless local area networks operate in the ISM bands. The industrial, scientific and medical (ISM) bands [26] are publicly accessed, and can be affected by interference caused by another system on the same frequency range, external noise, or some other co-located system. For example, microwave ovens and Bluetooth networks are the two most common sources of interference for IEEE 802.11b devices. The interference degrades the performance of wireless networks [27].

## **Limited Battery Life**

A wireless device is typically also a mobile device. Since mobility dictates compactness in size, and since there is no wired power connection, batteries are often the only means of power supply. Even the longest lasting batteries offer a very limited amount of power.

## **System Interoperability**

With wireless LANs, interoperability is taken as a serious issue. There are still pre-802.11 (proprietary) wireless LANs, both frequency-hopping and direct sequence 802.11 versions, and vendor-specific enhancements to 802.11 compliant products that make interoperability questionable. To ensure interoperability with these enhancements, it is often needed to use radio cards and access points from the same vendor. An organization, the Wireless Ethernet Compatibility Alliance (WECA) [28], ensures compliance among IEEE 802.11b wireless LANs through its Wi-Fi (Wireless Fidelity) test.

## **Network Security**

Security [29] is a big concern in wireless networking, especially with commercial applications. The mobility of users increases the security concerns in a wireless network. Current wireless networks employ authentication and data encryption techniques on the air interface to provide security to its users. Due to radio wave propagation the wireless network boundaries cannot be well defined or restricted which may allow an unauthorised person to access or connect to the network. In a large enterprise, an IP network level security solution can be used to ensure that

the corporate network and proprietary data are safe. The use of a virtual private network (VPN) is an option to make access to the fixed network reliable.

## Connectivity Problems in the Application layer

The use of traditional Internet protocols over wireless networks can introduce problems [30] with maintaining connections between the client applications and the applications residing on a server. Generally, the response time of wireless networks is higher than in wired networks and wireless networks often need to request retransmission due to data loss or corruption. These problems make wireless network less reliable in comparison to wired networks.

### 3.2.1 Issues with TCP/IP over Wireless LANs

Currently, most of the wireless LANs implement TCP/IP as the communication protocol. TCP/IP based protocols provide an excellent platform for high-speed wired LANs with constant connectivity; however, the use of TCP/IP over wireless LANs poses significant problems [31]. Some of them are discussed below.

- *High overhead*: Because TCP packets are cumulatively acknowledged as they arrive in sequence, out of sequence packets cause duplicate acknowledgements to be generated which do not contain real data. This additional overhead can cause bad performance over the wireless LAN due to the limited wireless bandwidth.
- *Incapability to adjust under marginal conditions*: the TCP protocol does not have further capabilities to re-establish the connection. TCP/IP protocol can terminate the connection between the AP and the wireless user device when a marginal connection (i.e. connection with very low bandwidth) exists causing an application or user re-connection intervention.
- *Difficulty in dealing with mobile node addresses*: The Mobile IP (or IP mobility) protocol [32] allows location-independent routing of IP datagrams on the Internet. Each mobile node is identified by its home address, disregarding its current location in the Internet. A home agent stores information about mobile nodes whose permanent address, is in the home agent's network. A foreign agent stores information about mobile nodes visiting its network. Foreign agents also advertise care-of addresses, which are used by Mobile IP. A

node wanting to communicate with the mobile node uses the home address of the mobile node to send packets. These packets are intercepted by the home agent, which uses a table and tunnels the packets to the mobile node's care-of address with a new IP header, preserving the original IP header. The packets are decapsulated at the end of the tunnel to remove the added IP header and delivered to the mobile node. Problems arise when a user device associated with an *IP address* roams to an access located in a different network domain that is separated from the original network domain by a router. This can confuse the router and possibly other devices located within the new network domain which can result in a network that cannot route packets.

### 3.3 Security Issues in WLAN Access

Wireless LAN security is work that is continuously under progress. Research indicates that the perceived insecurity of wireless networks is a major inhibitor to further market growth [33]. An institution wireless internal network will require strong user authentication to prevent unauthorized users, as well as strong privacy protection to keep information confidential. However, commercial hotspot networks do not need to provide privacy protection, but do need to restrict the use of the network to paying customers. Wireless APs must announce themselves using beacons frames, used to broadcast network parameters. Monitoring these beacons frames make easy finding a wireless network.

WLAN access applies security actions using a link-layer mechanism, which means that radio traffic is encrypted between the access point and the user device. The 802.11 standard provides two types of authentication (open system and shared key), and a privacy method (WEP). They do not support either end-to-end or user-authentication. The open system authentication is by default a null authentication service, i.e. no authentication at all. Any client can join the network. The second one is the shared key authentication, which is used to authenticate mobile stations joining a network that share the same secret key with an AP.

Privacy issues remain in Wired Equivalent Privacy (WEP) [34], which was the first link-layer security architecture used, but it was proven to be insecure. WEP was not designed for high security, but rather to be at least as secure as its wired counterpart. Consequently, the security problem was solved by a WLAN alliance without using the IEEE standardization body. The result was the Wi-Fi Protected Access (WPA) protocol which corrects the WEP deficiencies.

## Wireless Security Threats

Attacks on the wireless network can come from internal or external sources and they can be passive or active attacks. In passive attacks an unauthorized party gains access to a network and does not modify the resources on the network. An attacker simply collects various WLAN transmissions (eavesdropping) or monitors the traffic for communication pattern analysis. In active attacks, the attacker can try to modify the network resources, for example, by spoofing first the MAC address and/or IP address to gain network access. Types of attack include:

- *Masquerading*: term used to refer to a process where an attacker obtains certain unauthorized privileges (the use of spipofing, rogue APs, and redirections attacks).
- *Replay*: An attacker monitors the traffic and later retransmits the same message as a legitimate user.
- *Message Modification*: An attacker alters a part of the message by deleting, changing or adding. In addition, an attacker may modify some device configurations.
- *Denial of Service (DoS)*: An attacker can inject a large amount of traffic or use malicious commands to alter the common network management.

## Infrastructure countermeasures

WLAN signals can reach more than 500 meters, that means, some techniques such as wardriving/warwalking [35] can be used to attack vulnerable networks. The equipment needed is again a sniffing software downloaded from the Internet and a laptop or even a PDA. WLAN networks are secured in two common ways: on the one hand, the MAC address based authentication allows only known computers to join the network. However, attackers can eavesdrop the MAC and they reconfigure their WLAN devices accordingly. On the other hand, web based authentication can be required to authenticate users to the network. The user is required to give a correct user name and password, and is then given access to the Internet.

Basic countermeasures involve altering the security functions that are provided with the wireless equipment. These provide solutions such as the Remote Authentication Dial In User Service (RADIUS) protocol [36] or Virtual Private Networks (VPN)

[37]. VPN applies the existing wired network mechanism such as IP Security (IPSec) tunneling, and the 802.1x standard [38], which require user authentication before granting network access and incorporate the port-based network access control for 802.11 infrastructures. Even if 802.1x is imperfect, it is considered as a better user-authentication solution than WEP ever was [39]. 802.1x clients are now becoming available for many popular operating systems.

### 3.4 Data Mining and Privacy Policies

Privacy is a topic that is much debated today. Due to the rapid increase of personal information usage in the Internet and development of data mining techniques, the resulting privacy concerns have become a well known topic of discussion. Data mining tools produce data for marketing products and services to targeted customers. Furthermore, it often gives access to a large amount of personal information such as medical history, purchase transactions, and telephone usage statistics, among other aspects of the users personal lives. In recent years, data mining has been widely used in the area of science such as genetic engineering, medicine, education and electrical power delivery. The customers may not know that their shopping habits, names, address, and other information are being stored in a database. The data mining process has a number of advantages, but there are some ethical issues that are often raised.

The procedure that should be followed to provide security anonymity of users should answer question such as the purpose of the data collection, existence of any data mining projects, how the data will be used, who will be able to mine the data and use it, the security surrounding access to the data and how the collected data can be updated avoiding anonymity risks. The term data mining in this study has been used to search for apparent representative patterns in a large amount of data, while maintaining privacy concerns.

Haib and Dressler [40] discussed the requirements for anonymization techniques in the domain of network measurement and monitoring. The study reveals necessary privacy regulations in terms of law, and potential issues in allowing a publication of packet traces that are still comparable. In this thesis, we consider data payload as anonymous and base our Analysis of just header information. The IP address is the piece of information in question that can be related to a natural person, however, considering IP address belonging to the data processing entity's domain, this poses no problems. Because the dynamic IP address changes each time a host connects to

the Internet, it is more difficult to establish an on pattern profile user or to collect data of the same host because it cannot be re-identified. Therefore, many authors suggest that dynamic IP addresses are not capable of being linked to the user by others than the Internet access provider.

Privacy versus content tradeoffs may need to be made regarding data analysis to ensure security. Networking researchers and engineers rely on network packet traces for understanding network behavior, developing models, and evaluating network performance. In order to publish packet traces, organizations have to employ techniques that remove sensitive information before making them available. The most common process of anonymizing a packet trace involves removing a packet's payload and replacing source and destination IP address with anonymous identifiers [41]. By removing the sensitive information, the packet trace can be shared with others (for debugging or network analysis). Without data information and IP address, it becomes difficult to extract sensitive user information from the trace, but not impossible. The use of this information starts to be malicious or illegal when for example there are a marketing, surveillance, censorship or industrial espionage purposes among others. Kouskis, made an experimental study to assess the risk of publishing anonymized packet traces [42]. They investigate whether it is possible to break the address anonymization scheme and identify specific host addresses in the trace by using attacks based on web site footprints or port-scanning activity. Associating the inferred information with other sources, such as a web server's log files, could lead to significant privacy problems.

The application used in this thesis for tracing data anonymization has been the Anon Tool, an open-source implementation of an anonymization API [43]. This framework is composed of an anonymization application programming interface (AAPI) which offers anonymization primitives. It can specify anonymization policies with varying levels of granularity needed by the user. It is also easy to support anonymization for new application level protocols and different traffic sources such as Netflow. The Anon tool is more sophisticated and faster than tcpdpriv, which offers less functionality.

### **3.5 Wireless Network Access**

Hotspots provide local coverage for authenticated traveling users who want to connect to the Internet. This allows packets to flow between the client and the attached wired network. In order to access the Internet by using WLAN, it is natural to use

TCP which is the most used transport protocol in the Internet. Users that connect to the Internet often use several TCP connections in the same WLAN session. The amount of TCP traffic consist of numerous network applications that establish a session between the client and the server. However, TCP was not primarily designed to be used in wireless links, such as WLAN. RTT variation in wireless network can cause temporary TCP disconnections. This section describes important concepts related to wireless Internet usage in the scope of this thesis.

### 3.5.1 TCP Flow

The Transport Control Protocol (TCP) [44] is a reliable protocol designed to perform well in networks with low bit-error rates, such as wired networks. It is a reliable transport protocol that uses a retransmission mechanism. TCP assumes that there is a congestion when retransmissions have been frequently received. It adjusts its windows size (and transmission rate) when congestion is encountered, and retransmits the lost packets. In WLANs, however, packets loss is often caused by high bit-error rates air-links. Thus, the TCP window adjustment and retransmission mechanisms result in poor end-to-end performance.

TCP is a connection orientated protocol that provides reliable data delivery. A single TCP flow consist of several packets exchanged as part of a TCP connection between a client (i.e. the host that opened the connection) and a server (i.e. the host that accepted the connection). The TCP connection is identified by two endpoints (IP address and TCP port). In addition, several connections can share a single TCP port. The flow size can be consider as the amount of data flowing from the server to the client or in the other direction depending on who initiated the TCP connection.

TCP flow is constituted by a connection establishment, data transmission and a connection closing. This entire process can be called a complete connection. However, sometimes there exist TCP connections which have not realized the complete process. The TCP connection is established with a three part handshake task (SYN,SYN-ACK,and ACK). This packet exchange make it possible to fix the initial sequence number which will be stored in each endpoint. In addition, an incremental sequence number is associated with each acknowledgement (ACK) informing the sender about the last packet that was successfully delivered. If the sender receives the same ACK sequence number more than once, this can suggest that the data since the last byte is lost. The TCP procedure connection closing is full-duplex, so both directions need to be closed simultaneously.

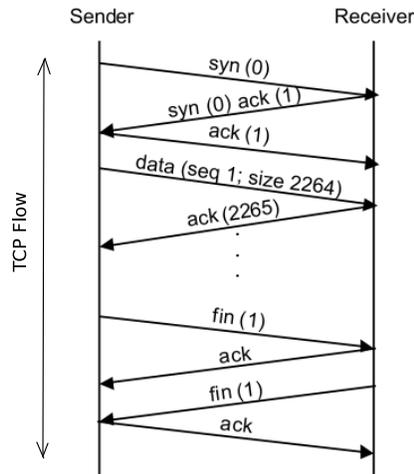


Figure 3.3: TCP flow control.

Figure 3.3 shows an example of TCP flow in which a user connects to a remote host. When a connection is established, each end allocates a buffer to hold the incoming data, and sends the size of the buffer to the other end. As data arrives, the receiver sends an ACK request, which is called a window advertisement, with the amount of buffer space available. The sender forwards the amount of data concerned (windows size), but if there is an error on the wireless link, one of the ACK packets from this data may not be received. Consequently, the rest of the packets can be duplicate ACKs. The sender detects a loss when multiple duplicate acknowledgements (usually three) arrive.

### 3.5.2 Round Trip Time

TCP detects losses using a time-out mechanism. Retransmission timers are continuously updated based on weighted average of previous RTT. It is also called round-trip delay, is the time required for a signal or packet to travel from a specific source to a specific destination and back again. The source is the host transmitting the packet and the destination is the remote computer that receives the packet and retransmits a response back to the source. TCP estimates with each received ACK the maximum RTT of the ongoing connection (called Retransmission TimeOut, i.e., RTO) and a timer with the estimated value. The result depends on various factors including the transmission medium (copper, optical fiber, wireless or satellite), the physical distance between the source and the destination, the amount of traffic on

the LAN to which the end-user is connected, the remote server, the speed with which intermediate nodes and the remote server function, as well as the presence of congestion in the links or interference in the wireless interface.

When a client sent packets over a TCP connection, the protocol predicts the future RTT by averaging earlier RTT samples into a smoothed round-trip time estimate (SRTT). The client uses a sequence of round-trip samples (latest RTT for acknowledged message):  $S(1)$ ,  $S(2)$ , ... to compute the new SRTT, as shown by Equation 3.1:

$$SRTT(i + 1) = \alpha * SRTT(i) + (i - \alpha) * S(i) \quad (3.1)$$

where  $\alpha$  is a constant between 0 and 1 that controls how rapidly the SRTT adapts to changes. RTT times are swift Poisson distributed, but with brief periods of high delay. If the SRTT had not swift enough adaptation, the TCP sender would unnecessary retransmit packets because the RTO was set too low.

For unreliable mediums, such as radio, *Karn's Algorithm* [45] can be applied because many retransmissions required producing many duplicate between the ACK. Karn's Algorithm does not require using retransmitted segments for updating the round trip estimate. Round trip time estimation is based only on unambiguous acknowledgments, which are acknowledgments for segments that were sent only once to avoid the retransmission ambiguity problem.

The WLAN environment consist of numerous link technologies, which makes it difficult to understand the latency fluctuations. This thesis consider RTT estimations to perform TCP connection analysis of a high level. In addition, web page retrieval delays are an important form of analysis of traffic patterns.

### 3.5.3 Sessions

Sessions can be defined in a specific manner depending on the protocol layer. This section defines some of them.

#### **Application sessions**

An application session is defined on a high level, that is, in the application layer. We define a session as a sequence of connections that have a common point of origin. The session involves a single local-host and a single remote-host, so implicitly all

connections require only these two hosts. For example, a particular user interacts with a server to retrieve a piece of web content. During a web application session, the unique identity of the user is maintained internally and it serves as a way to transport and maintain user data in web pages, such as forums, or e-commerce websites. The application associated with a session is of the type of the first connecting application. Figure 3.4 shows an application session example, where a client request a web page.

Web pages usually consist of many HTTP requests each for a small size document (image, text or script). The practice with HTTP/1.0 was to use a separate TCP connection for each HTTP request and response [46]. This, however, causes slow-start latency on each request because of the slow connection establishment [47]. Using persistent connections addresses this problem by reusing long-lived TCP connections, which were introduced with HTTP/1.1. The process required for exchanging a HTTP message request/response pair between the client and the server needs to map a domain name of the server's IP address (DNS query) in order to establish the TCP connection, which is used by the client and the server to send the HTTP messages. The HTTP transfer process consist on a client sending a request to the server over the connection and waits to read the server response. Thus, before responding to the request, the server may add latency due to generation of dynamic content, disk I/O, or reverse DNS query. A typical web page consists of several parts such as graphics or applets called application objects. Fetching each of these objects requires a fair number of communication cycles between the user's system and the web site server, each of which is subject to the RTT delay. Several connections can be required to response all the web page objects. Session objects can preserve user preferences and other user information in the web application.

Different applications vary widely in the behavior they exhibit from the point of view of the session structure. They can be singleton or homogeneous sessions, which are mainly started by SSH, FTP, SMTP, or HTTP client applications.

### **WLAN usage session**

The WLAN access is provided in public areas and it is available during continuous periods like in airports or during limited opening hours like in libraries. A WLAN usage session is the time a client is associated with an access point. Normally, there is no time limit on wireless sessions. However, the network connection can end after a time with no network traffic. When a user attempts to access a service through a

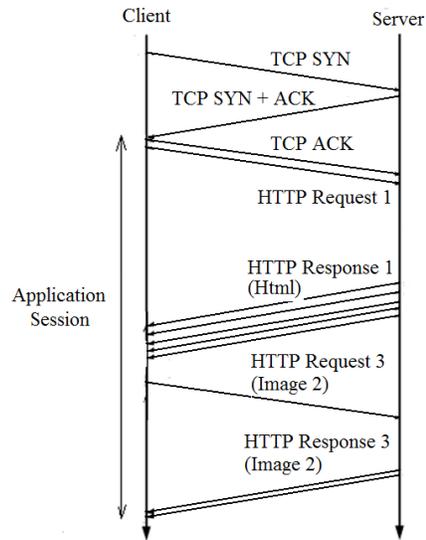


Figure 3.4: An example web application session.

public WLAN, the WLAN first authenticates and authorized the user access prior to granting network access. A session starts when the user send an Association request frame to the AP and this reply with an Association Response frame. Public locations often allows APs with unrestricted access to the network. Clients with wireless capability can use Internet without authentication and establish a WLAN session. Next, DHCP (Dynamic Host Configuration Protocol) [48] server in the AP grants network configurations automatically. The DHCP server gives IP addresses dynamically to clients when they become associated with an AP or in the other case, where the addresses lease becomes expired. For this reason, we can find different users using the same IP address in the captures and on the other hand, the same users using different IP addresses. The clients can use the Internet service for several purposes such as accessing the web or file transfer services. The user services are served by several connections within a single WLAN session. When the user decide not to use the network anymore, he may send a Disassociation frame to the AP.

In some cases, we may not capture all the packets required to determine the start and end of user session. In absence of Association frames, we started a new session whenever a packet from a new user was notices in the trace. Similarly, in the absence of Disassociation frames, the end time for the session was set to the time of the last packet seen.

## Summary

IEEE 802.11 standard characteristics and observations of the drawback issues in this technology bring us a background for capturing and analyzing traffic. These drawbacks are considered to avoid potential barriers to making a valuable capture of information. In addition, the perceived insecurity of wireless networks is a major issue debated today. Commercial hotspot networks do not need to provide privacy protection, therefore the procedure followed to offer security anonymity should avoid user privacy risk. The traffic characterization and monitoring is useful information to gain an understanding of application sessions in the network traffic. This basic WLAN session and application session knowledges can bring rich information for gaining insight into how the network is used. In addition, a characterization of network traffic can have different forms. Some of them are explicitly defined with sessions such as HTTP, FTP, etc, which arises from how end-user software, or the end-users themselves, drive network access and usage in order to perform tasks.

## 4 Capturing WLAN Traffic

In this section, we describe the procedure used to collect the trace data as well as the capture methodology. We also provide a global summary on data analysis. This applies as a general summary about the traffic collected in Helsinki and the tools used in the study. The collected data is presented in detail in order to explain the behavior in each of the places of the study. We specify the places where captures were collected and the most relevant free WLAN networks in each of them.

### 4.1 Challenges of Wireless Monitoring

It is difficult to guarantee that a single sniffer can capture all wireless packets between two end-points (user and AP). Many losses can occur as a result of signal strength variability, card variability or a combination of both. Due to these losses, a capture device is involved in a challenging problem of performing effective wireless monitoring.

Different sniffer positions have different viewpoints of the wireless medium. Jihwang Yeo, et al. [49] have explored the issues of monitoring traffic in IEEE 802.11 wireless networks. We have used their finding as input for our experiments in order to improve the capturing performance. The sniffer loss for from-AP traffic may be less than the loss for to-AP traffic. One explanation is that the AP has better hardware compared to clients and because of that, the signal seen by a sniffer from an AP is stronger than the signal seen from the client. Another aspect to consider is related to the physical location between the observed wireless client and the sniffer, which can affect the data collection from the particular user. If the distance between them increases, the signal decays, which can decrease the sniffing performance. If the power is very low for some clients, it can cause the packages to be successfully transferred from the AP towards the client, but the transmissions of the client are out of the range of coverage of the sniffer. This means that the sniffer can listen to only half of the communication. Figure 4.1 shows a wireless monitoring example, where we can see some associated users (A,B and C) and a possibly bad capturing client placement (B) in the trace collection. However, if the capture device is closer to the AP, it can capture frames that this badly situated client cannot. In addition, clients can fail to adapt to the capture bandwidth because of the signal strength between them and the access point.

In order to improve the capture performance, it could be feasible to merge the data

from different sniffers to obtain a better view of traffic and select the sniffer locations to achieve an acceptable capturing performance.

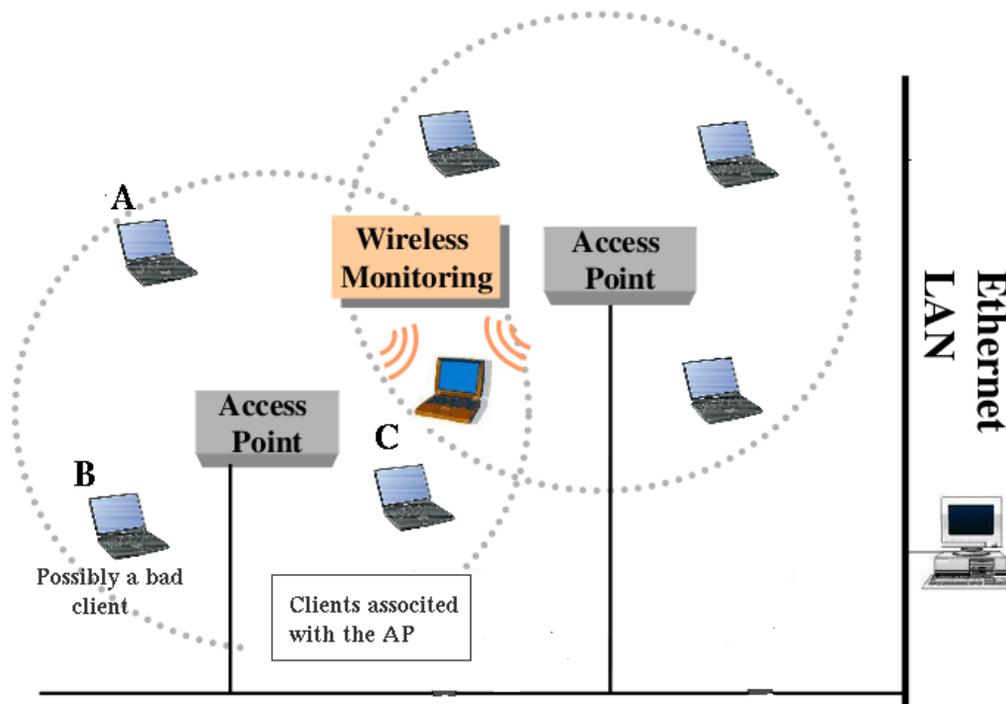


Figure 4.1: Monitoring wireless traffic, illustration based on [49].

## 4.2 Data Collection

The free Wi-Fi networks analyzed in the study consist of 14 APs distributed around the city of Helsinki. The experimental study in this thesis is based in a trace collection in different free hotspots in Helsinki. These APs have been chosen as a result from previous traffic captures, which were performed in several locations. The captures enabled establishing a criterion to reduce the capturing radius to obtain the relevant traffic from the measured APs. Some of the locations were not good enough to evaluate the important measures in the study, because of the low number of users and low amount of data traffic. In addition, some APs provided bad signal quality. Consequently, the study was reduced to collect traffic from three strategic places of the city, which include diversity in the number of APs and the possible users enjoying the networks. Table 4.1 shows the number of hotspots evaluated in each location, which have been chosen for evaluating the user diversity, being workers, students, traveling or statics users and the name of the WLANs chosen in

each location. The wireless traces consist of three different capture data sets from the hot locations.

Table 4.1: Main locations.

<b>Locations</b>	<b>Access Points</b>	<b>WLANs</b>
<i>Kamppi</i>	4	<i>Helsingin Kaupungin</i> and <i>Lasipalatsi</i>
<i>Library-Post</i>	6	<i>Helsingin Kaupungin</i> , <i>Hsverko</i> and <i>Stadinetti</i>
<i>Stockmann</i>	4	<i>Helsingin Kaupungin</i> and <i>Kafka</i>

We captured traffic data for approximately a 50 hours period between June 2009 and September 2009. More specifically, 10 hour capture was collected for each of the hot places in different times periods during a day and on different days. The captures were collected during a continuous period of 2 or 3 hours. The collected WLAN traffic came from the most relevant WLAN networks located in the place, and the traffic was from several users connected to these hotspots. The networks do not use MAC-layer authentication in the APs. Any client can associate with any free access point, and obtain a dynamic IP address. All hotspot do not share the same network name (BSSID), allowing wireless clients to obtain a new IP address from one AP or another. The APs are exposed to cover outdoor spaces, but also many indoor buildings have coverage. Because of the number and geographical distribution of APs, the structure of the networks, and the volume of the traffic, it was not possible to capture all of the wireless traffic. Appendix 6 (A.1, A.2 and A.3) shows the WLAN traffic summary for each of the main locations in this study.

The traces were collected with the *Airodump-ng* tool [50] which collected 802.11 frames from the free APs for later analysis. Some packets were dropped by the operating system kernel due to a lack of buffer space in the packet capture machine, which was running the Ubuntu operation system.

Table 4.2 shows the the WLAN traffic statistics analyzed in the study. In addition, the total traffic column shows the amount of data traffic captured in each of the zones where the traffic was collected. As can be seen, each of the WLAN networks presents a specific fraction of the captures. The percentage reveals which of the networks were used the most by the clients to transmit data packets. On one hand, the traces have been captured following the same patterns. The sniffer was in a specific position where we could find the best traffic but this can be relative depending on the clients position (see Section 4.1). More precisely, the rest of the captures in the study come from these strategic positions. On the other hand, the distance between the

hotspots and the sniffer was different in each location due to the signal quality. The city of Helsinki has extensive 802.11 coverage within which clients can connect to the Internet. The users in our trace resided approximately within a 200m radius area located around specific points, which offered wireless coverage from these networks.

Table 4.2: Summary of collected WLAN traffic.

Locations	WLANs	Traffic Percentage(%)	Total Traffic (MB)
<i>Kamppi</i>	1 and 2	70 and 30	449.28
<i>Library-Posty</i>	1,2 and 3	74.32 19.466 and 2.48	198.4
<i>Stockmann</i>	1 and 2	74.32 and 19.46	550.4

*Helsingin kaupungin WLAN* contains the most relevant free network based on the quantity of traffic transmitted during the trace collection. This free WLAN is established in several places, however, some other free WLANs were more used by the clients in the places where data was collected due to better quality of the signal. That is a likely the reason for that we can find a high percentage by users using the *Stadinetti* network in the *Library-Post* location and using *Kafka* network in the *Stockmann* location.

The sniffer was equipped with a Wireless Networker 802.11 card and the detected client device was capable fo establishing a WLAN session in the observed AP, which offered a connection to the Internet. We identify users according to their wireless card MAC address, and assume that there is a fixed one-to-one mapping between a user and a wireless card. The mapping is anonymous. We have no mapping of MAC address to user identifier. Moreover, the clients were associated with over 6 unique network SSID in our traces, but there were 14 unique APs (BSSID addresses). Nevertheless, these MAC address can be used to have an idea which devices were using the WLAN network. The clients were connected to these APs in different intervals of time and they have been classified by using the BSSID address association value.

### 4.3 Methodology Used for Analysis

As explained in the previous section, the captures were not continuous in time. In this case, the *mergecap* [51] tool was applied to combine the capture files into a single output file. The packets were merged in chronological order based on each timestamp to simulate a continuous trace. Consequently, the analysis of the information is more rapid. This program is applied to all the captures resulting in a 10 hour

trace from each location. Due to the user anonymity, the packets were truncated in the payload content using *editcap* tool, leaving out the information which can risk client privacy. The next step followed is trace filtering, which aggregates IP packets belonging to the specific APs in an area. In addition, we apply specific header information filters. Basically, we take into account only selected header parameters needed for the analysis that is presented in the next section, and we never observe information contained in the payload. We use the *Tshark* [52] application to create these filters. The program dumps and analyzes network traffic resulting in a .txt file, which contains the most relevant fields extracted from the captured frames. The file structure incorporates the following fields: frame timestamp, BSS Id, source and destination MAC address, source and destination IP address, and protocol version, among others.

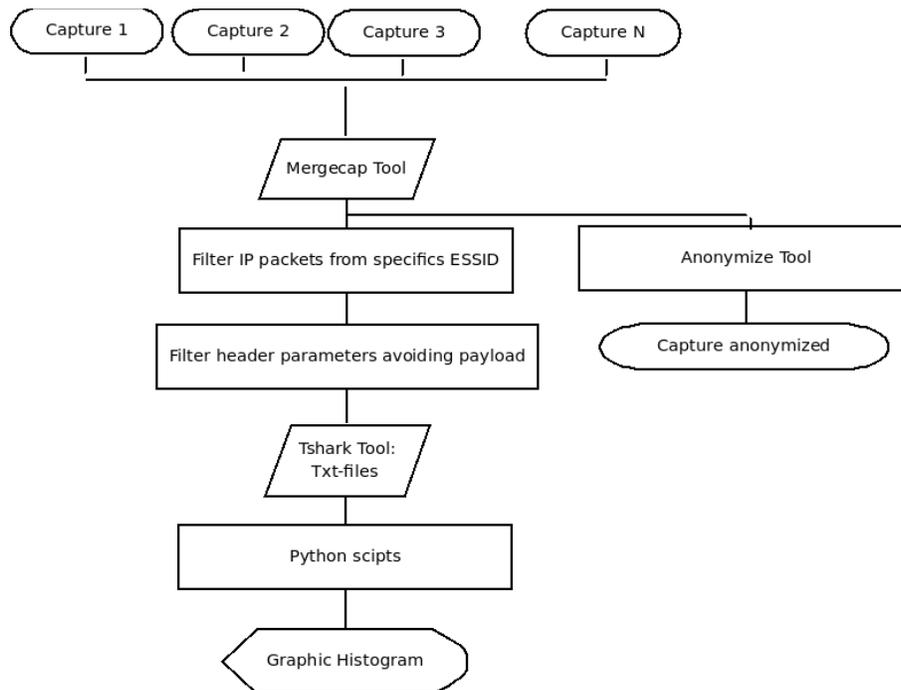


Figure 4.2: Flow chart used to obtain the results.

In order to analyze the previously obtained txt files we have used *python programming language* [53] to build several scripts that can automatically operate on the values to obtain the aggregated results. In addition, some standard libraries, packages developed as an Open-Source modules have been used to facilitate the script efficiency and to program easier. As a standard library, *regular expression operations* have been applied to control the data analysis. Apart from that, some other pack-

ages have been used to provide a compelling environment for numerical analysis and computation. *NumPy* and *Scipy* packages [54] are needed for scientific computing. Furthermore, we generate plots, histograms and bar charts using the *matplotlib* library [55] which is a 2D plotting library to produce high quality figures.

Concluding with trace anonymity, it is recommended to apply an anonymization tool. Anonymization is a technique to overcome privacy issues in sharing trace data, and to enhance the cooperation between different organizations by enabling the exchange of monitored data. Today, several organizations provide anonymized traces using different mechanisms for fast on-line anonymization. We use the *AnonTool* [40] which provides a flexible anonymization framework and extended functionality to allow fine-tuning of privacy protection level. Finally, the traces have a higher level of user anonymity and they can be used for other experiments too. Figure 4.2 shows a summary about the process followed in the data analysis.

## Summary

WLANs grow in size, scale, and complexity, consequently, the challenges for WLAN traffic measurement also grow. The primary challenges for WLAN measurement include the geography diversity of WLAN deployments and the physical proximity for WLAN packet capture. We used open-source monitoring tools to collect wireless traffic from 3 selected locations and 14 APs in Helsinki center. The clients were connected to these APs in different intervals of time and they have been classified by using the BSSID address association value. The methodology to process the traces combines different tasks and tools in order to maintain user privacy and complete the WLAN analysis. The post-processing is realized developing Python scripts to obtain the aggregated results.

## 5 Measurements

This chapter discusses wireless Internet usage in public WLAN access points in the hot locations. For researchers to obtain an understanding of application sessions in the network traffic, we search for characteristics of typical mobile usage. This section evaluates some statistics to gain an understanding of the user behavior. Moreover, it proposes a comparison between the results obtained in the strategic places, in order to understand the client diversity in public WLAN hotspots.

### 5.1 Definitions

One goal of this study is to understand the user behavior. We imagine user "sessions" in which a user (MAC address) joins the network, uses the network and leaves the network. For this, the following definitions are needed:

**Card:** a wireless network interface card, identified by a MAC address.

**Endpoint:** is the logical endpoint of separate protocol traffic of a specific protocol layer. TCP endpoint is a combination of the IP address and the TCP port used, so different TCP ports on the same IP address are different TCP endpoints.

**Active card:** a card involved in a session, during the day, or at the hotspot locations.

**SSID:** is a name that identifies a particular 802.11 wireless LAN.

**Session:** A session starts when a card associates with an access point.

**Inbound:** traffic from-AP to the client.

**Outbound:** traffic sent by the client to-AP.

**Remote host:** an end point, which can offer web content, data streaming, phone call service or a file download.

**Domain Name System (DNS):** is a hierarchical naming system for devices connected to the Internet.

## 5.2 User View

This section discuss about the user view and diversity at the various locations. We were interested in knowing which locations had the most users, and how usage varied at each location.

### 5.2.1 User Duration

User durations help us to understand connection patterns, the amount of time users are active, and the traffic generated. These results can be used to develop synthetic workload generators for performance evaluations of the WLANs. We identified 153 unique users, 355 unique users and 142 unique users in the *Kamppi*, *Library-Post* and *Stockmann* locations respectively, among which 66, 151 and 84 users had a user durations time less than 2 min. These users could not be associated with any single AP. We conjecture that such users were on the periphery of the range of the sniffer, so few packets sent or received by the users were captured. In *Kamppi* location about 90% of the user durations ended within 2.9 h. The median user duration was approximately 32 min, and the mean user duration was about 1.2 h. The median user duration in *Library-Post* location was 32 min, while the mean for user durations was 51 min. About 90% of all user durations ended within 4.13 h. The median user duration in *Stockmann* location was 30 min, while the mean for user durations was 45 min. About 90% of all user durations ended within 3.5 h.

### 5.2.2 User Devices

Most laptops and handled devices these days have built-in wireless Network Interface Card (NIC). We used MAC addresses to classify the wireless NIC by vendor. A MAC address is a 6 bytes long with first 3 bytes representing the vendor ID. We used the vendor ID list from IEEE to classify the entire number of clients in this study. We found that the devices percentages were:

In *Kamppi* location: Apple (53%), Intel (20%), Nokia (5%), Hon Hai Precision (6%), Askey (2%), AzureWave (2%), GemTek (2%) and others (10%).

In *Library-Post* location: Apple (46%), Intel (27%), Hon Hai Precision (8%), GemTek (5%), Nokia (3%), AzureWave (2) and others (8%).

In *Stockmann* location: Apple (63%), Intel (13%), Nokia (10%), AzureWave (3%), GemTek (3%), Belkin (1%), Askey (1%) and others (5%).

### 5.2.3 Traffic Patterns per User

This section provides a study about the inbound and outbound traffic per user associated to each network. Usually, a client must specify on SSID explicitly to "ANY" to associate with any available network, or it cannot associate with the AP. Users with the same BSSID address are counted as connected in the sub-network. This study is based on clients IP packet analysis which belong to these sub-networks. The collected IP traffic comes from different links, from client to AP (outbound) and APs to client (inbound). Each graph shows the IP traffic as measured for each customer in each location.

The script written for the following analysis realizes several filters based on the needs of the study. First of all, we filter each packet that is not associated with the SSID of the network in the specific location. Moreover, in order to organize the traffic per user we focus on MAC-IP address pairs to know which of the connected clients are new. The traffic for each client is distributed as inbound and outbound traffic based on whether the MAC address is the sender or not (source MAC address). We quantify the IP traffic in each direction using the IP length header field which is the size of the IP packet and it is used to accumulate the traffic of each user. The final result is a python dictionary containing the traffic information classified by the network, link direction, and user. The script is applied for each location resulting in a bar histogram which represent the IP traffic pattern by the users.

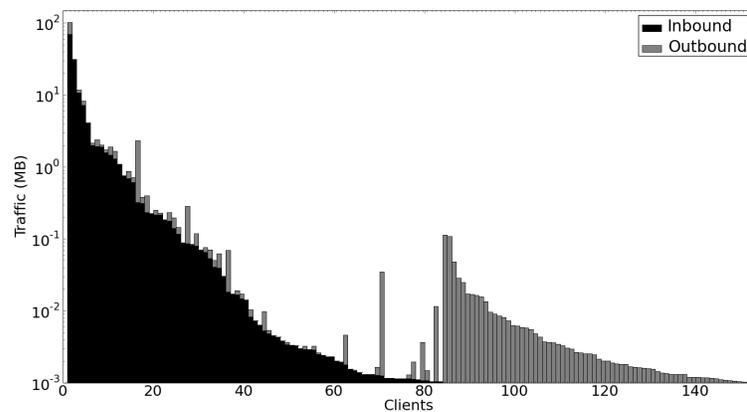


Figure 5.1: User traffic patterns in *Kamppi* location with four APs providing service.

Figure 5.1 plots the total number of unique MAC connected in the networks evaluated near *Kamppi*. The amount of inbound traffic seen was 138.5 MB and for outbound traffic 40.1 MB respectively. From the figure, we see that the total num-

ber of user is 153 but they were not equally found in each AP. In the 10 hours trace we found 107 clients in *Lasipalatsi* network (80 clients in one AP and 27 in the other) and 65 in *Helsingin kaupungin WLAN* (58 in one AP and 7 in the other). As we can note, the total number of customers has been reduced, this is the reason for finding the same MAC addresses in different networks, and these were combined as a unique traffic item. Based on the quantity of traffic per client we can suppose two user classifications, mobile and stationary. We could estimate 20 associated users as stationary clients. However, the majority of customers are probably mobile users due to the low amount of traffic found. For example, travelers just checking some web content in their WLAN session. The most intriguing part of the graph is the elevated number of users with just outbound traffic and the small amount of transmitted traffic. Considering some NAT drawbacks, some services may not work properly with some implementations of NAT. The traffic from the Internet cannot reach the user directly. However, this does not necessarily prevent two-way services like Internet Relay Chat (IRC) and email. Moreover, as we mentioned in the previous section, that such users could be in the sniffer periphery but not associated with the WLAN.

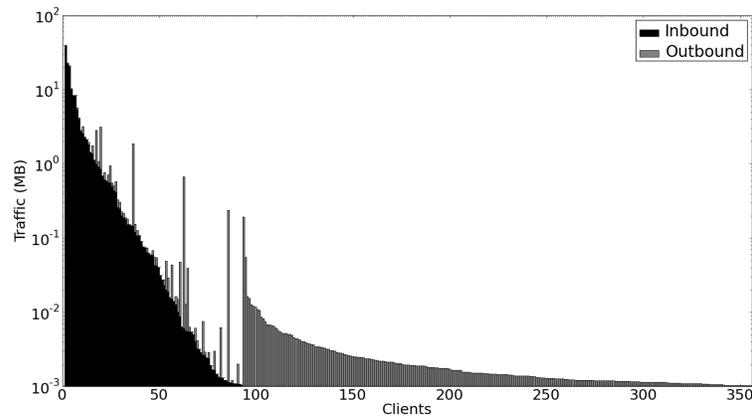


Figure 5.2: User traffic patterns in *Library-Post* location with four APs providing service.

Figure 5.2 shows a total number of 355 users. The total inbound IP traffic is 140.8 MB and the outbound is 13.6 MB. The APs see a different number of users over the 10 hours period. The number of customers found from *Stadinetti* APs is 188 (24 clients in one AP and 164 in the other), 45 in *Helsingin kaupungin WLAN* (6 and 39 clients) and 213 in *HSverkko WLAN* (with 96 in one AP and 120 in the other). This user distribution reflects the topology of the networks and the setting, where many users were gathered in a small area during the scheduled times. Each

network can be related to a specific user population due to the APs location in the capture area and based on our observations. The *Stadineti* and *HSverkko* networks offer service a high number of users than *Helsingin kaupungin WLAN* which can be possible caused the good signal quality for these WLANs. We can observe many users with just outbound traffic, as was seen before, the observation could be caused by NAT drawbacks. Examining the traffic by the user, there are some customers with an elevated inbound traffic. This is interesting considering that hotspot clients are not limited by the capacity of the connections.

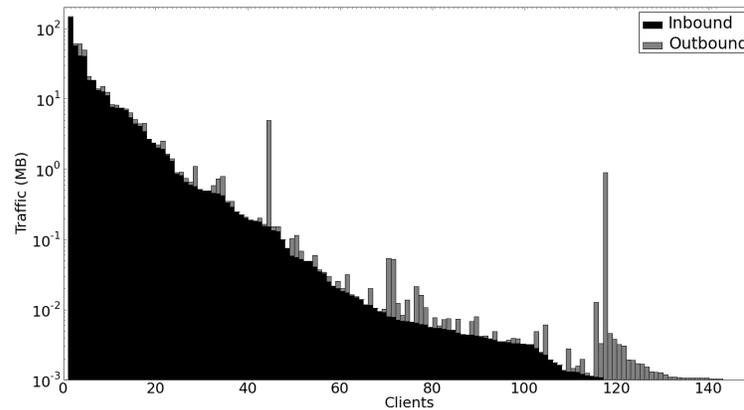


Figure 5.3: User traffic patterns in Stockmann location with four APs providing service.

Figure 5.3 plots an inbound and outbound traffic distribution by users in the *Stockmann* location. Similar to previous locations, the number of users vary depending on the AP. The total number of customers has been 143, distributed as 53 clients in *Kafka* WLAN and 90 in *Helsinki kaupungin WLAN*. The total inbound traffic is 435.2 MB and the outbound is 50 MB. Figure 5.3 shows a low number of clients with just outbound traffic compared with the *Kamppi* and *Library-Post* locations. This observation can be caused by a proper NAT service implementation, but apart from that due to a low usage of mail clients. The total inbound traffic shown in the graph can reveal that static users, such as home clients or workers, were connected to these APs and cause elevated traffic. Home users may be associated in these networks for long period of time.

### 5.3 Traffic Statistics of Public WLAN Traffic

This section presents a network view of the traces analysis. We were interested in knowing how much user traffic the WLANs handles, how this traffic varies with time, how much traffic is accounted by each location, and how much traffic is accounted in each direction. We also wanted to understand the load on the WLANs.

#### 5.3.1 IP Traffic Load

In this section we study the variations in the pattern of IP traffic over a 10 hours period. We perform the same analysis for each location in the measurement study. Moreover, we correlate data rate used with the traffic load on the networks. Figures 5.4, 5.5 and 5.6 show the IP traffic captured every 10 minutes. Each 10 minute is a point in the plot.

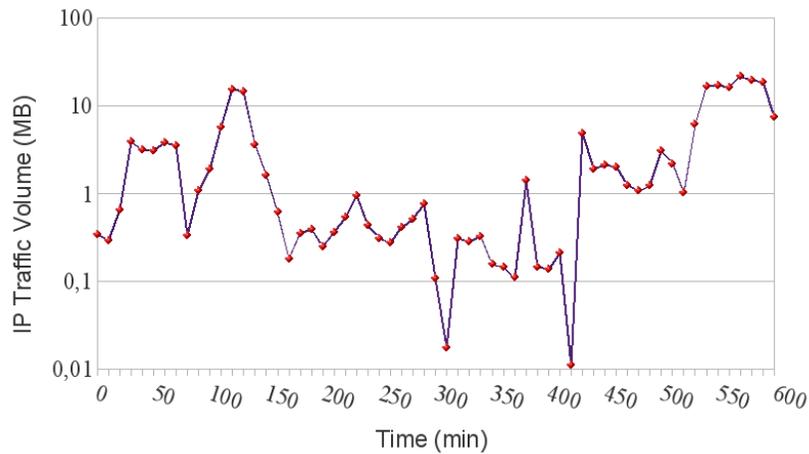


Figure 5.4: Traffic IP pattern in *Kamppi* location with four APs providing service.

The network load was not constant during this interval of time. In addition, we can see a high load variation between locations. The dips represent period of time when there were few users connected to the APs. The IP traffic was fairly low, with the peak 10 min throughput observed of 22 MB in *Kamppi*, 16 MB in *Library-Post* and 90 MB in *Stockmann* for a 10 minutes interval. These corresponds to 300 Kbps, 218 Kbps and 1,2 Mbps respectively. Note that the throughput calculations are based on IP traffic only. The actual WLANs load would be higher if management and control frames were considered. The small amount of traffic transferred for some hotspots during these hours can indicate that there is lot of bandwidth not consumed by any client.

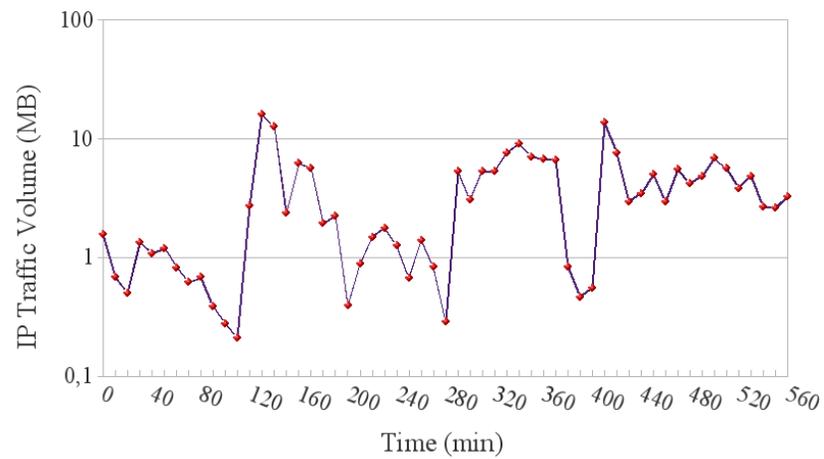


Figure 5.5: Traffic IP pattern in *Library-Post* location with six APs providing service.

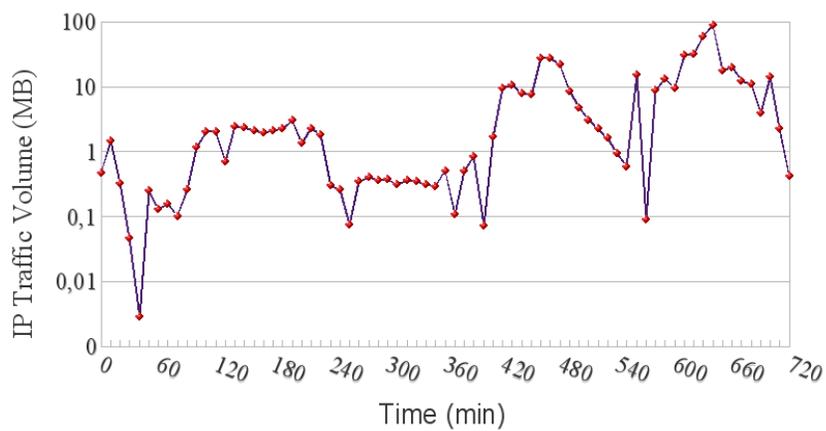


Figure 5.6: Traffic IP pattern in *Stockmann* location with four APs providing service.

### 5.3.2 Analysis of User TCP traffic Load

In this section we study the user TCP traffic load in each WLANs to understand the AP load and the user diversity. Most access points see only little traffic, but several see significant amounts of data, this is illustrated in Figures 5.7, 5.8 and 5.9. The figures show the amount of TCP traffic that active clients generate in each AP. Each figure plots a cumulative distribution function (CDF) and shows the probability that the users (in a specific hotspot) generate less than or equal to the amount of traffic shown by the CDF. Each CDF shows the total inbound and outbound data transferred by the clients during a 10 hour period. Over the life of the trace, the APs show different amounts of handled TCP traffic per hour.

For each place analyzed in this study, figures show several metrics to estimate the user behavior. Furthermore, we have analyzed the traffic of the most relevant SSID, filtering out the part of the traffic from other SSIDs which can confuse the results. The script algorithm analyze each TCP packet associated with the SSIDs (MAC address) of the networks in each location. We organize the traffic per client (unique MAC address) based on MAC-IP address pairs, because the IPs can be provided to other MAC addresses too. The user TCP flows are separated based on inbound and outbound traffic and each user is grouped in the appropriate network. Finally, the CDF is calculated based on the clients TCP flow in each WLAN network for each location.

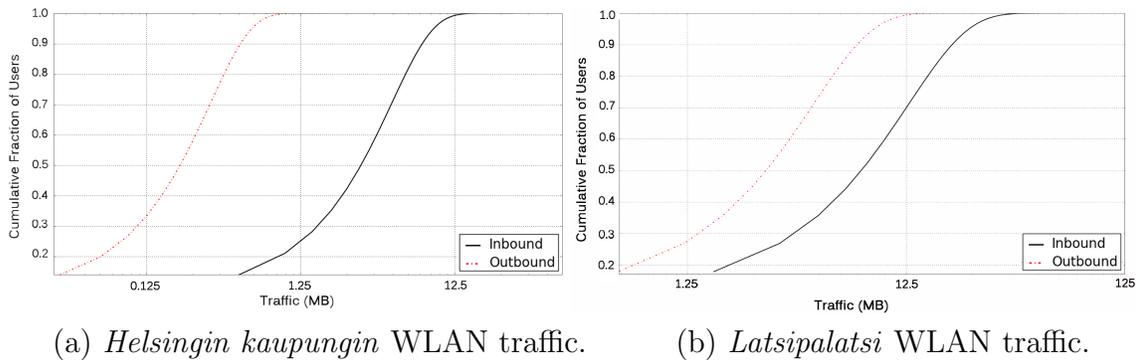
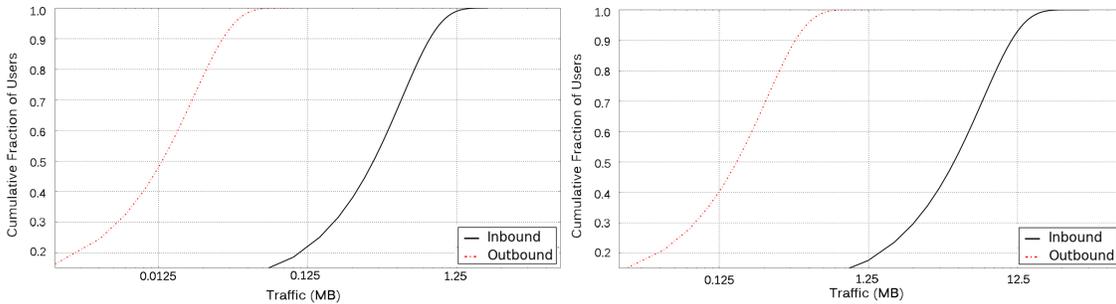


Figure 5.7: A TCP flow measurement from *Kamppi*. Distribution across 10 hours analysis with 51 connected clients (MAC) address in *Helsinki kaupungin WLAN* and 29 addresses in *Lasipalatsi WLAN*.

Figure 5.7 shows the TCP data captured in the most accessed APs in the *Kamppi* location. We collected TCP traffic from 80 users. They transferred a mean of 3.33 MB (*Latsipalatsi WLAN*) and 1.28 MB (*Helsingin kaupungin WLAN*) of TCP traffic in the *Kamppi* area. We can see that for the 95 percent of the users connected to

*Helsingin kaupungin* WLAN Hotspot, the amount of traffic during a trace capture over 10 hours period is less than or equal to 31.72 MB (29.84 MB inbound and 1.88 MB outbound). However, in *Lasipalatsi* WLAN traffic is around 90.65 MB (66.12 MB inbound and 24.5 MB outbound). Furthermore, the number of clients connected in Figure 5.7(a) (51 clients) is larger than in Figure 5.7(b)(29 clients). The differences found in these two WLANs can be explained by the client population diversity. Many of the clients connected to *Lasipalatsi* APs can be statics with a necessity of transferring a large amount of data while the users associated with the *Helsingin kaupungin* hotspots could be mobiles and their sessions do not need to transfer much data.

Both figures show a negligible outbound traffic compared to the inbound, making the amount of the outbound data less significant than the traffic demanded and issued directly by users. In addition, the figure 5.7(b) shows that a ratio of download remains constant across the period in our trace.



(a) *Helsingin kaupungin* WLAN traffic.

(b) *Stadinetti* WLAN traffic.

Figure 5.8: A TCP flow measurement from *Library-Post*. Distribution across 10 hours analysis with 25 MAC addresses in *Helsinki Kaupungin* WLAN and 55 MAC address in the *Stadinetti* WLAN.

Figure 5.8 plots the TCP traffic captured from the most accessed APs in the *Library-Post* location. The mean traffic handled by the clients in the *Helsingin kaupungin* WLAN was 0.2 MB of in/outbound traffic , 3.17 MB in *Stadinetti* WLAN and 0.75 MB in the *Hsverkko* WLAN. As can be seen from Figure 5.8(a) 95% of users had an inbound/outbound traffic less or equal to 2.11 MB in *Helsinki Kaupungin* network. However, the same percentage of the users transferred less or equal to 39.12 MB in the *Stadinetti* WLAN and 8.58 MB in the *Hsverkko* WLAN (with 28 users). These traffic differences can be explained by evaluating the usability of the APs. Figure 5.8(b) plotted in hotspots can offer connectivity through a static (worker/student) population due to the high traffic compared to Figure 5.8(a) which can support traveling clients as mentioned before.

These plots show the CDF of the traffic volume (inbound and outbound). One intuitive interpretation is that very likely most outbound traffic consisted of request packets (e.g, HTTP GET), which are generally smaller than inbound response packets which contain application payload. Therefore, most clients in these APs tend to send smaller requests and wait for larger responses. Such characteristics are typical in web browsing, news group reading, and accessing email services.

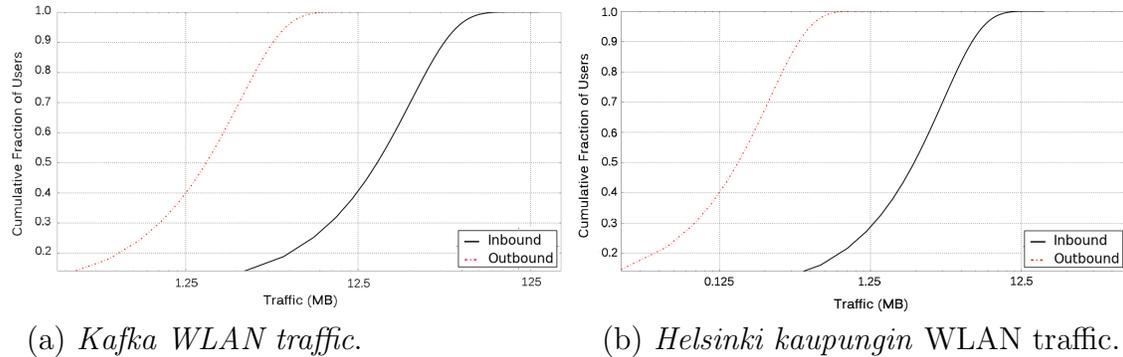


Figure 5.9: A TCP flow measurement from *Stockmann* location. Distribution across 10 hours of capture with 48 MAC addresses in the *Kafka* WLAN and 50 addresses in the *Helsingin kaupungin* WLAN.

Figure 5.9 shows the data collected in the most significant hotspots in the *Stockmann* location. The captured traffic reveals a total number of 98 clients served by the APs. The mean inbound/outbound traffic consumed by users during the collection period in these APs was 8 MB in the *Kafka* WLAN and 1.13 MB in the *Helsingin kaupungin* WLAN. These figures illustrate that 95% of users has a inbound/outbound traffic less or equal to 142.46 MB in the *Kafka* WLAN and 6.85 MB in the *Helsinki kaupungin* WLAN. This can indicate that many of the clients associated with these hotspots were likely traveling consumers in the *Helsinki Kaupungin* WLAN due to the small amount of data that was captured compared to *Kafka* WLAN. Surprisingly, some users in *Kafka* WLAN transferred a significant amount of traffic, which can reveal the presence of static customers.

Moreover, we can observe in the figures that a fraction of users between 15% and 20% by amount of submitted data, do not use the service offered by these free APs. A good reason may be the users did not use the AP at all but their phones or laptops just automatically performed address resolution in the AP. The address resolution is used for finding a host's link layer (hardware) address when only its Internet Protocol (IP) or some other network layer address is known. Because our environment contains a mixture of residential and working users, these plots show

a mixture of the usage patterns during workday as well as residential usage.

Two observations can be made from this section. First, load is unevenly distributed across APs in the WLAN. Secondly, the traffic load on APs is loosely related to the number of users. Traffic load depends on the type of users, and the applications they use. Non-uniform AP usage seems to be an inherent characteristic of WLANs.

### 5.3.3 Round Trip Time Measurements

This section analyzes the TCP traffic captures at each location during the trace collection and focuses on round trip time estimations. The RTT represents the time to send and receive a message, from the client to the server and back to the client. We realize round trip estimations based on a packet-pairs (the time to acknowledge the segment that was sent). This measures the delay experienced by traffic that was sent to the network. RTT is an important metric in determining the behavior of a TCP connection, and it is useful in measuring the available transfer characteristics of a path. This information helps to determine factors that limit data access. Knowing the RTT characteristics for a given set of TCP connections helps to understand the mobile user behavior. Wi-Fi hotspots providing local coverage to traveling users with the ability to access web content and RTTs estimations can give information on how these hotspot networks provide services to their customers. This section is based on the TCP connection analysis in order to understand the experienced delays when clients use wireless services and HTTP responses delay analysis to observe the responsiveness of the servers and how long the clients were retrieving content from different domains.

#### TCP Open connections analysis

This study measures RTTs at WLAN network locations where several users contribute to the network traffic. The tool used to estimate the RTT values for this task is called *Ethereal* (it has been demonstrated that Wireshark has an error which can lead to interpret the RTT value for an ACK packet in an incorrect way), which is the previous version of the Wireshark sniffer. The method uses TCP timestamps to associate data segments with the acknowledgements that trigger them. The RTT average estimation is calculated considering the up-link and downlink, the bytes sent to the remote host and received from it in all the connection achieved during the session.

Clients can have several RTT values during a WLAN sessions. On one hand, this study estimates the path propagation delay in TCP connection by halving the minimum RTT estimate. We use the minimum RTT because low bandwidth links can have very large variability in queuing delay. The minimum RTT is typically estimated by measuring the three-way handshake. Thus, we estimate the RTT average using values between the client SYN and the first server SYN-ACK since packet queuing delays are usually small at this point. On the other hand, we are interested in ACK RTT values during the observed TCP sessions in order to evaluate application responses delays even when if we find high asymmetries.

We save the RTT estimations of the SYN/ACK segments, which were stored using an statistic *tethereal* tool, in a txt file in order to process them by a script file. The script algorithm basically calculates the RTT mean for all three-way handshakes RTT values and plots the results in a CDF histogram.

Figures 5.10, 5.11 and 5.12 show RTT variations for each location with values that are taken at different times of a day. Each figure consist of the following by (a) represents the values estimated by observing the opened TCP connections (SYN/SYNACK pairs), and (b) plots the RTTs estimations for ACK packets from the user applications. We consider that mobile terminals can cause large delay because they often have limited processing power or due to reason of user mobility (if the client is moving to another location and the packet is buffered in the network). However, freezing TCPs connections may lead to degraded performance [56]. Since different connections may have different RTT values, this adds to the difficulty in accurate predictions.

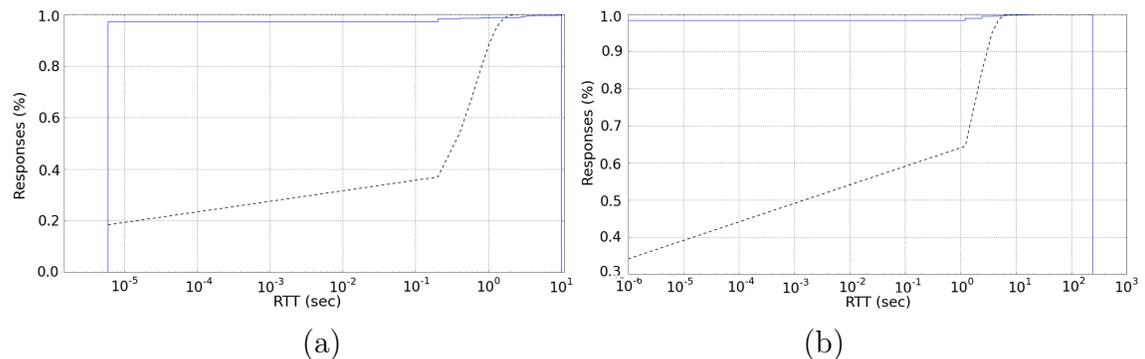


Figure 5.10: Analysis of the RTTs in the *Kamppi* location.

Figure 5.10 show RTT estimations in the *Kamppi* location. In Figure 5.10(a) the RTT mean is 118ms and the most usual interval varies from 1ms to 150ms. At least 40% of TCP connections appear to have low delay around 100ms. Besides,

the 80% of the TCP connection have a value inferior to 700ms. We can see in the figure some connections (10%) which had an elevated delay around seconds with a maximum value of 10.11 sec. These can be caused by several users trying to connect with servers which are not available, file transfers, or some system problems. Figure 5.10(b) plots a response delay CDF constituted by several user applications. The mean is 122ms with a maximum value of 242sec. 60% of responses have a delay less than 100 ms which can lead to low transfer capacity. Furthermore, 30 % of responses have an elevated delay which in some cases cannot perform well for applications.

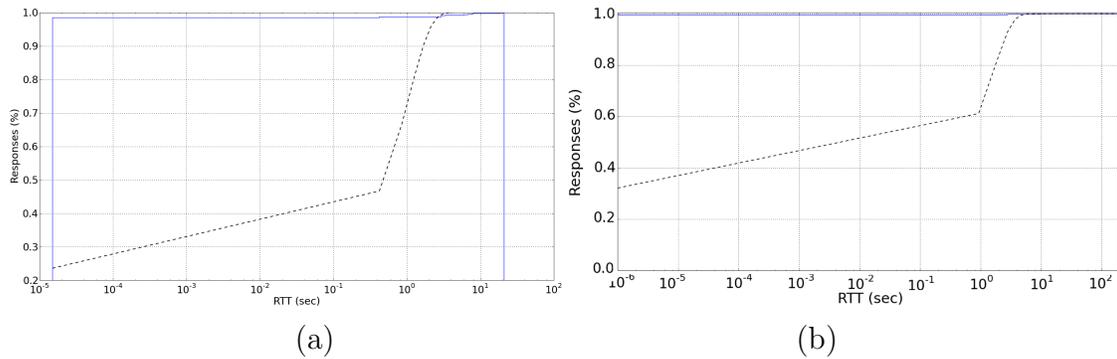


Figure 5.11: Analysis of the RTTs in *Library-Post* location.

Figure 5.11 shows the RTT estimations in *Library-Post* location. 5.11(a) plots the RTT CDF of wireless TCP connections in *Library-Post* with a mean of 142ms and a maximum value of 21sec. Close to 50% of TCP connections have a delay less than 100ms which indicates that server response delays were experienced in TCP connection establishments. However, there is an interval with 30% of connections around a second that is a relevant percentage to consider. These responses can indicate network problems. Figure 5.11(b) CDF shows response time from user applications where the mean is 68ms and a maximum value was 184 sec. 60% of responses have a delay lower than 500ms, which indicates an acceptable value with which users can use a two-way conversation without interrupting each other. We can observe 20% of responses with high RTT values which can be caused by several reason such as file transfer or congestions. The interpretation could be further improved if responses were analyzed in more detail focusing on user applications.

Figure 5.12 shows the RTT estimations in the *Stockmann* location. Figure 5.12(a) shows a mean RTT value of 148ms and a maximum value of 47sec. The performance is similar as commented previously. However, in Figure 5.12(b) the percentage of responses (30%) with a low delay is less than in both *Kamppi* and *Library-Post* locations. The maximum value observed is 899sec, which could be caused by changes

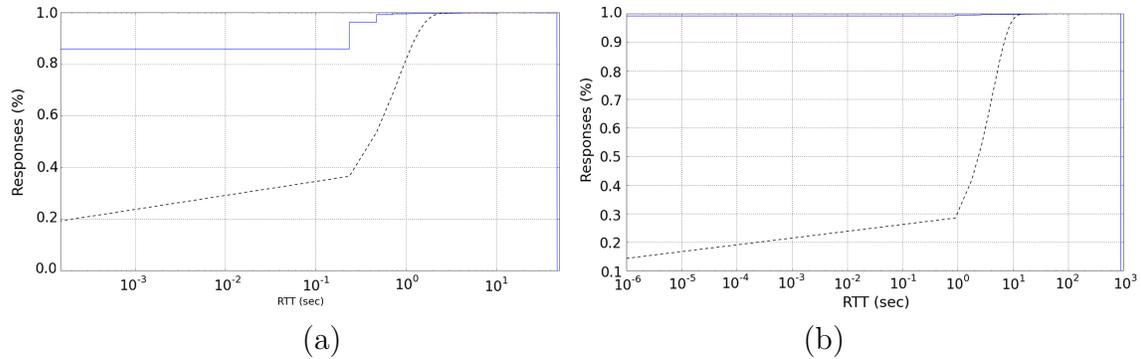


Figure 5.12: Analysis of the RTTs in the *Stockmann* location.

in network conditions where buffers cannot adapt rapidly.

The mean RTTs in each location are less than 150ms, and the average latency itself is not a significant issue. However, as we can see from the previous figures, WLANs experience sometimes significantly higher delay, with network jitter and packets loss. Consequently, when several users are connected to the same AP, congestion easily occurs and jitter can be very significant if large packets are presents. The International Telecommunication Union’s Standardization Sector, or ITU T [57], recommends in standard G.114 that the one way delay should be kept lower than 150 ms for acceptable conversation quality or audio/video quality. VoIP calls can be placed in environments with RTT as high as 500ms and above but the delay will be noticeable to callers.

### HTTP level analysis

For RTT measurements HTTP layer analysis allows observing the server responsiveness by measuring the response time between the client HTTP request and the server response to return a web page. This is the time spent by a user while waiting for a web pages that was requested. Measuring the delay experienced by customers is important to web sites. The measurement allows the site to evaluate the latency experienced by the clients (regardless of their network location). This analysis is conducted on the HTTP level in order to understand application performance and to present RTT statistics for the web traffic found in the trace. The tool used to estimate the RTT values is called *tcptrace*[58], which generates an RTT graph from HTTP traffic samples (taking into account RTT from ambiguous ACKs) representing the minimum, average and maximum RTT values observed in the measurement interval. We evaluate the HTTP response delays using a continuous trace of 3 hours

in each location, thus the latency estimate is from these locations only. The disadvantage of this method is that the measurements may not reflect the real web user experienced in other places. Figures 5.14 (a), 5.14 (b) and Figure 5.15 show the HTTP request/response latency observed in each location. The figures show three lines which track the minimum, the average and the maximum observed RTT.

In addition, we investigate on the web domain connection times which can explain a user's behavior when they are connected to a web sites. The time a client was connected with a domain can constitute of the time to retrieve a web page and the time possibly used to visit different web pages from the same domain. The time taken for a web page to be completely displayed on a web browser depends on several factors such as bandwidth, web page size, RTT. The number of objects downloaded. We make an estimation of the time a client is moving for different pages.

We estimate the web domain connection time by measuring the total duration that the user is connected with a domain. This time can be made up of by several application sessions which were requested by the client when he was visiting other pages from the same domain. Thus, the time is the user connection period in this specific domain. The algorithm that is applied to the traces, analyzes HTTP request/responses for every user HTTP session and avoids retransmissions and out of order request/responses. We consider the first GET request to retrieve the web contents and the last 200 OK responses from the domain which were sent by the server. The interval can be continued or in some case the user could not transmit any request after a specific period. However, this time interval time without transmission is omitted in the time estimation. Figure 5.13 shows an example of the web domain connection time for a user and a server.

The processing times (server and client) are not separately considered in the web connection times measurements, which can vary depending on the web page, but they contribute to the aggregate result. On the server side, pages with static content require minimal processing compared to interactive web pages. On the client side, the browser has to load and run the correct interpreter for the web page which can incur some delay. These factors are difficult to estimate because they depend on the browser and server characteristics. We consider that they are not relevant for this study, consequently, and do not look closer into them.

Figure 5.14 (a) shows the HTTP response delay statistics for a 2.53 hour trace in the *Kamppi* location. The trace has 2909 persistent HTTP connections with 4581 B/s. We can observe that the RTT values vary between 50ms and 240ms over the

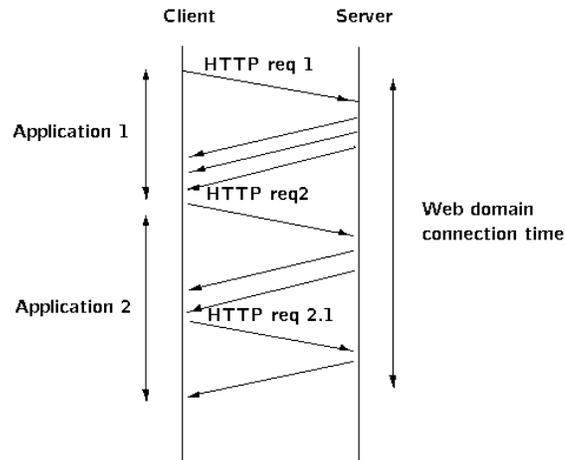


Figure 5.13: Web domain connection time example.

period with an average RTT of 110 ms.

Figure 5.14(b) shows the HTTP response statistic from the *Library-Post* location during a 2.45h trace. The trace contains 1266 persistent HTTP connections with 2070 B/s. The RTT values vary between 40ms and 230ms over the trace with an average RTT of 84ms. These locations show large variation in the RTT, which is often typical for web traffic characteristics. We can interpret from the figures that the users were interested in pages containing large responses such as the images of web pages.

The observed delay variations might be likely to be caused by network and server load characteristics. Furthermore, some responses can be faster due to caching services. HTTP connection caching can underestimate transmission time by assuming that there is no penalty for slow-starting for later requests and managing persistent connections.

Figure 5.15 plots an average RTT of 80.74ms during a 2.50h trace period. The number of persistent HTTP connections was 1200 with a bandwidth of 472 B/s. The RTT duration varies from 40ms to 280ms. However, these durations are larger during the first trace hour which indicates that network could be more congested than the next period of the trace. This can indicate that users were connected with busy web pages or there were some server response difficulties. Comparing the HTTP RTT figures, we observe that the server response delays were not very

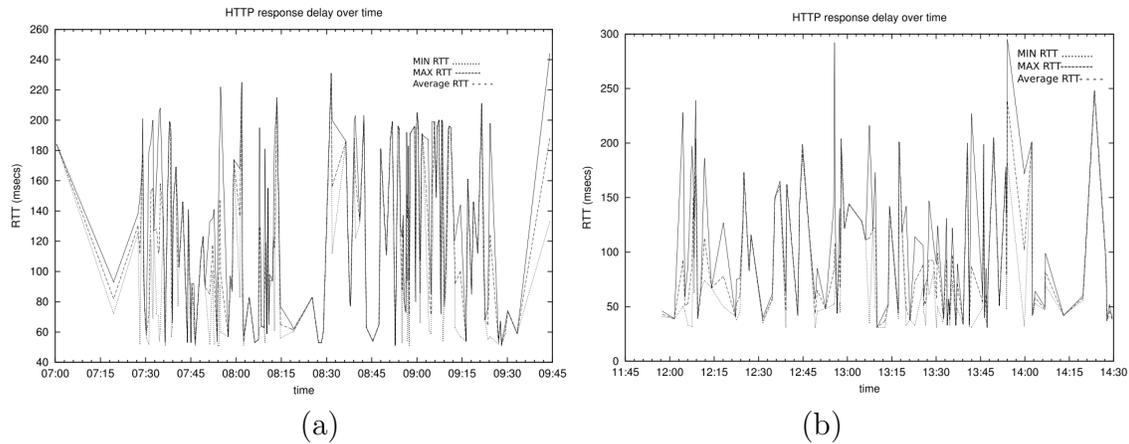


Figure 5.14: Analysis of HTTP response delays in the *Kamppi*(a) and *Library-Post* (b) locations during a continuous trace of 3h.

similar, which can be interpreted that user behavior can depend on the location where they connect to web services.

Figure 5.16 (a) and (b) show a CDF user distribution of the web domain connection times for each of the HTTP sessions in the *Kamppi* and the *Library-Post* locations. We can observe in the figures that 60% and 40% of web connections (*Kamppi* and *Library-Post* respectively) had a duration of at least of 15 minutes. This can show that users spent a short interval visiting a domain during their connections, which can indicate traveling or worker users who just quickly check some information. In addition, we can observe a fraction of web connections with a duration time between 25 minutes and 2.15 hours which can reveal the presence of more static users population connected with to the web.

Figure 5.17 shows a CDF figure of web domain connection times in the *Stockmann* location. We can observe that the figure has similar results compared to *Kamppi* location. 60% of web connections has a duration of less than or equal to 15 minutes which manifest in short user sessions. However, as in previous locations some users can be considered more static (workers or home users) because their web domain connection times has a high duration in the order of hours.

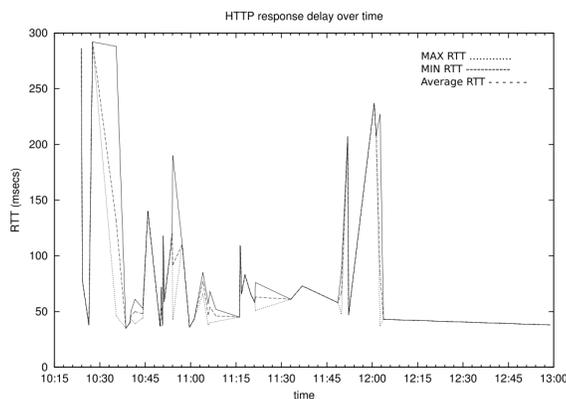


Figure 5.15: Analysis of HTTP response delays in the *Stockmann* location during a continuous trace of 3h.

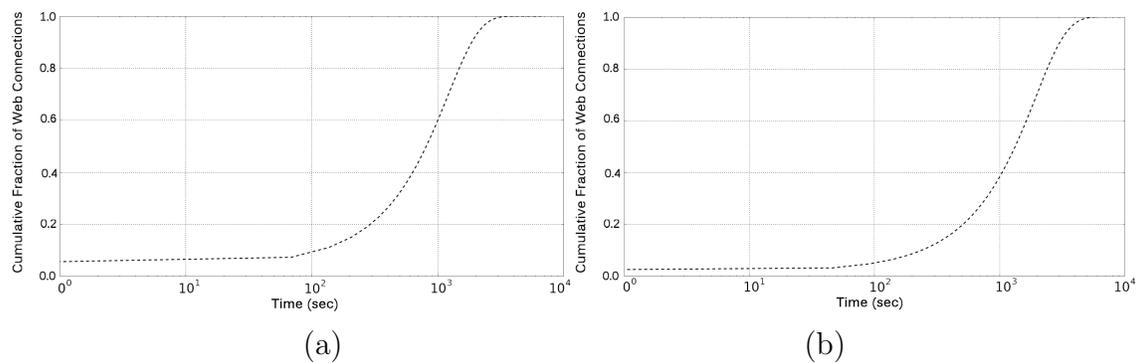


Figure 5.16: Analysis of web domain connection times in the *Kamppi* and *Library-Post* locations during a continuous trace of 3h.

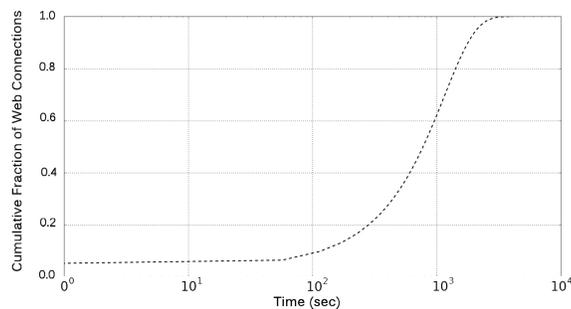


Figure 5.17: Analysis of web domain connection delays in the *Stockmann* location during a continuous trace around 3h.

## 5.4 Internet Usage in Wireless Hotspot

This section reports our analysis about the most used protocols that are responsible for the traffic. Figure 5.18 shows the distribution of network traffic load by users in the measured WLANs. We see that most of the traffic is generated by TCP in every location (*Kamppi* with 174.08 MB, *Library-Post* with 157.44 MB and *Stockmann* with 486.4 MB). Apparently, the small amount of UDP traffic seems to indicate small presence of real-time video/audio applications. The rest of the traffic is sent by ICMP and other protocols. With the predicted fast growth of real-time video/audio applications, in addition to voice over IP (VoIP), there might be requirement from users for the hotspots to be provisioned to satisfy tight delay constraints.

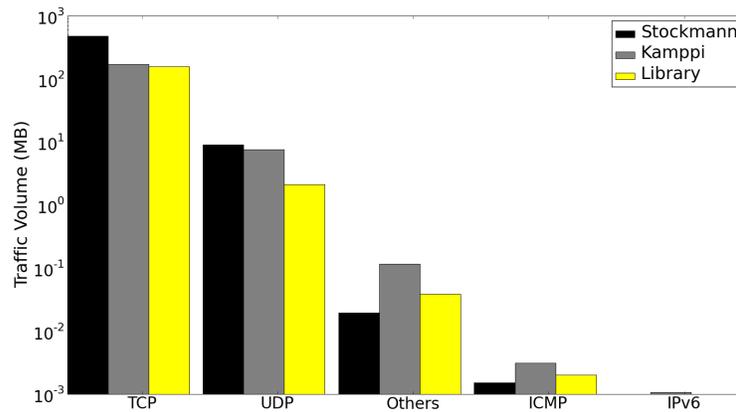


Figure 5.18: Traffic protocols distribution from 10 hours trace in the *Kamppi*, *Library-Post* and *Stockmann* locations.

### 5.4.1 Analysis of User Applications

Next we present a study of the applications that are the most used by the entire user population that is connected during the period of traffic capturing. We base our Analysis of TCP/UDP traffic and plot application histograms per user and show the traffic volume. Knowing the traffic mix can help researchers to model user traffic better, which is relevant when simulations are used to evaluate mobile protocols.

The Internet Assigned Numbers Authority (IANA) is responsible for maintaining the official assignments of port numbers for specific uses. The following application analysis is based on the well known ports managed by IANA. These ports are

assigned numbers from 0 through 1023. The list specifies the ports reserved for assignment by the Internet Corporation for Assigned Names and Numbers (ICANN) that are used by the application endpoints using TCP or UDP. Each different type of application has a designated port number.

The most interesting ports scanned during the data analysis are HTTP and HTTPS (port number 80 and 443, respectively), which indicate the presence of web applications. The web applications can include shopping carts, forms, login pages, dynamic content, discussion board, and blogs. Their well-known ports that are studied include POP3, (Post Office Protocol version 3) application, IMAP (Mail Access Protocol), SMTP (Simple Mail Transfer Protocol), which are commonly used by email applications.

Each TCP/UDP packet is examined in order to find the source and destination ports (TCP/UDP source port, TCP/UDP destination port) of the communication. If either of the ports was not well-known, we associate the packet with an unknown service. In opposite the case, the packet is associated with the corresponding protocol based on the IANA port numbers. However, many of the client ports match a random value that cannot be matched, but in most such cases the server is using a low-numbered port (80 for HTTP). Consequently, this traffic is associated with the application according to server port. A high number of different applications are used by less than 0.5% of the clients, we classify those as *other applications*.

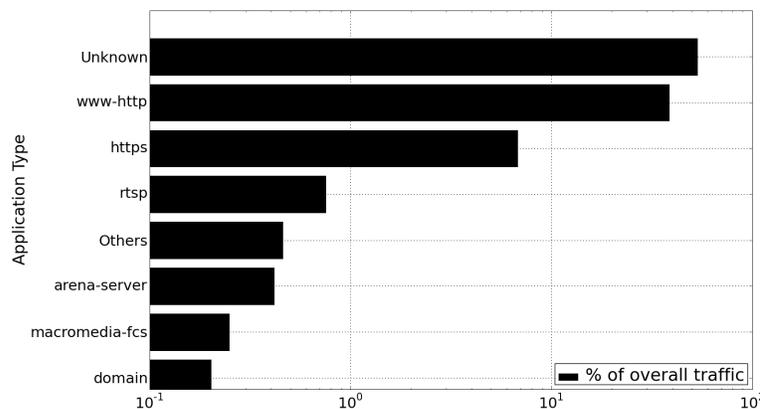


Figure 5.19: Analysis of the use of applications by volume connected to the *Kamppi* AP location.

Figure 5.19 details the protocol traffic generated by several well-known applications in the *Kamppi* location. Clearly, HTTP and HTTPS) dominate this hotspots net-

work usage, with 38.5% and (7%) of the total traffic. Consequently, web browsing is by far the most popular application. About 53% of the traffic is contributed by unknown applications, which use unassigned port numbers but are running on either TCP or UDP. Application-level protocols is analyzed to identify this portion of traffic, resulting that a majority of this application percentage was generated by programs that dynamically establish connections via arbitrary ports, for example, P2P applications. Note that we show only the applications with more than 0.5% traffic contribution. The remaining applications that top these values are *arena-sever* for an interactive game, *Macromedia flash communications* for audio/video applications, and *domain* for domain name server.

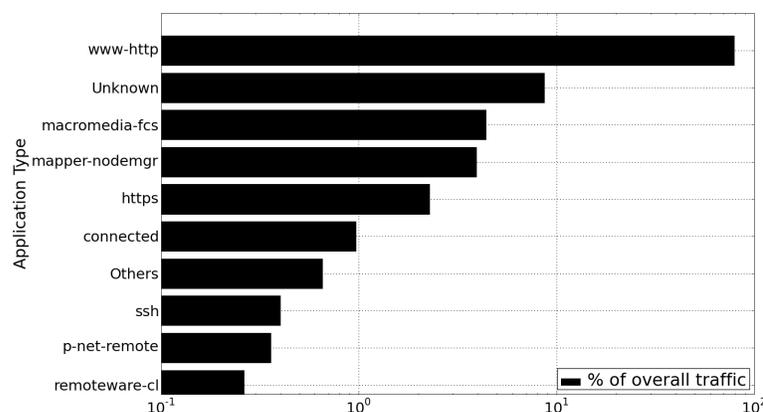


Figure 5.20: Analysis of the use of applications by volume connected to the *Library-Post* AP locations.

Figure 5.20 shows web browsing (HTTP) as the most popular application, contributing 79% of the total bytes transferred, followed by *unknown* applications which use unassigned port number (8.56%) and *Macromedia flash communications* (4.3%). News applications appear in the *Library-Post* location but they contribute only a small percentage of the overall traffic. The *Mapper nodemgr* application has constituted 3.9% for node management tasks, which may be covering capacity planning, firewall maintenance or cross-platform tasks. Close to 1% of the traffic is used by a connected service where users have been connected online. We can see in the figure that clients did not use a lot security protocols such as *ssh* and *https*.

In Figure 5.21, web browsing is the most popular application like in the previous locations with 81% of overall traffic. However, we can observe that new services like *tripe*, *imaps*, *ipp* and *mdns* have been used. *Tripe*, with 1.22 % popularity, is a

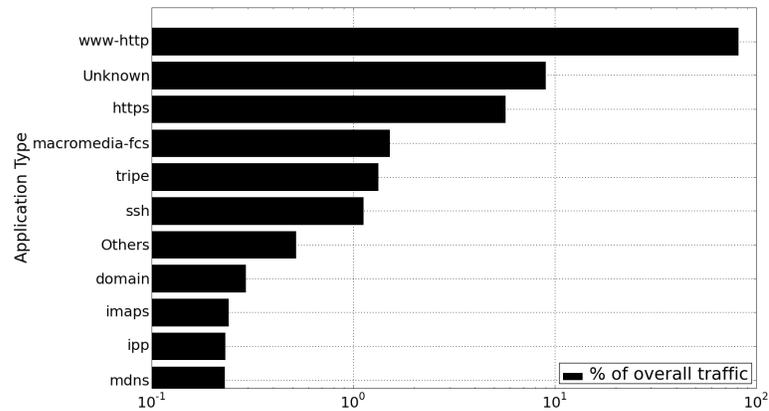


Figure 5.21: Analysis of the use of applications by volume connected to the *Stockmann* AP locations.

trivial IP encryption application, *ipp* is executed as an Internet printing protocol, *imaps* as a messaging protocol and *mdns* for multicast DNS.

It is important to point out that the usage pattern closely depends on the presence of certain user groups. For example, no *imaps* and *tripe* was observed from the *Kamppi* location trace. However, all of the traces confirm the predominant position of the HTTP protocol. It can be stated that applications which generate the highest traffic during the connection, are web applications and *Macromedia* applications. They have been used by many of the users connected to different AP locations in Helsinki. Moreover, an important *unknown* traffic component percentage is found in every place analyzed, this is represented by P2P applications (e.g, Gnutella, BiTorrent, KaZaA) which establish connections via arbitrary ports.

### Application Popularity by User

Tables 5.1, 5.2 and 5.3 list the most common classes of applications by number of users and total number of transferred packets and bytes in each location. Unsurprisingly, basic services applications such as *netbios*, *bootp* and *dns* are used by several clients in each location. However, we can observe differences in the number of clients in these services, reflecting different user patterns in each location. The high number of *netbios* and *mdns* users indicates that almost half of the users run some version of Windows or Mac system in their devices. As we can see from the tables, the *netbios* and *mdns* traffic is inbound and the protocol is used as a datagram service. In *netbios* the User Datagram Protocol (UDP) is used to transfer data from device

to device. Datagrams are used when a reliable delivery and two way communication are not required (as opposed to the session service). The datagrams can be sent to just one application, to a group of *netbios* applications, or to all applications. A datagram message will only be received, if the station is registered to receive a datagram, otherwise it will be ignored. The machines use *netbios* to register their names on the network by announcing it to the rest of machines of the sub-network. This auto-presentation is carried out by means of broadcasting.

HTTP and HTTPs are the most popular of end-user applications, (with 84 users (*Kamppi* location), 123 users (*Library-Post* location) and 130 users (*Stockmann* location)). Also interesting is the number of people who use their devices to run direct mail client (POP3 and IMAP), only 9 in *Kamppi*, 10 in *Library-Post* and 11 in *Stockmann*. An explanation to this can be that the users prefer to open a web-browser application as a mail clients instead of connecting directly to the device. Moreover, a low number of users still use session applications such *ssh*, showing that users do still need to connect to some other machines, which indicates a tendency to use devices as mere terminals.

Table 5.1: The most common applications by number of users in the *Kamppi* location, incoming traffic (MB), outgoing traffic (MB).

Application	Users	Volume In	Volume Out
Unknown	36	60.23 MB	29.23 MB
Http	51	59.62 MB	5.06 MB
Https	33	10.73 MB	0.49 MB
Domain	54	0.12 MB	0.04 MB
Pop3	3	69.6 KB	0.69 KB
Bootps	38	21.5 KB	11.6 KB
Imaps	3	24.5 KB	2.17 KB
Ntp	6	0.49 KB	0.43 KB
Hpvirtgrp	2	1.45 KB	0.17 KB
Teredo	3	0.061 KB	0.67 KB
Netbios-ns	54	0.0	85 KB
Mdns	60	0.0	109 KB
Ipp	3	0.0	94.2 KB
Netbios-dgm	30	0.0	51.2 KB

In Table 5.3 we find also some clients with house-keeping applications such *ntp* which is time synchronization software ported to almost every platform from PCs to supercomputers and embedded systems. The application mix can be run by mobile and static clients at different times of the day. We can conclude that there can be different application combinations from devices running in Windows and Macintosh OS, where the most commonly used service is web browsing, but other applications exist frequently too.

Table 5.2: The most common applications by user in the *Library-Post* location, incoming traffic (MB), outgoing traffic (MB).

Application	Users	Volume In	Volume Out
Http	64	111.27 MB	3.79 MB
Unknown	97	9.98 MB	2.48 MB
Macromedia-fcs	3	6.26 MB	0.0023 MB
Https	59	2.11 MB	1.04 MB
Domain	66	95.2 KB	4.09 KB
Teredo	5	12.08 KB	15.36 KB
Imaps	7	55.29 KB	2.15 KB
Tripe	3	94.2 KB	6.14 KB
Bootps	40	2.25 KB	19.04 KB
Pop3s	3	35.84 KB	1.19 KB
Http-alt	2	1.12 KB	2.17 KB
Nsp	2	0.328 KB	0.616 KB
Ntp	5	0.16 KB	0.81 KB
Sunwebadmin	2	0.23 KB	0.36 KB
Blackjack	2	0.28 KB	0.92 KB
Netbios-ns	121	0.019 KB	87 KB
Mdns	167	0.0	125 KB
Netbios-dmg	66	0.0	87 KB
Netbios-ssn	14	0.0	43.41 KB
Osu-nms	10	0.0	3 KB

Table 5.3: The most common applications by user in the *Stockmann* location, incoming traffic (MB), outgoing traffic (MB).

Application	Users	Volume In	Volume Out
Http	78	338.93 MB	23.92 MB
Unknown	27	37.48 MB	2.26 MB
Https	52	17.43 MB	7.58 MB
Macromedia-fcs	2	5.06 MB	0.023 MB
Domain	90	0.525 MB	0.332 MB
Tripe	2	5.48 MB	0.018 MB
Imaps	7	55.29 KB	60.41 KB
Ntp	20	29.69 KB	25.6 KB
Bootps	64	13.51 KB	72.6 KB
Pop3	4	17.2 KB	0.262 KB
Ssh	3	0.037	4.53 MB
Mdns	67	0.0	0,581 MB
Netbios-ns	36	0.0	0,150 MB
Osu-nms	21	0.0	12.28 KB
Netbios-dmg	10	0.0	10.74 KB

### 5.4.2 Analysis of HTTP traffic

HTTP (Hypertext Transfer Protocol) is a client-server communication protocol used to transfer web documents over the Internet. This section presents a study on HTTP web traffic analysis, because it has been the most important traffic component in our traces as we showed in the previous section. The following findings are based on analysis of HTTP pairs which consist of HTTP requests and HTTP responses. A HTTP client, most of the time a web browser, sends a HTTP request to a web server with the well-known Uniform Resource Locator (URL) field to locate the file. The web server answers with an HTTP response and provides to the client the sought after web page. However, it is usual that the number of responses can be different compared to the number of requests, if some retransmission were needed or the web page has large size.

We analyze the remote hosts which have been mostly used by the entire user population during the traffic capture period. The results allow us to estimate the interests of users connecting to the Internet. Several requests are found to apply for common services, including retrieving web content and some streaming connections, for example, music or video. In addition, clients can connect to a number of static services through a network, these include, for example, distributed file systems, distributed databases, and distributed multimedia systems. The servers generally wait for a request initiated by the users and simply process these request and return the results.

As a result, we show a table which collects the remote host diversity in each of the AP locations. It builds about 3 columns, containing the DNS-name of the remote host, the number of HTTP requests per domain and the request percentage applying for this particular HTTP host. Tables 5.4, 5.5 and 5.6 show the most popular HTTP hosts found in each location based on the number of HTTP requests applying for each service. These requests have been successfully sent by users connected to the observed networks. We summarize the number of requests for each host applied by users. Due to the high server diversity, we filter those with a low number of request on them. These estimations can indicate the user interests in each location.

When many users browse the same site, an optimal solution can be to share a cache. The proxy caches help by serving information which was previously demanded and they can avoid re-sending requests to the server. Consequently, server cache control

Table 5.4: The most popular HTTP hosts in the *Kamppi* location.

Domain Host	Num. Http requests	Percent (%)
mail.google.com	220	7.26
www.ning.com	177	5.83
static.ak.fbcdn.net	105	3.81
www.facebook.com	91	2.91
Photos-ak.fbcdn.net	88	2.9
www.palkkalaskari.fi	83	2.74
www.hel.fi	73	2.41
www.wikimedia.org	46	1.52
www.kikari.com	45	1.48
feeds.feedburner.com	73	2.41
idisk.mac.com	42	1.39
www.google-analytics.com	40	1.32
www.hif.fi	38	1.25
www.wired.com	36	1.19
www.google.com	31	1.02
www.youtube.com	30	0.92

can improve the site performance and reduce bandwidth consumption. If the proxy has the document in its cache it can just return the document, and if it does not, it can submit the request on behalf of the browser, store the result and relay it to the browser. So the proxy can reduce network traffic and in more advanced deployment the technology can be used as a hierarchical proxy cache. These tables show the candidates to be cached in the proxies that could be shared by customers. Note that, e.g, google mail cannot be cached, since it is private information.

Apart from that, the user HTTP requests reveal that social networks such as *Facebook* or *Ning* can be considered to have an important value in each location. Consequently, this suggests increment of these services in the social population. The observed patterns are suggesting that customers connect to the network motivated

Table 5.5: The most popular HTTP hosts in the *Library-Post* location.

Domain Host	Num. Http requests	Percent (%)
k.h.google.com	453	8.95
khmdb.google.com	227	4.48
www.kauppalehli.fi	164	3.24
www.hs.fi	119	2.35
www.google-analytics.com	111	2.19
profile.ak.fbcdn.net	97	1.92
www.facebook.com	78	1.54
profile.facebook.com	77	1.52
web.archive.rog	75	1.48
mw2.google.com	71	1.40
statik.ak.fbcdn.net	65	1.28
b.statik.ak.fbcdn.net	65	1.28
www.ytv.fi	30	0.59
mail.google.com	27	0.53
www.tkk.fi	21	0.41

Table 5.6: The most popular HTTP hosts in the *Stockmann* location.

Domain Host	Num. Http requests	Percent (%)
profile.ak.fbcdn.net	1656	4.14
statik.ak.fbcdn.net	1629	3.7
mail.google.com	1164	3.18
www.facebook.com	796	2.84
www.helsinki.fi	577	2.66
swscan.apple.com	387	2.38
clients2.google.fi	430	2.56
www.google.com	411	2.41
cha nnel.facebook.com	374	2.31
mw2.google.com	71	1.40
b.statik.ak.fbcdn.net	269	2.21
ww.google.fi	179	1.93

by fun purposes. However, the domain *mail.google.com* is very popular in both *Kamppi* and *Stockmann* locations, revealing that users had established email sessions. Table 5.5 shows services such as *ytv*, which indicates the presence of traveler users and *google* services for the student user population. However, workers can be related as a portion in each of the locations because of the presence of the *google-analytics* service which is used by the clients to analyze online marketing statistics and to make their sites more profitable.

From the point of view of caching, information about HTTP response packet sizes in the user sessions, can help to establish an understanding about what cache size should be efficient in the APs to offer service to users by using just the content stored in the APs. In that analysis, we set our attention on finding the largest sizes in the HTTP responses and which content items are popular. The user can have different application sessions during a WLAN session, thus, the process consist of analyzing all the packets sent by the server to acquire an average response size and to obtain the total packet size in order to understand how to best build the caches.

Several HTTP responses were fragmented because of the protocol data unit (PDU), which is a unit of Information of the protocol in the lower layers. It is used for the data exchange inside a layer of an OSI model. An application-layer message can be broken into several TCP segments so that all of them have uniform size. For instance, a single PDU contains all the segments that contribute a message. In order to estimate the HTTP response size, we use a reassembled PDU which combines all the TCP segments of an application message. We base on the HTTP responses with a code equal to *200 OK* which are those responses that were received successfully. The algorithm applied for this filter file is based on the content-length size for the

responses and they are included based on their content-type.

As explained earlier, several places with different AP locations were used to capture the traffic. Consequently, each location is analyzed separately to gain detailed understanding on the cache needs in the place where the traffic was captured. Tables 5.7, 5.8 and 5.9 show the HTTP responses size distribution from servers as analyzed in each location. Each row summarizes the percentage of packets found with a specific size. The HTTP responses are measured from the HTTP content-lengths and can be one of the object (turns) in a specific web resource size.

Table 5.7: Statistics of HTTP response by type in the *Kamppi* location.

<b>Kamppi</b>	<b>image</b>	<b>text</b>	<b>application</b>	<b>video</b>
<b>Pakt Length (bytes)</b>	<b>Percent (%)</b>	<b>Percent (%)</b>	<b>Percent (%)</b>	<b>Percent (%)</b>
smaller than 739	35.57	32.91	8.44	0.09
740-2559	9.02	2.75	1.8	0
2560-5119	4.15	0.81	0.45	0
5120-10239	1.39	0.67	0.49	0
10240-20480	0.54	0.27	0.18	0
bigger than 20480	0.09	0.13	1.17	0

We list 2215 HTTP 200 OK (reassembled) responses provided by servers and a total of 403 HTTP responses with content-length 0 which contain only encoding or encrypted information in the *Kamppi* location. Table 5.7 shows HTTP responses size less than 740 bytes as the most usual size sent by the servers. However, several responses contained large content length such as images with values over 5120 bytes and even more than 20480 bytes.

Table 5.8: Statistics of Http response by type in the *Library-Post* location.

<b>Library-Post</b>	<b>image</b>	<b>text</b>	<b>application</b>	<b>video</b>
<b>Pakt Length (bytes)</b>	<b>Percent (%)</b>	<b>Percent (%)</b>	<b>Percent (%)</b>	<b>Percent (%)</b>
smaller than 739	30.05	13.78	7.68	0.32
740-2559	12.79	4.26	4.12	0
2560-5119	5.62	1.82	0.7	0
5120-10239	2.57	1.26	1.07	0
10240-20480	0.75	0.28	0.28	0
bigger than 20480	0.14	0.14	0.23	0

We estimate the number of 2133 HTTP 200 OK responses in total provided by servers and a total of 289 HTTP responses with content-length 0 (encoding or encrypted information) in the *Library-Post* location. Table 5.8 shows HTTP response sizes between 740 and 2559 bytes as the most usual packet size sent by the servers. However, some image responses showed a large content length. As we can see from the video column, some HTTP responses with video content were collected with sizes less than 739 bytes which indicates that the servers retrieved flash videos.

Table 5.9: Statistics of Http response by type in the *Stockmann* locations.

<b>Stockmann</b>	<b>image</b>	<b>text</b>	<b>application</b>
<b>Pakt Length (bytes)</b>	<b>Percnt (%)</b>	<b>Percnt (%)</b>	<b>Percnt (%)</b>
smaller than 739	30,48	21,37	10,63
740-2559	11,86	5,15	3,53
2560-5119	9,02	2,54	1,23
5120-10239	3,96	1,67	0,86
10240-20480	2,8	0,88	0,62
bigger than 20480	0,24	0,78	0,68

Table 5.9 shows 11220 HTTP responses from the *Stockmann* location and 1295 with encoded content. We can observe that video HTTP responses were not collected from this location and the most usual responses sent by the server have a size between 740 and 5120 bytes. This result is consistent over all traces and it is the content HTTP size response most used by the servers when they want to send an object which pertain to a web resource. The number of HTTP responses more usual in each location were image responses, which indicate that the image constituted most of the object found in the web page visited by the clients. As we can see, the sizes responses seem to be smaller than expected due to the web pages commonly containing large objects.

In addition, we are interested in the average web resource size in order to estimate what size of cache should be suitable in our environment. As we observed before, several responses came from the same servers which provided the content for distinct GET request. We collect all the HTTP 200 OK responses which are related to a single GET request to avoid retransmissions. In addition, we focus on the server which were visited by more than one client as they are the best candidates to be chosen as web content to be cached. The total web resource size calculation for each domain from traces is not a exact result due to the packet losses. Therefore, we make an estimation about the retrieved content for a server when a client request their services. The algorithm developed follows the HTTP flows to enable mapping responses to a related request. The HTTP responses for each web resource are identified by using the destination port (client port) and it is also supposed that the web browser does not provide the same client port if the web resource is visited again. Each of the HTTP responses constituted an object from the web resource requested. The web resource size is the sum of all HTTP responses belonging to the same flow and respectively from the same server. The responses as is explained in section 3.5.3 can be sent by different HTTP connections. Tables 5.10, 5.11 and 5.12 show the web resource size estimations visited by the users for each location analyzed

in the study. The third column identifies the possible web resources content to be cached. These tables show the percentage of the servers visited by clients and the content retrieved from them.

Table 5.10: Statistics of web resource sizes in *Kamppi* location.

<i>Kamppi</i>		
Web resource size (bytes)	Servers with 1 client (%)	Servers with more than 1 client (%)
smaller than 2560	38.28	24.84
2560-10239	15.06	10.04
10240-20480	4.06	3.13
larger than 20480	2.09	1.88

Table 5.10 has a number of 478 web services with an average web resource size of 7214 bytes from the servers with more than one client. We can observe that at least 1.88% of the server visited by more than one client have a size more than 20480 bytes, which can indicate the minimum cache needs in *Kamppi*. The most popular web resource content has a small size which reveal that they might contain just a piece of text information or an image.

Table 5.11: Statistics of web resource sizes in the *Library-Post* location.

<i>Library-Post</i>		
Web resource size (bytes)	Servers with 1 client	Servers with more than 1 client
smaller than 2559	37,31	26,51
2560-10239	13,91	9,49
10240-20480	3,43	2,94
larger than 20480	3,92	2,45

The number of web services applied by the clients in the *Library-Post* location is 611 and 1075 in the *Stockmann* location. Table 5.11 and 5.12 show an average web resource size of 3049 bytes and 5648 bytes respectively for the servers with more than one client. 2.45% and 9.39% of the servers visited by one than more client both in *Library-Post* and *Stockmann* locations have a size more than 20480 bytes which indicate that the web contents were larger with objects such as images or applications.

Table 5.12: Statistics of web resource sizes in the *Stockmann* location.

<i>Stockmann</i>		
Web resource size (bytes)	Servers with 1 client	Servers with more than 1 client
smaller than 2559	27.25	21.48
2560-10239	9.2	11.9
10240-20480	4.65	6.32
larger than 20480	9.76	9.39

As we can see from the tables, the server responses vary considerably in size depending on the location. We have estimated different average web resource sizes

which indicates that a different cache size should be built in each location to serve user requirements. However, these results seem to be very small than normal use of Internet nowadays because of the large web content in the web pages (image, video, etc). The presented estimations are not accurate but could bring an orientation about the client usages in different locations, but in order to obtain the best cache size in the proxy it should be necessary to perform more specific captures by using HTTP only filters in every location. This could have reduced the global packet loss and drop.

## Summary

The WLANs were used by heterogeneous devices in each location. They were associated to the AP during short and long periods of time. The users can be classified as mobile and statics depending on the session duration and the traffic consumed. There is an important difference in the traffic consumed by the clients regarding to the location, revealing that the WLANs usage varied in each location. This can be caused by the user diversity in each of the analyzed APs. A high quantity of users did not transmit too much traffic even though they were continuously connected to the hotspots. This seems to indicate that a larger part of population connected to them may be mobile users. Because of the smaller amount of outbound traffic compared to inbound traffic volume which is a typical characteristic of web browsing, news group reading, and accessing email services. Wi-Fi hotspots provide local coverage to mobile users bringing services to their customer with acceptable RTT values. In addition, some web connections had a duration of less than or equal to 15 minutes. Finally, there were different application combinations from devices, where the most common service is HTTP, but other applications exist frequently too.

## 6 Conclusions and Future Work

This Master thesis presents an analysis of the Internet use in a free 802.11 hotspot network in Helsinki, Finland. We examine more than 50 hours of traffic traces from different Wi-Fi hotspot networks. We focus our analysis of locations that were used enough to enable evaluating the most important measures in the study. We combine analysis for three different geographical regions, 14 access points, traffic captures and times of the day. For these, we show statistics to evaluate and to gain an understanding of the user behavior. We show a comparison between the results obtained in the most used AP locations in order to understand the WLAN usage in public hotspots. Our measurements were gathered and aggregated using commercially-available tools. These tools help overcome the challenge of measuring and monitoring a large and heterogeneous WLANs.

Our study demonstrated the feasibility and effectiveness of remote non-intrusive wireless-side measurement in a public WLAN environment. This data collection approach enabled multi-layer analysis from the wireless layer to the application layer. Analysis of our traces identified several trends, including usages patterns, diverse network application usage and user activity. The analysis of our WLANs hotspot traces identified similarities and differences in WLAN behavior across heterogeneous sets of users. The number of customers associated with these APs is seen to be different depending on the WLANs and locations, and so is likely the interest of users choosing the network. This distribution reflects the setting of the network, where, for example, many users are gathered in a small area during scheduled times. Moreover, at least some customers have elevated inbound traffic, which is not limited by the capacity of the connection. Some clients can be identified as static or moving users based on the amount of traffic and connection duration in every location.

Analysis at different WLAN access points show variations in the WLAN usage, and the number of connected users is seen to vary between locations. This is caused by the user population diversity and challenges in AP usability. Clients connected to some WLANs can be statics transmitting large amount of data and travelers in the case of short sessions (e.g. checking email). Nevertheless, the inbound traffic collected in the hotspots is higher than the outbound traffic, so that the outbound traffic likely is requests, with only a small size, and the inbound traffic contains large responses with payload. For these reason, these APs served clients on web browsing, news group and email services. Moreover, a fraction of users (based on the amount of submitted data) did not use the services offered by the APs, and just performed

address resolution in the AP, suggesting that their devices were detectable, default settings.

Round trip time measurements for HTTP traffic were used to understand the mobile user experience where hotspots provide services to clients. On the one hand, a portion of users responses experienced high delay which could be caused by servers with availability or system problems. In this case, the delay can be also caused by packet loss and network congestion. On the other hand, user applications often had an acceptable value of delay. However, some clients experienced an elevated delay which could be caused by network conditions. In addition, HTTP delay disparity could be caused for instance, by large responses from images and caching services causing low-start penalty. Considering web session durations, some web connection times were estimated in the order of minutes by some clients. Moreover, a fraction of web session durations were lasting hours revealing that clients could be connected to work related services.

Analysis of wireless hotspot TCP and UDP traffic showed that web access is a major application, and there is only a small fraction of real-time audio/video applications. In particular, the HTTP protocol dominated traffic at every hotspots network. Thus, web browsing applications were the most used by clients during their WLAN sessions. In addition, unknown applications (port not matched to IANA numbers) were applied by users, which are likely to be P2P applications with arbitrary ports. However, the usage pattern mostly depended on the group of users present at each location. For this reason, applications such as interactive games were found just at the *Kamppi* location suggesting travelers were killing time while waiting for metros, trains and buses, etc, Macromedia applications at the *Library-Post* location or Internet printing applications in *Stockmann* location. In addition, the devices connected to these APs run different Operation systems, and so the traffic contains basics services (netbios and mDNS) for these systems.

HTTP traffic was the most important component in the traces, and it was analyzed to reveal the most common services retrieved by the users. This confirmed assumption that clients connecting to the networks at every location were more interested in web based email services than regular email clients, work related services and social networks. These locations could be candidates to deploy a cache (not in the case of email services) reducing the bandwidth consumption. In the web services provided by servers, image responses were the most popular objects found in each location with variability in the size. In addition, these images significantly increase

the web resource size average for domains. Based on the results, the suitable cache size should be considered based on location in order to serve users requirements more effectively.

A lot of research on user behavior and network performance in free WLAN networks has been done in the past, leading to numerous Internet usage characterizations. The capture methodology explained in this thesis can serve as a basis for further work and improvements, when a diverse mechanisms such as data analysis evolves. For example, the popularity of the streamed content indicates string interest to use such services. For this reason, new traffic captures focusing on this content with new mechanism of traffic collection seem interesting future directions. Although we have only collected and analyzed data from some WLANs, we believe that many of our observation are applicable in other settings. For example, many of the characteristics are an artifacts of human behavior, which can be extended in some WLAN environments. We believe that our results are helpful to a wide audience.

However, mobility habits in WLANs could be analyzed providing understanding about the user mobility restriction and giving knowledge of how users were leveraging the flexibility offered by a public WLAN hotspots. Specifically, knowing how many clients move, how they move, where they move, and how often they move.

Considering the challenges in monitoring and, for example, to improve the capture performance, it could be feasible to merge captures from different sniffers to obtain a better view of the network traffic. Another important factor that an efficient network performance analysis tool should consider is the distance of the APs from the clients and the sniffer respectively. In fact, a sniffer, in order to collect the highest amount of traffic, should be positioned near both AP and client in a strategic location.

Finally, the higher layer traffic capture serves as a valuable input to a final decision for some analysis. For example, in order to understand how caches should be built for web usage, traffic collected from clients should be analyzed by taking into account references to other protocols. This can help to gain deeper understanding about which caching mechanism are needed in each location.

## References

- [1] Ieee. 802.11b/d3.0 wireless lan medium access control (mac) and physical layer (phy) specification. August 1999.
- [2] D.Tang and M.Baker. Analysis of a local-area wireless network. pages 1–10. In Proceedings of MobiCom 00, August 2000.
- [3] D.Kozt and K.Essien. Characterizing usage of a campus-wide wireless network. Dartmouth College, March 2002.
- [4] Helsinki wlan hotspot. <http://ptp.hel.fi/wlan>, June 2009.
- [5] tcpdump - dump traffic on a network. <http://www.tcpdump.org>, June 2009.
- [6] Wireshark award-winning network protocol analyzer developed by an international team of experts. <http://www.wireshark.org/>.
- [7] Passive and active measurement conference. <http://pam2009.kaist.ac.kr/callpapers.html>.
- [8] D. Tang and M. Baker. Analysis of a metropolitan-area wireless network. pages 13–23. In Proceedings of ACM MobiCom 00, August 1999.
- [9] D.Kozt and K.Essien. Analysis of a campus-wide wireless network. In *In Proceedings of ACM MobiCom 2004*, pages 11:115–133. 2005.
- [10] David Kotz T.Henderson and I.Abyzov. The changing usage of a mature campus-wide wireless network. In *In Proceedings of ACM MobiCom 2004*, pages 187–201. Philadelphia, PA, Sept. 2004.
- [11] M.Balazinska and P.Castro. Characterizing mobility and network usage in a corporate wireless local-area network. 2003.
- [12] T.Henderson D.P.Blinn and D.Kotz. Analysis of a wi-fi hotspot network. Dartmouth College, Hanover, 2005.
- [13] G.M.Voelker A.Balachandran and P.Bahl. Wireless hotspots: current challenges and future directions. In *In: WMASH 03: Proc. of the 1st ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, page 19.

- [14] M.G.Voelker A.Balachandran and P.Bahl. Wireless hotspots: current challenges and future directions. In *In:WMASH '03:Proc. of the 1st ACM international workshop on Wireless mobile applications and services on WLAN hotpots*, pages 1–9. New York.
- [15] M.McNett and G.M.Voelker. Access and mobility of wireless pda users. In *SIGMOBILE Mob.Comput.Commun.Rev. 7(4)*, 2003.
- [16] K.Karvonen and J.Lindqvist. Usability improvements for wlan access. In *Department of Computer Science and Engineering, Helsinki University of Technology*. Finland, 2008.
- [17] M.S.Gast. 802.11 wireless networks: The definitive guide, sebastopol, ca:. In *O'Reilly & Associates*, 2002.
- [18] L.L.Peterson and B.S.Davie. *Computer Networks: A Systems Approach*. 2000.
- [19] The european conference of postal and telecommunications administrations. <http://www.cept.org>.
- [20] IEEE Standard for Information Technology IEEE, "IEEE Std.802.11a-1999. Telecommunications and information exchange between systems- local and metropolitan area networks- specific requeriments- part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications. Amendment 1: High-Speed Physical Layer in the 5 GHz Band, 1999.
- [21] IEEE Standard for Information Technology IEEE, IEEE Std.802.11b-1999. Telecommunications and information exchange between systems- local and metropolitan area networks- specific requeriments- part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications. Amendment 2: High-Speed Physical Layer Extension in the 2.4GHz Band, 1999.
- [22] J.Terry and J.Heiskala. *Ofdm wireless lans: Atheoretical and practical guide*. Indianapolis, Indiana:Sams, 2001.
- [23] Hiroyuki Yomo Petar Popovski and Ramjee Prasad. Strategies for adaptative frequency hopping in the unlicensed bands. pages 1–8. Aalborg University, December 2006. <http://kom.aau.dk/~petarp/papers/DAFH-AFR.pdf>.
- [24] D.Nations. Wireless local area networks. Sept 1997. <http://www.ac.wvu.edu/~n9649918/wlans.html>.

- [25] T.S. Rappaport. Wireless communications: Principles and practice, 2 edition. Upper Saddle River, Nj: Prentice-Hall, 2001.
- [26] International telecommunication union. <http://www.itu.int/ITU-R/terrestrial/faq/index.html>.
- [27] John A. Stankovic Gang Zhou, Tian He and Tarek Abdelzaher. Rid: Radio interference detection in wireless sensor networks. pages 1–11. University of Virginia, Charlottesville 22903, December 2006.
- [28] Wireless ethernet compatibility alliance. <http://www.weca.net>.
- [29] Jeong Geun Kim Brian P.Crow, Indra Widjaja and P.T.Sakai. Ieee 802.22 wireless local area networks. IEEE Communications Magazine, Sept 1997.
- [30] Joerg Ott. Application protocol design considerations for a mobile internet. Helsinki univesty of technology. Networking Laboratory.
- [31] G.Xylomenos and G.C.Polyzos. Tcp performance issues over wireless links. Athens University of Economics and Business, Greece, 2001.
- [32] Networ Working Group. Rfc. ip mobility support for ipv4. Nokia Research Center, August 2002.
- [33] Ian Goldberg Borisov, Nikita and David Wagner. Intercepting mobile communications: The insecurity of 802.11. In *Published in the proceedings of the Seventh Annual International Conference on Mobile Computing And Networking*, July 16-21, 2001.
- [34] Ian Goldberg Nikita Borisov and David Wagner. Security of the wep algorithm. In *Analysis of the Wired Equivalent Privacy (WEP) algorithm, which is part of the 802.11 standard*. <http://www.isaac.cs.berkeley.edu/isaac/wep-faq.html>.
- [35] H.Berghel. Wireless infidelity, acm. war driving.
- [36] Rubens et al. Rignay, Willats. Rfc 2865. remote authentication dial in user service. 2000.
- [37] A. Nagarajan. Rfc 3809. generic requirements for provider provisioned virtual private networks (ppvpn). June 2004.

- [38] Arunesh Mishra and William Arbaugh. An initial security analysis of the iee 802.1x security standard. February 6, 2001. <http://www.cs.umd.edu>.
- [39] Jim Geier. Implementing 802.1x security solutions for wired and wireless networks . April 14, 2008.
- [40] S.Antonatos D.Koukis and K.G Anagnostakis. Anonymization of measurement and monitoring data: Requirements and solutions. Autonomic Networking Group, Dept. of Computer Science 7, University of Erlangen, 2003.
- [41] G. Minshall. Tcpsdpriv: Program for eliminating confidential information from traces. <http://ita.ee.lbl.gov/html/contrib/tcpsdpriv.html>, 2005.
- [42] S.Antonatos D.Koukis and K.G Anagnostakis. On the privacy risk of publishing anonymized ip network traces. Infocom Security Department, Institute for Infocomm Research, Singapore, 2006.
- [43] D. Antoniadis E.P. Markatos D. Koukis, S. Antonatos. Anontool. a generic anonymization framework for network traffic. institute of computer science (ics). <http://www.ics.forth.gr/dcs/Activities/Projects/anontool.html>, 2006.
- [44] A.S.Tanenbaum. Computer networks, third edition. Englewood Cliffs.
- [45] Phil Karn and Craig Partridge. Improving round-trip time estimates in reliable transport protocols. In *Harvard University/BBN Laboratories*, 2005.
- [46] H. Kaplan E. Cohen and J. D. Oldham. Managing tcp connection under persistent http. 1999.
- [47] A. Baird-Smith E. Prudhommeaux H.W. Lie H. Frystyk Nielsen, J. Gettys and C. Lilley. Network performance effects of http/1.1. In Proceedings of the ACM SIGCOMM97 Confer-ence, 1997.
- [48] R.Droms. Dynamic host configuration protocol. Network Working Group with Category of Standars Track, March, 1997.
- [49] Moustafa Youssef Jihwang Yeo and Ashok Agrawala. Characterizing the iee 802.11 traffic: The wireless side. pages 1–29. University of Maryland, March 1, 2004.

- [50] Tool used for packet capturing of raw 802.11. August 2009. <http://www.aircrack-ng.org>.
- [51] Scott Renfro. Mergecap. merges two or more capture files into one. <http://www.ethereal.com>.
- [52] Gerald Combs. Tshark - dump and analyze network traffic. <http://www.wireshark.org>.
- [53] Van Rossum. Python. high-level programming language. <http://www.python.org/>, June 1991.
- [54] Travis Oliphant. Numpy and scipy documentation. <http://docs.scipy.org/doc/>.
- [55] John Hunter. Matplotlib 2d plotting library. <http://matplotlib.sourceforge.net/>.
- [56] Ajay Kr Singh and Sridhar Iyer. Improving tcp performance over wireless environments. pages 1–5. IIT Bombay. <http://www.it.iitb.ac.in/~sri/papers/atcp-mwcn02.pdf>.
- [57] International Telecommunication Union. <http://www.itu.int/en/pages/default.aspx>.
- [58] Shawn Ostermann. Tcptrace. <http://www.tcptrace.org/>, Ohio University.

## Acknowledgements

Finally my graduation came which I looked forward to it. But I have to remarks that the entire study preparation during all these years have absolutely been perfect, even with the hard moments, pressures and stresses. At the end, all these feelings helped me a lot to be a better person and confront every situation successfully. Moreover, I was lucky meeting all the people on my way- from Spain to Finland, those who experienced with me all the moments. The future is yet to come, and this final experience give me the illusion to continue fighting hard.

Foremost, I would like to thank Professor JÖRG OTT for giving me the chance to join his research group and complete my Master's Thesis. A really good boss, always taking care of everything. Additionally, I would like to thank MIKKO PITKÄNEN for being my Instructor, always helpful to suggest ideas, and WILLIAM MARTIN, the faculty thesis revisor, for his textual revision work. And I am deeply grateful to the Aalto University School of Science and Technology for the chance to study in this university, specially JENNI TULENSALO who helped me with all the technical and bureaucratic stuff, always with a smile on her face.

Nevertheless, I have to be grateful for everything to my family, for the whole support during these years. My mother MARI has overturned all her life for mine has done a perfect job, always advising me and giving me freedom to fight for my own dreams. She has been really the best mother in the world and, even from thousand kilometres away, has always given me the love and words I needed. I also thank to my grandmother CARMEN for everything, the most sweet person possible, my aunt LOLI for her love and help studying in Barcelona and my aunts SONIA and BIBI, for all they have always given me.

A really special thank goes to my boyfriend JOSE, who has always been my right hand supporting me in bad moments, giving me love to make this dream possible, helping with all the problems and surrender me the warm words for calm to face up to every situation.

Also, I wish to thank everybody with whom I have shared experiences in life. From the people who first persuaded and got me interested into the study of telecommunications, especially those who also played a significant role in my life, to those whom with the gift of their company made my days more enjoyable and worth living. Pointing out VICKY, in particular, for sharing with me a very special friendship, years of lessons and enjoyable moments, really the best sister/mate possible.

And last, but not least, I really want to give a big thank to all the people with whom I have shared this finish experience. Those people whose have been my family during these years. First of all, I have to thank MARIA for always being such a nice company and friend. I also thank FISH and JING, my sweet Chinese roommate and my lovely Chinese girl. Thank you so much every body, with all my love.

Otaniemi, March 1th 2010

Miriam González

## WLAN traffic summary

<b>KAMPPI PLACE</b>	<b>(bytes)</b>		<b>(Mbits)</b>
<b>3 Hours</b>	<b>TCP</b>	62333167	498,67
	<b>UDP</b>	418695	3,3496
	<b>Total</b>	62751862	502,01
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage</b>
Helsinki Kaupungin WLAN	81666	-44	97,05
Others	2479		2,95
<b>1:41 Hours</b>	<b>TCP</b>	2145053	17,16
	<b>UDP</b>	51398	0,41118
	<b>Total</b>	2196451	17,572
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage</b>
Helsinki Kaupungin WLAN	3575	-44	92,35
Helsinki Kaupungin WLAN	126	-79	3,25
Others	160	-82	4,4
<b>2:13 Hours</b>	<b>TCP(bytes)</b>	4698594	37,589
	<b>UDP</b>	129191	1,0335
	<b>Total</b>	4827785	38,622
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage</b>
Helsinki Kaupungin WLAN	7065	-53	88,78
Lasipalatsi	351	-83	4,41
Helsinki Kaupungin WLAN	121	-82	1,52
Lasipalatsi	377	-47	3,95
<b>3 Hours</b>	<b>TCP</b>	112458479	899,67
	<b>UDP</b>	8359783	66,878
	<b>Total</b>	119835155	958,68
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage</b>
Helsinki Kaupungin WLAN	571	-42	38,9
fifiwifi	792	-73	2,1
Lasipalatsi	148326	-74	43
Lasipalatsi	1327	-87	16

Figure A.1: Traffic summary in *Kamppi* location

<b>STOCKMAN PLACE</b>		<b>(bytes)</b>	<b>(Mbits)</b>
<b>1:30 Hours</b>	<b>TCP</b>	5994202	45,732
	<b>UDP</b>	79989	0,63991
	<b>Total</b>	6074191	48,594
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Kafka	9045	-45	92,37
Helsinki Kaupungin WLAN	398	-82	4,06
Helsinki Kaupungin WLAN	295	-82	3,01
Others	0	-82	1,61
<b>2:10 Hours</b>	<b>TCP</b>	1761185	14,089
	<b>UDP</b>	91508	0,73206
	<b>Total</b>	1852693	14,822
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Kafka (7)	3062	-41	79,86
Helsinki Kaupungin WLAN	387	-78	10,09
Helsinki Kaupungin WLAN	355	-76	9,26
Others	30	-81	0,79
<b>2:45 Hours</b>	<b>TCP(bytes)</b>	25141443	201,13
	<b>UDP</b>	804214	6,4337
	<b>Total</b>	25945657	207,57
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Helsinki Kaupungin WLAN	15612	-76	35,7
Kafka	23900	-59	54,66
Others	3398	-80	9,64
<b>1 Hour</b> <b>Total Users: 10</b>	<b>TCP(bytes)</b>	2050307	16,402
	<b>UDP</b>	16860	0,13488
	<b>Total</b>	2067167	16,537
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Helsinki Kaupungin WLAN	3866	-78	100
<b>2:40 Hours</b>	<b>TCP</b>	18800003	143,4326
	<b>UDP</b>	303033	2,3119
	<b>Total</b>		
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Kafka	34397	-68	45,57
Helsinki Kaupungin WLAN	1029	-78	28,57
Others	486	-82	10,6

Figure A.2: Traffic summary in *Stockmann* location

<b>LIBRARY POSTY</b>		<b>(bytes)</b>	<b>(Mbits)</b>
<b>2:30 Hours</b>	<b>TCP</b>	7701691	58,7592
	<b>UDP:</b>	569421	4,5554
	<b>Total:</b>	8271112	66,169
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Hsverkko	32216	-75	100
<b>2 Hours</b>	<b>TCP</b>	53131762	405,363
	<b>UDP</b>	1883506	14,37
	<b>Total</b>	55015268	454,1707
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Stadinetti	46659	-69	56,98
Hsverkko	32222	-82	39,35
Others	3003	-80	3,67
<b>2 Hour</b>	<b>TCP</b>	53973790	431,79
	<b>UDP</b>	349908	2,7993
	<b>Total</b>	54323698	434,59
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Stadinetti	56119	-47	82,61
Hsverkko	5090	-78	7,59
Stadinetti	4442	-82	6,63
Helsingin kaupungin WLAN	1204	-82	1,8
<b>3 Hours</b>	<b>TCP</b>	71064961	568.52
	<b>UDP</b>	2639988	21.12
	<b>Total</b>	73704949	589.64
<b>SSID</b>	<b>DataPackets</b>	<b>Power</b>	<b>Percentage(%)</b>
Stadinetti	74691	-62	83,37
Hsverkko	10267	-57	11,46
Helsingin kaupungin WLAN	1772	-74	1,98
Others	2856	-80	3,19

Figure A.3: Traffic summary in *Library-Post* location